University of Sheffield

Enabling Better Healthy Food Recommendation By Incorporating Contextual Knowledge

Mengyisong Zhao

Submitted in Partial Fulfilment of the Requirements for the Degree of Doctor of Philosophy

in the

Information School

April 8, 2025

To my dearest and loving mother, father and husband

Acknowledgements

First and foremost, I would like to express my deepest and most heartfelt gratitude to my funding body, Mrs. Song and Mr. Zhao - my dearest mother and father. They have been the kindest supporters throughout every step of my PhD journey. I am immensely thankful to them for being the best parents in the world and for giving me the freedom and opportunity to pursue what I truly wish and desire. Their unwavering love and support have empowered and equipped me to face every challenge with courage and resilience.

My PhD journey began during the COVID pandemic, and I owe immense gratitude to my loving husband, Mr. Cai (soon to be Dr. Cai), who has been by my side every step of the way. I consider myself luckier than most, as with him, I have always been able to be my true self. He has supported every major decision I've made, offering everything he has and everything he can. Being with him make me believe that there is someone in this world who will love everything about you - you just need to love yourself and find that person.

More importantly, I would like to express my sincere gratitude to the best supervisors in the world: Dr. Morgan Harvey, Dr. Dave Cameron, and Professor Frank Hopfgartner. Their unwavering support, encouragement, and inspiration in every meeting have made my PhD journey truly rewarding and fulfilling. With a blend of humour, empathy, and open-mindedness, they created a supportive and nurturing environment that allowed me to play to my strengths, offering encouragement and inspiration for every decision I made rather than being overly critical or pushy. They celebrated even my smallest achievements and patiently tolerated my mistakes, lifting me up whenever I lost confidence. My supervisors could always see through my worries, offering wisdom and reassurance that kept me grounded. It has been a great honour and privilege to be their student, and I feel incredibly fortunate to have had them as supervisors. By their side, I am motivated to keep striving and becoming the best version of myself. I wish they could always be my supervisors, so that I could remain the happiest student. No matter where I go or what stage of life I am in, I will always feel proud and grateful about my PhD journey because of their guidance and support.

I am also deeply grateful to Dr. Sophie Rutter for her caring, invaluable guidance and support in both my teaching and personal life, which has helped me grow in confidence. Without her support and endless encouragement, I would not have achieved what I have today. I want to express my heartfelt thanks to my Iron Man, Dr. Suvodeep Mazumdar, for his kindness, sense of humour, and constant support. His invaluable guidance has been a cornerstone of my project, and I am truly grateful for his encouragement every step of the way. I would like to sincerely thank Dr. David Elsweiler for his support and for consistently providing me with opportunities to present my work and connect with others, helping me become more self-assured. His insightful discussions and invaluable guidance have greatly enriched my project, and I am truly grateful. Special thanks to Dr. Pamela Abbott for her encouragement and reassurance during times when I felt uncertain about the road ahead, and to Dr. Monica Paramita and Dr. Itzelle Medina for their insightful discussions and support on my PhD research and personal growth.

I would like to thank my friends, Xuanning, for her companionship, unwavering support and exceptional encouragement. With her, I feel more confident about the way forward. She often helps me and guides me to make better decisions. I am also grateful (not too grateful) for her feeding, which, although delicious, caused me - already not exactly slim - to put on a few more pounds. Sincerely thanks to Omaima for her friendship, patience, and encouragement. She has been like a sister to me, tolerating my quirky ideas and hobbies. Her care and support (sometimes with a touch of sarcasm) made my PhD journey truly memorable. I constantly miss our free-flowing conversations and hearty laughter. I would also like to thank Jie, who greatly enhanced my sense of belonging in the United Kingdom. She guided me in quickly learning the local culture. I am especially grateful for her encouragement during times when I doubted myself. Special thanks to Zied for his honesty in expressing bold ideas and for the incredibly insightful and joyful conversations. Heartfelt to James and Moritz for their endless support whenever I needed it. Finally, sincere thanks to Yuyang, Haoyu, Siyang, Ziyong, Jess and Xinyu for all the support and joy they have brought to both my research and personal life.

I am also grateful to the Information School family and all the staff who have supported me throughout my PhD. At last, I extend my great gratitude to all the participants in my research. Their involvement made this work possible.

Publications

The second stage of this research has been published in a peer-reviewed conference, as listed below. Two papers are currently in preparation. The third stage of the research is targeting the *Information Processing & Management journal*, while the first stage of the research is targeting the *iConference*.

Peer-reviewed conference publications

Zhao, M., Harvey, M., Cameron, D., & Hopfgartner, F. (2024). The Effect of Simulated Contextual Factors on Recipe Rating and Nutritional Intake Behaviour. In *Proceedings of the 2024 Conference on Human Information Interaction and Retrieval* (pp. 97-107).

In preparation

Zhao, M., Harvey, M., Cameron, D., & Hopfgartner, F. Enabling Better Healthy Food Recommendation by Incorporating Contextual Knowledge (*Ready to Submit, Target: Information Processing & Management Journal*)

Zhao, M., Harvey, M., Cameron, D., & Hopfgartner, F. How Contextual Factors Shape Individual Food Choice and Intake Behaviour (*Ready to Submit, Target: iConference*)

Abstract

Food recommender systems (FRS) are increasingly recognised as valuable tools for simplifying food decision-making, while also promoting healthier eating habits. However, accurately predicting user preferences in this context remains a challenging task: choosing what to eat is a multifaceted and highly complex process. The greater challenge lies in recommending food that users will not only enjoy, but that is also healthy, requiring work that goes beyond the typical recommender system (RS) algorithms, which focus on maximising expected ratings and typically may encourage people to make less healthy choices.

This thesis seeks to improve food recommendation performance and address the taste-healthiness trade-off challenge and by integrating several forms of contextual knowledge, including situational context. Very little extant research has systematically explored the impact of multiple dynamic factors on influencing people's eating habits, recipe ratings, and nutritional intake behaviours. Most existing research focuses on single contextual factors, typically simple extrinsic ones such as location and time.

This research addresses these gaps by understanding daily eating habits through semi-structured interviews, followed by a large-scale experimental study in simulated contexts to examine how these contexts influence recipe rating behaviour and, therefore, implied nutritional intake. The results highlight the importance of developing context-aware recommender systems (CARS) leading to the development of novel one-stage and two-stage contextual modelling approaches with multimodal feature sets. An innovative method for generating healthy recommendations and a novel evaluation approach are also proposed.

Key findings suggest that dynamic contextual factors like emotions, busyness, seasons, and physical activities influence food choices and decision-making. People's preferences for healthy recipes vary with context, especially under stress. Incorporating such contextual features into RS algorithms significantly improves recipe rating predictions and permits healthier recommendations to be made. A contextual healthy recommendation approach and a novel evaluation metric, RMSEh, are introduced to align recommendations with individual preferences whilst promoting healthier choices.

This research has implications for the future development of FRS, and shows that emotion-aware systems could lead to better healthy food recommendations. These findings provide valuable insights into the design of more sophisticated healthy food recommender systems and offer a promising framework for the development of context-aware recommender systems across various application domains.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background and Motivation

In today's digital age, where the Internet permeates nearly every aspect of our daily lives, the accessibility and availability of social media, content creation platforms, e-commerce, and digital communication have led to an exponential growth in the volume of data generated globally. This rapid increase in data has resulted in information overload, making it increasingly difficult to obtain specific knowledge and resources that meet individuals' needs (Khusro et al., 2016). Recommender Systems (RS) have been shown to be a potential solution to empower users in overcoming information overload problems in diverse domains, such as books (Chandak et al., 2015; Alharthi et al., 2018; Afchar et al., 2022), music (Yoshii et al., 2008; Schedl, 2019), movies (Christakou et al., 2007; Katarya & Verma, 2017), or restaurant recommendations (Park et al., 2008). In some leading companies, RS have played a vital role in their business models, as in the cases of Amazon.com, Spotify, Netflix, IMDB, and LinkedIn.

Although seminal work in RS has traditionally focused on other application domains, research in food recommendation has garnered increasing attention in recent years (Trattner & Elsweiler, 2017a; Min et al., 2019a). Proper food consumption is vital to sustaining human life, yet making daily food choices, while seemingly simple, can be quite challenging. Individuals are often overwhelmed by the vast number of available food options, leading to decision fatigue (Sobal et al., 2006). Additionally, the need to balance personal preferences with dietary diversity can further complicate the decision-making process (Furst et al., 1996), making food recommendation systems all the more essential and valuable (Trattner & Elsweiler, 2017a). More importantly, in today's fast-paced world, food marketing strategies and convenience often drive subconscious food choices, further discouraging individuals from prioritising healthier options (Chandon & Wansink, 2012). Consequently, there is growing concern over diet-related health issues, such as diabetes and obesity (Mozaffarian, 2016). According to data from the World Health Organization, in 2022, approximately 1 in 8 individuals worldwide were living with obesity, with over 2.5 billion adults aged 18 and older and 37 million children under the age of 5 classified as overweight (Organization, 2024). Furthermore, approximately 463 million people suffered from diabetes in 2019, and the prevalence rate is projected to double by 2045 (Aschner et al., 2021). Evidence suggests that unhealthy dietary patterns may significantly associated with obesity (Seifu et al., 2021; Jayedi et al., 2020). These alarming statistics have heightened awareness regarding the importance of adopting a healthy lifestyle and embracing nutritious eating habits to enhance both psychological and physical well-being (Sobal et al., 2014).

Food recommender systems (FRS) equipped with the capability to sift through vast amounts of information, could potentially excel in suggesting food products that align with users' dietary preferences while meeting their essential biological and physiological needs. Technological advancements and increased access to information have provided individuals with more opportunities to obtain nutritional guidance, plan their diets, and monitor their food intake. As a result, online food portals have become increasingly convenient and user-friendly. Additionally, the increasing availability of open-access recipe and user interaction data has made the development and implementation of FRS more feasible. But with the overwhelming variety of food products available and the influence of advertising, people can easily be enticed by visually appealing yet unhealthy options (Trattner et al., 2018; Starke & Trattner, 2021). Under these circumstances, FRS are expected to play a pivotal role in promoting healthier lifestyles by recommending nutritionally balanced meals tailored to individual health goals. Yet, fulfilling this role is accompanied by considerable complexities (Harvey et al., 2012; Min et al., 2019b; Berkovsky & Freyne, 2010).

As understanding the nature of food choices involves a complex interplay of dynamic contextual, multilevel, integrated and diverse factors (Sobal et al., 2014; Elsweiler et al., 2017), making accurate recipe recommendations becomes a significant challenge. Moreover, balancing users' preferences with their nutritional requirements presents an additional challenge (Trattner et al., 2017a). For instance, users who generally prefer unhealthy food options might not find value in a recommendation system if it consistently suggests healthier recipes that do not match their tastes. Research by Trattner et al. (2017a) demonstrates that internet recipes tend to be less healthy compared to those developed by professional chefs and even ready-made meals from leading UK supermarkets. Subsequent studies revealed that this issue is further exacerbated by users' tendencies to choose less healthy recipes, along with recommendation systems' biases toward suggesting popular recipes, which are often unhealthy (Trattner et al., 2018; Starke et al., 2020). As a result, food recommendation systems face the challenge of finding healthier alternatives while still meeting users' taste preferences, particularly when using traditional recommendation techniques (Starke, 2019).

From a user behaviour perspective, individuals may not consistently maintain an unhealthy diet in all circumstances; specific contextual situations might trigger or encourage particular healthy or unhealthy eating habits (O'Connor et al., 2008; Cameron et al., 2015). For example, emotional fluctuations - feeling stressed or overwhelmed by a busy schedule - or attending parties or group activities may cause people to deviate from their normal eating patterns. Therefore, examining whether users exhibit varying dietary and nutritional intake patterns in different contexts could provide a valuable breakthrough in avoiding blanket health recommendations that might overwhelm or frustrate users. Moreover, it is important to identify when it is most effective to offer healthy recommendations and how to incorporate nutritional information in a way that minimises the likelihood of recommendation rejection.

1.2 Problem Statement

Most recommender system (RS) approaches in the food domain are based on two-dimensional models (users $\times$ items), such as Content-Based Filtering (CBF) and Collaborative Filtering (CF) (Fakhri et al., 2019; Siddik & Wibowo, 2023; Arulprakash et al., 2024). CBF relies on the characteristics of the items themselves, while CF draws simply on users' past interests, gathered through either explicit (direct user ratings) or implicit feedback (e.g. click through data, time spent on items). However, relying solely and predominantly on traditional CF and CBF algorithms to compute similarities between recipes or users may not adequately predict users' potential needs. Furthermore, these approaches often struggle to promote healthier eating habits due to their inherent limitations. Specifically, these algorithms are not typically designed to consider factors such as health objectives or nutritional content when generating recommendations, thus limiting their ability to guide users toward healthier choices (Starke, 2019; Elsweiler et al., 2017).

The rise of mobile technologies and the Internet of Things (IoT) has enabled RS to integrate users' contextual factors (e.g., time, location, physical activity, or electrocardio signals) into traditional two-dimensional models, allowing for more personalised and context-sensitive recommendations (Vallejo-Correa et al., 2021). These context-aware recommender systems (CARS), which incorporate domain knowledge of contextual information, may provide opportunities to simultaneously satisfy user preferences and meet nutritional needs, thus providing more intelligent and tailored recommendations. After all, individuals' food preferences extend beyond taste; factors such as seasonality, emotional status, and physical activities may all influence food decision-making and dietary expectations (Kusmierczyk et al., 2015a; Macht, 2008).

CARS may enhance predictive capabilities by incorporating and aggregating contextual information to refine recommendations, as user preferences may evolve in response to changing contexts (Abusair et al., 2021; Aghdam, 2019). For instance, recipes favoured during a stressful week at work may not hold the same appeal during a relaxing weekend, and those popular on hot summer days may not elicit the same interest during cold winter periods (Aghdam, 2019). As CARS have the capacity to construct more comprehensive user profiles by accounting for preferences and nutritional intake behaviours across multilevel contextual situations. Such approaches may offer a viable solution to the longstanding health versus taste conundrum. By incorporating expert nutritional knowledge, these systems may be able to generate recipes that are both nutritionally balanced and appealing, tailored to specific contextual conditions, thus may enhance user acceptance. Moreover, the integration of multilevel contextual factors has the potential to markedly improve the accuracy of rating prediction models, aligning recommendations more closely with users' historical preferences. Then, refining recommendation lists based on nutritional requirements and health considerations may result in more appealing, health-conscious recommendations that are likely to be better received by users. These approaches represent untouched territory within the FRS domain.

Cross-disciplinary insights from psychology and food science highlight numerous factors that may influence people's food choices and nutritional intake behaviours. Food choices may not be solely driven by taste; factors such as cost, convenience, self-identity and health concerns can also sway the decision-making process (Connors et al., 2001). More specifically, among various types of factors, contextual factors - both static (relatively constant over time) and dynamic (changing based on context or time) - may significantly reshape people's food and nutritional preferences. Such as emotions and mood may further regulate eating and nutritional intake behaviours, and vice versa (Bove et al., 2003; Sun et al., 2011; Macht, 2008). For instance, a positive mood may point to healthier food choices (Cameron et al., 2015), whereas negative emotions and moods may lead to indulgence in snacks and increased caloric intake (Dingemans et al., 2009). Given the vast number of contextual factors present in real-life situations, capturing and measuring all of them is extremely challenging. To date, the most influential factors on people's food choices and nutritional consumption behaviours remain unclear, as no comprehensive studies have been conducted to address this complexity. Such insights would be crucial for the development of context-aware healthy food recommender systems.

Few studies in FRS has made noteworthy progress in exploring the relationships between contextual factors and an individual's (online) food choices. Factors such as gender (Cavazza et al., 2015), hobbies (Trattner et al., 2017b), time (Kusmierczyk et al., 2015a) and location (De Choudhury et al., 2016; Cheng et al., 2017) have been highlighted. However, previous research has primarily focus on investigating the user's online rating and uploading behaviour under a single contextual condition, rather than making healthy food recommendations. Most of these studies have predominantly centered on ingredient-based recommendations (Harvey et al., 2013; De Pessemier et al., 2013), while some have focused exclusively on addressed healthy recommendations without considering users' past preferences, or vice versa (Gao et al., 2019; Ueta et al., 2011).

Recently, a large body of research in other application domains of RS has focused on developing novel algorithms of context-aware recommender systems, including for music (Gong et al., 2020; Wu & Sun, 2024; Grigorev et al., 2024) and Point of Interest (Livne et al., 2019; Noorian, 2024) recommendation. However, up to now, no research has focused on experimented with various dynamic contextual factors and explored whether people's food choices and nutritional intake behaviours varied under these contextual situations. This gap is particularly evident when considering psychological factors, such as emotions, stressfulness, which are complex and difficult to control. As a result, challenges remain in identifying the most influential contextual factors that impact (online) food choices and the healthiness of the selected food and recipes. More critically, integrating these contextual factors into RS and addressing them algorithmically to deliver personalised, healthier recommendations remain additional challenges (Trattner & Elsweiler, 2017a; Rokicki et al., 2017). More specifically, in the field of FRS, it remains unclear whether incorporating dynamic contextual factors - those that fluctuate over time - would enhance model performance. Moreover, the optimal combination of feature sets for achieving superior results is still uncertain, and the effectiveness of providing healthy recommendations based on diverse contextual scenarios has yet to be fully understood.

This research seeks to address existing gaps by investigating how novel contextual factors—such as emotional status, physiological state, and physical environment—affect online food choices (as reflected in ratings) and nutritional intake behaviours. The insights gained will inform how these factors can be effectively integrated into CARS to enhance performance over traditional CF and CBF models, and to determine which feature combinations with contextual factors yield the best model performance. Additionally, to address the trade-off between user preferences and recipe healthiness, this study proposes a novel approach that integrates weighted dynamic contextual features and recipe health levels into predicted ratings, aiming to facilitate more adaptable healthy recommendations. The overarching research questions of this work are as follows:

1. What contextual factors affect people's (online) food choices?
2. What impact do these same factors have on people's nutritional intake?
3. Can integrate these contextual factors enhance the performance of recommendation systems?
4. How to combine this knowledge of contextual factors to recommend people healthy recipes that they will enjoy?

1.3 Research Aims and Objectives

The aims of this research can be summarised in three-fold. This first aim is to broadly and qualitatively explore how contextual factors affecting and reshaping people's food (online) choice and nutritional intake patterns, to understand people's perceptions and needs towards food decision-making process. After identifying and refining the potential contextual factors, the next aim is to quantitatively examine how specific types of contextual factors, such as *"seasonal", "emotional", "busyness" and "physical activity"* affect people's recipe rating and implied nutritional consumption behaviour. Lastly, this research aims to determine whether the integration of contextual factors can improve the performance of recommendation system models, and further proposes a context-aware healthier recipe recommender system that balances users' food preferences with nutritional needs to promote healthier eating behaviours.

This will be achieved by first interviewing individuals about their daily eating expectations and the factors that influence their decision-making processes. Following this, an experimental study will be designed to gather user recipe ratings under various simulated contextual scenarios for subsequent analysis. Next, the best rating prediction model will be identified through the evaluation and integration of both static and dynamic contextual features, alongside multi-modal user demographic data, recipe content, and recipe image information. These features will be incorporated into three machine learning models—Extreme Gradient Boosting (XGBoost), Ridge Regression, Support Vector Regression (SVR)—and a deep learning model, Multilayer Perceptron (MLP). A novel re-ranking approach will be proposed to adapt healthy recommendations to various contextual situations, aligning with a preference and health balance evaluation metric. The research objectives are as follows:

1. To understand people' perceptions towards food choice and needs under different contextual situations.
2. To identify the most influential (novel) contextual features that might affect people's food choice and acceptance of recommendations.

3. To explore how people's implied nutritional intake behaviours vary under the identified contextual features.
4. To examine state-of-the-art context-aware recommendation techniques, particularly in the food recommendation domain, and determine whether contextual features have demonstrated improvements in model performance.
5. To develop novel contextual healthy food recommendations that effectively address the trade-off problem between user preferences and nutritional needs.

Note that a common issue in food recommender systems work is the difficulty in directly measuring users' cooking and eating behaviour, as preparing food requires a significant amount of effort (Min et al., 2019a). Biologically speaking, the urge to eat and cook a certain food is usually driven by a liking or preference for a particular food or recipe (Rozin, 2007). As such, a high rating can be considered a proxy for the future intention to consume a certain recipe. The behaviour measured in this research is implied behaviour, and the discussion and implications of this research are based on this assumption.

1.4 Importance of the Study

This research not only builds upon existing literature but also introduces several novel perspectives. To the best of our knowledge, this is the first comprehensive study to explore how both static and dynamic contextual factors influence individuals' eating and nutritional intake behaviours. This was achieved by gathering insights directly from participants' experiences, reflections, and expectations. No prior research has examined the impact of dynamic contextual factors on recipe ratings and nutritional intake behaviours from the perspective of FRS. By conducting a large-scale between-subject experimental study, simulated contextual scenarios were created to collect user ratings, highlighting how food choices and nutritional intake behaviours vary across different contexts. These findings offer critical insights for incorporating contextual and health information into FRS. Moreover, previous studies have yet to fully explore the integration of multiple contextual factors when developing recommender systems in the food domain. This study advances the field by proposing a novel context-aware contextual modelling approach for FRS that integrates multi-modal features to address the trade-off between health and taste, while promoting the potential for increased acceptance of healthier recommendations. The three main ways in which this thesis advances the field of food recommendation are outlined in detail below:

- *To fill in the gaps and provide new insights into peoples' food intake perceptions and needs, particularly in exploring how static and dynamic contextual features influence their food choice and nutritional intake behaviour from the benefit of FRS perfective.*

  Most existing research focuses on investigating users' recipe rating and upload behaviour instead of making healthy food recommendations. Although contextual information has been considered in previous studies, most extant research only adopted single contextual scenarios. Given the vast number of contextual factors present in real-world situations, collecting and monitoring their impact on people's behaviour is challenging. Rather than randomly selecting contextual features to build a context-aware recommendation model, it is crucial to first identify the most influential factors, based on users' experiences and feedback, from the large number of available options.

- *Introduction and provision of a novel large-scale dataset to study context-aware healthy food (recipe) recommender systems, illustrating novel context-aware contextual modelling rating prediction algorithms for food recommendation.*

  Through the careful design of a between-subject experimental study, this study made in-depth exploration and development of context-aware food recommender systems possible. User ratings were collected under seven contextual scenarios, alongside a context-free group, enabling researchers to make clear comparisons. This study highlights a promising direction in developing context-aware FRS by emphasising the critical role of integrating dynamic contextual features. This study pioneers the capture, handling, and modelling of several dynamic contextual features within recommender system algorithms.

  To achieve more accurate rating predictions, it is imperative to move beyond traditional factors, such as taste and ingredients alone. Incorporating cooking directions and image features, may both contribute to improved model performance. This research holds potential benefits for researchers and developers in recommender systems and natural language processing domains, encouraging exploration of various model structures to achieve more satisfactory results in modelling and predicting individuals' eating and recipe selection behaviours.

- *To propose a novel approach for providing adaptive and context-sensitive healthy recommendations by integrating the weight of contextual information and recipe nutritional levels to balance people's food preferences and nutritional needs.*

  Instead of focusing on merely improving recommendation performance, which has already been shown to lead to unhealthy eating behaviour. This research focuses on seeking a new means of addressing the health/taste trade-off. While integrating the nutrition information is likely to reduce the prediction accuracy to some extent, incorporating contextual information beforehand may allow for flexibility in finding the threshold that provides both tasty and healthy recipes. The core idea of providing healthy recommendation in this research involves identifying the best rating prediction model to align with a user's preferences, and then refining these recommendations by substituting healthier options under specific contexts. This research has provided a novel approach to delivering healthily recommendations in food recommender systems. In this approach, recommendations are designed to align closely with the user's preferences while subtly shifting towards healthier options under specific contexts, users are also granted the autonomy and flexibility to decide the extent to which they wish to adhere to their personal health and nutritional expectations.

1.5 Structure of the Thesis

This thesis is organised into seven chapter, as follows:

- **Chapter one:** Provides background information and outlines the research motivation, current challenges and issues within the field, the research aim and objectives, the importance of the study, and the thesis structure.

- **Chapter two:** This chapter begins with an introduction to the definition and history of recommender systems, along with a discussion of the main challenges facing the recommender systems community. Following this, key research on recommender systems across various application areas is discussed, with a focus that narrows to the food recommendation domain. The chapter then introduces the development of context-aware recommender systems, along with a detailed discussion on designing healthy food recommender systems. Finally, evaluation methods for recommender systems, as well as identified research gaps and limitations, are presented.

- **Chapter three:** This chapter establishes the theoretical framework and research approach, starting with a discussion of relevant philosophical assumptions and research methodology. It then introduces the overall research method, followed by a detailed explanation of why semi-structured interviews, an experimental study, contextual prefiltering, and contextual modelling approaches were chosen as appropriate methods for this research. Finally, ethical considerations for the study are presented.

- **Chapter four:** This chapter presents the findings from the first stage of qualitative research, which used semi-structured interviews to explore how contextual factors influence people's food choices and nutritional intake behaviours.

- **Chapter five:** This chapter presents the findings of an experimental study that examines the relationship between individuals' recipe rating behaviours and their implied nutritional intake across seven simulated contextual scenarios and a context-free control group. The analysis identifies the most influential contextual factors impacting these behaviours and demonstrates how they benefit rating prediction model performance.

- **Chapter six:** This chapter demonstrates the proposed novel contextual modelling approach, in combination with the novel contextual healthy recommendation method. The optimal feature combination for achieving the best model performance is identified and discussed.

- **Chapter seven:** This chapter concludes the thesis, providing a summary of the research findings on each stage, highlighting the contributions made, discussing potential directions for future work, and reflecting on the limitations of the study with suggestions for improvement.

Chapter 2

Literature Review

This chapter provides an overview of the current state-of-the-art algorithms for RS development, with a particular focus on their implementation in the FRS domain. The research gaps in integrating contextual knowledge, as well as health and nutritional information, have been identified. The potential influential contextual factors influencing individuals' food choices and dietary patterns have been reviewed from anthropological, ecological, and psychological perspectives. The chapter begins with a brief overview of the RS field, including definitions, a concise history, current challenges, and popular application domains in Section 2.1. A review of the predominant RS techniques and their implementation in FRS is presented in Section 2.2, the discussion will begin with an overview of state-of-the-art algorithms and then narrow down to the area of food recommendations. Furthermore, the importance of integrating contextual knowledge in RS, particularly in FRS, along with typical approaches and challenges of implementing CARS is highlighted in Section 2.3. In the meantime, how food choices and dietary patterns are influenced and driven by anthropological, ecological, and psychological aspects is discussed in Section 2.4. The trade-off between user preferences and healthiness in FRS is emphasized in Section 2.5. The chapter concludes with a discussion on RS model evaluation in Section 2.6 and a summary of the research gaps identified in Section 2.7.

2.1 Recommender Systems

2.1.1 Definition and History of Recommender Systems

In recent decades, the exponential growth of information, particularly through the Web and e-commerce platforms, has generated an overwhelming volume of readily accessible data. As a result, recommender systems have become essential tools, widely adopted for their effectiveness in addressing information overload. These systems provide valuable insights aligned with user preferences and are capable of reducing cognitive load by simplifying the decision-making process. By presenting relevant and personalised suggestions, they minimize the unnecessary mental effort required to navigate large amounts of information, ultimately enhancing user satisfaction during browsing, searching and overall engagement (Resnick et al., 1994; Resnick & Varian, 1997; Aggarwal et al., 2016). Burke (2007) defined the recommender system as: "a personalised information agent, which provides suggestions and recommendations for items that could be liked or used by a user". The basic concept of an RS was established through cognitive science research (Rich, 1979), information retrieval (Van Rijsbergen, 1979) and artificial intelligence (Ricci & Werthner, 2006; Sharma & Gera, 2013a).

It can be argued that the earliest known RS application was called 'Grundy', a computer-based librarian that considered users' preferences to recommend books through an interactive interview process. the system would classify users into stereotype groups and recommend the same books to all users within the same group, using a relatively simple method (Rich, 1979). Goldberg et al. (1992) introduced a similar system called Tapestry, which focused more on user interactions and explicit feedback to provide recommendations. It allowed users to mark and comment on the articles they were reading, influencing further recommendations. From 1994 onwards, RS began to emerge as a distinct research direction, when the GroupLens research group launched a News RS (Resnick et al., 1994). The major contribution of their work was the introduction of collaborative filtering as a method for completing the recommendation task, along with the establishment of an open architecture for solving the recommendation problem. This innovation spurred the development of other systems and brought real-world challenges into the research sphere, leading to the creation of systems such as Firefly for music recommendations and the BellCore movie recommender (Resnick et al., 1994; Jannach et al., 2010). Among the most significant early commercial applications of RS was Amazon's item-based collaborative filtering algorithm, introduced in the late 1990s. Unlike traditional user-based collaborative filtering, Amazon's approach focused on item-item similarity, enabling the system to efficiently recommend products based on the relationships between items rather than user behavior. This method allowed for relevant recommendations even for users with limited purchase histories. Amazon's pioneering system scaled RS to unprecedented levels, serving millions of users and processing millions of products (Linden et al., 2001).

CBF and CF approaches were developed around the same time, but CF initially gained more attention due to its ability to leverage user behaviour for recommendations. Developing a well-functioning content-based recommender system is particularly challenging, as it requires detailed domain knowledge to understand the characteristics of the items and determine the factors that influence user preferences (Ricci et al., 2015; Jannach et al., 2010). Arguably, one of the first well-known content-based systems was the Music Genome Project, developed in 1999. This system analysed over 450 musical attributes, allowing it to recommend music with similar characteristics to the user's preferences (Joyce, 2006). However, this technique relied heavily on large amounts of user feedback to effectively identify individual preferences from the outset. Additionally, it struggled to extrapolate users' tastes beyond their prior experiences, as it focused solely on the properties of the items themselves. To address the limitations of both collaborative filtering and content-based approaches, a hybrid recommender system named Fab was developed by a Stanford University student (Balabanović & Shoham, 1997). By combining these two approaches, Fab aimed to solve the cold start problem by using content-based filtering to generate initial recommendations until enough user data was collected to apply collaborative filtering. This approach was influential in demonstrating how a combination of both methods could enhance recommendation accuracy and user satisfaction, and its underlying principles remain foundational in modern recommender systems.

Much algorithmic research in this space was driven by the Netflix Prize in 2006, a landmark competition that aimed to improve the accuracy of Netflix's movie recommendation algorithm (Bennett et al., 2007). After three years of competition, in 2009, a team called BellKor's Pragmatic Chaos won the prize by achieving a 10.06% improvement over Netflix's existing Cinematch algorithm. The winning solution was a complex ensemble matrix factorisation-based model (Koren et al., 2009), which, compared with classic nearest neighbour techniques offered superior incorporation of additional information, such as implicit feedback, temporal effects, and confidence levels. The Netflix Prize fostered cross-disciplinary collaboration, attracting experts from fields such as machine learning, statistics, data science, and artificial intelligence. It played a pivotal role in advancing recommender systems and machine learning by providing a large real-world dataset and emphasising the importance of data-driven innovation. One of the most significant contributions of the Netflix Prize was the development and widespread adoption of matrix factorisation techniques, particularly Singular Value Decomposition (SVD) (Funk, 2006), SVD++ (Koren, 2008) and related methods. The impact of these techniques on recommender system algorithms continues to shape the field today (Amatriain & Basilico, 2015). In the meantime, the success of ensemble models and matrix factorisation methods inspired further developments in recommender systems, including the incorporation of contextual data in later years.

In 2007, Joseph A. Konstan hosted the first Association for Computing Machinery (ACM) Recommender Systems Conference, where he highlighted the growing interest in making recommendations based on various contexts. He suggested that research should focus on understanding how users interact with different activities and situations, and how RS could better support these interactions (Jannach et al., 2010). This emphasised that context-awareness would be an inevitable development for the next generation of RS (Khusro et al., 2016). The rise of mobile computing, increased access to contextual data, and advances in machine learning have since made context-aware recommender systems both feasible and scalable. In the 2010s, Spotify became an early adopter of CARS, incorporating contextual factors such as workouts and time of day (e.g., morning commutes) to offer more personalised playlists. These features significantly improved the user experience, making music discovery feel more natural and intuitive, ultimately leading to higher user retention and loyalty. The future of CARS is promising, with advancements in real-time health insights and emotion detection poised to enhance personalised digital interactions, paving the way for the next generation of intelligent and adaptive recommender systems (Adomavicius & Tuzhilin, 2011).

2.1.2 Overview of Current Recommender Systems Research and Identification of Challenges

Current research in the field of RS has become increasingly diverse, spanning several key application domains. In terms of typical techniques, Felfernig et al. (2013) identified five commonly-used recommendation approaches: CF, CBF, knowledge-based recommendations (KBR), group recommendations (GR). CF remains one of the most widely-used techniques, with matrix factorization (MF) approaches like Singular Value Decomposition (SVD) being particularly popular due to their effectiveness in handling large datasets and sparse user-item interactions. Content-based filtering also plays a crucial role, especially in domains where item attributes (e.g., movies, books, or music) can be leveraged to provide personalised recommendations based on user preferences. Ontology and knowledge based systems, where ontologies can represent a subset of domain knowledge or be added as an additional semantic layer to improve the relevance of recommendations (Vijayakumar et al., 2019; Werner et al., 2013; Carrer-Neto et al., 2012). This approach is particularly popular in e-learning applications, where domain-specific knowledge structures play a crucial role in providing accurate recommendations (Zhuhadar et al., 2009; Tarus et al., 2018).

Several emerging technologies have been integrated into the RS domain, offering innovative approaches to enhancing recommendation accuracy and personalisation. Techniques such as Machine Learning (ML) and Data Mining (DM) enable the incorporation of more nuanced user behaviour data and contextual information, shaping research directions toward context-aware systems and resulting in more personalised recommendations. Lee et al. (2017) proposed a music streaming RS based on human activity detection on machine learning techniques such as Naive Bayes (NB), Support Vector Machine (SVM), Random Forest (RF). Li et al. (2016) employed a series of data mining and social computing models to analyse user activities within social network patterns, enhancing user rating prediction performance in the television and movie recommendation domain. In recent years, Deep learning-based approaches are also increasingly being adopted to model complex user-item interactions. Techniques such as Multilayer Perceptron (MLP) perform well in learning nonlinear transformations and hierarchical feature representations. Convolutional Neural Networks (CNNs) excel at processing data with grid-like topologies, while Recurrent Neural Networks (RNNs) are particularly suited for modeling sequential behaviors, making them valuable in capturing user activity over time (Ouhbi et al., 2018; Zhang et al., 2019; Kiran et al., 2020). Another area of growth is preference-based and active learning techniques, where user feedback is continuously integrated to refine recommendations in real-time. These methods allow systems to dynamically adapt to changing user preferences, leading to more engaging and interactive recommendations (Ono et al., 2009; Rubens et al., 2015; Qomariyah, 2018; Yakhchi, 2021).

Despite the advances in the RS field, significant challenges such as data sparsity, cold start problems, scalability, context awareness, diversity and explainability continue to pose obstacles for the RS community, as will be discussed in detail below.

Challenges:

- The Cold Start Problem: As the most commonly used RS algorithms are based on a user's historical behaviour, or how the user community has interacted with a given item. When a new user or a new item enters the system, there is an information blank which triggers the cold start problem. There are three typical cold start problems: new user problem, new item problem and new system problem. CF approaches suffer from both new user and new item problems as they provide recommendations based on users' past ratings (Park et al., 2012; Khusro et al., 2016). Content-based algorithms generate recommendations based on item attributes or features, which can help mitigate the cold-start problem to some extent. However, these recommendations rely heavily on comprehensive domain knowledge datasets, which can be challenging to acquire and may result in insufficient effectiveness. Cross-domain recommender systems present a promising solution to the cold-start issue by leveraging data from multiple sources to provide recommendations. Nevertheless, challenges such as data alignment and domain relevance must be carefully managed to ensure the effectiveness of these systems (Zang et al., 2022). The cold-start problem remains one of the most significant challenges in the field of recommender systems, necessitating further in-depth research.

- Sparsity: Sparsity is another major challenges facing the RS field, and can often be treated as a derivative of the cold start problem (Sharma & Mann, 2013; Sharma & Gera, 2013b). Sparsity issues occur when there is insufficient data concerning user preferences and/or item characteristics. Most users interact with only a small subset of available items, leading to a situation where the combination of user and item pairs u,i often lacks existing ratings. This results in a the sparse user-item interaction matrix, complicating the identification of meaningful patterns and similarities within the data. Particularly, in real life datasets such as MovieLens 1M, a widely used benchmark dataset in the area of movie recommendations, the sparsity rate is 95.5%, meaning more than 19 out of 20 cells in the user-item matrix are empty (Harper & Konstan, 2015). In others, like the Book-Crossing dataset, the sparsity rate is in excess of 99.99% (Kiran et al., 2020). Although, matrix factorisation techniques, such as, Singular Value Decomposition (SVD) (Sarwar et al., 2002; Zhou et al., 2015; Guan et al., 2017) and SVD++ (Jia et al., 2014; Cao et al., 2015) are able to reduce dimensionality and uncover latent factors even in sparse data, they often function as black boxes, resulting in lower transparency and explainability. While these methods address the sparsity problem to some extent, the lack of interpretability can hinder user trust and understanding of the recommendations generated.

- Scalability: Scalability refers to the ability of a system to handle large and continually growing data. With the information explosion online, the volume of data has grown extremely fast, making it challenging for RS to provide accurate and speedy results. In particular, for the CF approach, the largescale data will certainly cause a burden to computations (Sharma & Gera, 2013a). In dynamic environments, user preferences and item availability can change rapidly. Maintaining scalability while ensuring real-time recommendation generation presents considerable challenges. Methods proposed for facing this challenge are based on approximation mechanisms; even though speed would improve, most of the time there will be an accuracy degradation (Papagelis et al., 2005). In addition, incorporating contextual information into recommendations adds complexity and further presents scalability challenges (Adomavicius & Tuzhilin, 2011).

- Contextual Awareness: Context-aware recommender systems integrate various categories of user-related information that can be helpful in generating high-quality recommendations, such as current location, movement patterns, user activities, and time of day. From an operational standpoint, context awareness is an inevitable development for the next generation of recommender systems, as it allows for a dynamic understanding of user behaviour and enables the learning of preferences in both long-term and short-term scenarios (Khusro et al., 2016). However, incorporating contextual information into recommender systems introduces notable complexities, particularly due to the inherent increase in dimensionality and data sparsity. Furthermore, obtaining comprehensive contextual information is difficult, as real-life situations are hard to replicate under experimental conditions. Identifying useful contextual data and determining the appropriate weights to balance personalisation with the influence of contextual factors poses another critical challenge (Zheng et al., 2013; Felfernig et al., 2013). After all, overemphasising contextual information may lead to less personalised recommendations, undermining user satisfaction.

- Diversity: Once the RS starts to generate recommendations for users' personal expectations, the results can be easily narrowed by users' interest(s). For example, if a user shows interest in basketball, then only basketball matches will be recommended. This is where diversification is needed. Otherwise, based on the explicit feedback that the user provides, the RS algorithm would tend towards being more specific and focused on the user's preference, but unable to detect other types of interest since there are not any interests have been shown to the system (Lu et al., 2015). The graphtheoretic approach (Adomavicius & Kwon, 2011) and a hybrid ranking method called Total Diversity Effect Ranking (TDER) (Premchaiswadi et al., 2013) has been put forward as a proposed solution to this challenge. Therefore, future research could focus on the evaluation measure that elicits the average user's perception of diversity (Kunaver & Požrl, 2017).
- Explainability: Explainability is a hot topic in most recent research. It has been proven that integrating the explainable results will increase the user trust and effectiveness of the system (Wang et al., 2021). However, determining the optimal approach for adding explanations, whether through a supervised attention mechanism, a conversational agent, or an automatic recommendation based on a complex reinforcement learning strategy, remains a challenging and intricate topic (Tintarev & Masthoff, 2015).

This research prioritises addressing the challenge of context awareness in food recommender systems, an area that has not been thoroughly explored. Evidence supporting this focus will be summarised in the following sections. In developing context-aware food recommender models, this research also tackles issues of sparsity, scalability and explainability to some degree, as the proposed approaches are able to emphasizes and reveals the importance of model features, with the vision of enhancing model transparency and improving user understanding. Additionally, the proposed single-stage approach may offer potential solutions to mitigate the cold-start problem. Furthermore, the study aims to provide relatively diverse recommendation results by introducing a novel contextual healthy re-ranking approach.

2.1.3 Applications of Recommender Systems

According to literature reviews by Park et al. (2012) and Lü et al. (2012), popular application areas for RS include movie and music recommendations, television programs, books, documents, websites, conferences, tourism, and learning materials. These domains encompass a range of sectors, including e-commerce, e-learning, e-libraries, e-government, and e-business services. Felfernig et al. (2013) provide a novel classification of new and emerging applications of recommender systems, including Software Engineering, Data and Knowledge Engineering, Knowledge-Based Configuration, Persuasive Technologies, Smart Homes, Help Services, and Innovation Management. Their classification takes user preferences into account and categorizes application domains from the standpoint of personal assistants. Additionally, their study highlights several areas for future research in RS, including an emphasis on the user perspective, sharing recommendation knowledge, enhancing context awareness, and exploring psychological aspects.

Different recommendation scenarios present distinct challenges, as the factors influencing people's decision-making and preferences can vary widely. For instance, individuals may choose to listen to certain songs due to appealing melodies, lyrics, or soundtracks (Schedl, 2019; Shakirova, 2017). In contrast, motivations for watching movies might be tied to specific actors, themes, or regional characteristics (Christakou et al., 2007; Katarya & Verma, 2017). The choice of a particular restaurant may depend on factors such as the location, the time of day, and prior reviews (Ramirez-Garcia & García-Valdez, 2014). Rather than relying solely on the unexplainable hidden features of traditional recommendation algorithms, integrating various factors into recommender systems may yield significant benefits in uncovering user preference patterns and understanding the reasons behind these decisions. Table 2.1 below summarises recent research across different application domains, focusing on the features and algorithmic techniques employed. It highlights that although CF and CBF remain dominant algorithms, context-aware and deep learning techniques have garnered increasing attention in nearly all important application domains, supporting the findings of Felfernig et al. (2013).

Table 2.1: *Applications of recommender systems*

RS fields	Recommendation techniques	Features Considered	Reference
Books	Content-based	User, item, temporal	(Rana & Jain, 2012)
	Hybrid-based	Ontology for user profiling	(Chandak, Girase & Mukhopadhyay, 2015)
	FUCL (improved association-rule mining)	User profile, book category and book loan	(Jomsri, 2018)
Movie	Collaborative filtering	Genre, Actor, Director, Year, Rating	(Kumar et al., 2015)
	k-means clustering adopting cuckoo search optimization algorithm	Movielens dataset	(Katarya & Verma, 2017)
	Hybrid deep autoencoder network	Social characteristics and behaviors on Twitter	(Tahmasebi, Ravanmehr & Mohamadrezae, 2021).
	Several unsupervised machine-learning algorithm	Movie genre and tags	(Cintia Ganesha Putri, Leu & Seda, 2020)
Music	Collaborative filtering	Similarity of user and item	(Shakirova, 2017)
	K-meansH	Electroen- cephalography (EEG) feedback	(Chang, Huang & Hui, 2017)
	Context-aware	UGP (user genre profile) and UCP (user content profile)	(Takama, Zhang & Shibata, 2021)
	Contextual post-filtering approach	Spotify Audio feature, activity (time of day and mood.	(Gong, Kaya & Tintarev, 2020)
	Emotion-aware	Users' keystrokes and mouse clicks patterns	(Yousefian Jazi, Kaedi & Fatemi, 2021)
Video	Context-aware	Age, daytime and weekday.	(Abbas et al., 2017)
	Hidden Markov model (HMM)	Sports, Round, Nationality and Athletes	(Hasan, Jha & Liu, 2018)
News	Collaborative filtering and SVM	Old user, new user and news stories	(Fortuna, Fortuna & Mladenić, 2010)
	Biased matrix factorization model	Temporal dynamics of user preferences and news taxonomy	(Raza & Ding, 2019)
Food	Pre-filtering	Nutrition information and user preference	(Toledo, Alzahrani & Martinez, 2019)
	Knowledge-aware	Demographics, affect, domain knowledge, behaviour data and health data	(Musto, Trattner & Starke, 2020)
	Health-aware	Namely, recipe retrieval, user health profiling	(Wang et al., 2020)
	Random Forest, Logistic Regression and Naive Bayes	Title, image, ingredients, popularity and nutrition	(Elsweiler, Trattner & Harvey, 2017)
	Content-based, collaborative filtering and hybrid-based	Recipe and food item	(Freyne & Berkovsky, 2010)

Continued on next page

RS fields	Recommendation techniques	Features Considered	Reference
	Content-based, collaborative filtering and hybrid-based	Various contextual factors (including but not limited to main meal, breakfast, following days)	(Harvey, Ludwig & Elsweiler, 2013)
Drugs	Collaborative filtering (Neighbourhood-based method and the Restricted Boltzmann machine-based method)	Number of approved drugs, number of side effect terms, and the number of side effects of the approved drugs	(Zhang et al., 2016)
	Naive Bayes and Core NLP	Heart rate, blood pressure, fever, cold, headache etc	(Gujar et al., 2018)
Tourism	Rank-GeoFM	Weather-related features such as temperature, cloud cover, humidity and pre- cipitation intensity.	(Trattner et al., 2018)

2.2 Recommender System Techniques

There are three key approaches in the RS area: content-based filtering, collaborative filtering, and hybrid (a combination of the prior two). As illustrated in 2.1 below, each approach has a different implementation and mechanism of operation. The content-based filtering approach recommends items based on their inherent attributes, typically utilising two principal techniques: statistical analysis (e.g., t-tests, cosine similarity) and machine learning algorithms. These techniques facilitate the identification of items that are analogous to those previously favoured by a user, thereby enhancing the precision of recommendations. Conversely, collaborative filtering relies on prior user interactions and behavioural data, recommending items based on the preferences exhibited by similar users or the historical interactions between users and items. Collaborative filtering can be further categorised into three primary types: user-based (which identifies similar users), item-based (which finds analogous items), and model-based approaches. Hybrid systems integrate both content-based and collaborative filtering methodologies to leverage the advantages of each (Ricci et al., 2015). This chapter is dedicated to introducing and critically discussing the current research that employs these three techniques.

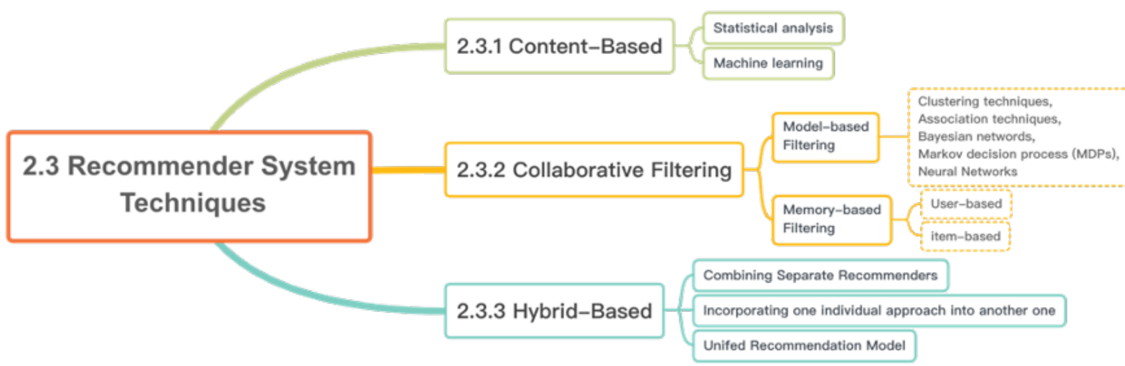


Figure 2.1: *Summary of recommender system techniques*

2.2.1 Content-Based Recommender Systems

A **content-based RS** emphasises the analysis of the attribute characteristics of items, generating predictions based on this information. In this technique, recommendations are derived from features extracted from item content, informed by the user's past preferences, such as purchase history, browsing behaviour, and favourite tags. The system recommends potential items that are most similar to those rated positively by the user. The workflow of a content-based recommender system is illustrated in Figure 2.2. For instance, if a user purchases spaghetti, bread, and cake, while also expressing interest in cheese, the algorithm is capable of identifying the ingredients of these purchased and interested items. It then matches them with items that share similar ingredients, potentially leading to a recommendation for pizza. Due to its algorithmic mechanism, a content-based approach is particularly effective for recommending textual items, such as web pages, publications, or news articles, as these items are easily "understandable" by computers (Isinkaye et al., 2015; Herlocker et al., 2000).

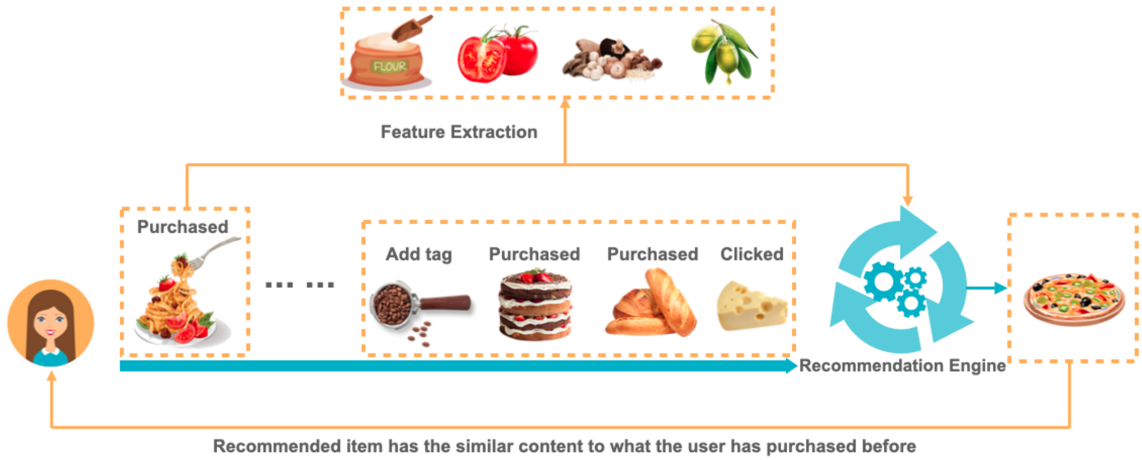


Figure 2.2: *General framework of Content-based recommender system*

2.2.1.1 Content-Based Approaches

There are mainly two different ways of generating similarity between items and providing meaningful recommendations. One is through statistical analysis, for example, Vector Space Model such as TF-IDF (Ghauth & Abdullah, 2009; Musto, 2010; Di Noia et al., 2012; Musto et al., 2016). The second is through ML probabilistic models such as Naive Bayes Classifier (Nidhi & Annappa, 2017), Decision Trees (Li & Yamada, 2004) or Neural Networks (Van den Oord et al., 2013; Musto et al., 2018) to model the relationship between different documents within a corpus. CBF does not need to link other users' profiles to the target user because they will not influence the recommendation result. Even if the user profile changed, the model still has the potential ability to adjust the recommendation within a short period.

Various research studies have attempted to employ the CBF approach to address domain challenges. During the past 15 years, most content-based systems have been conceived as text classifiers and keyword-based representation based on the evidence of user interests. Although accurate recommendations are achieved, sometimes there is the 'lack of intelligence'. Therefore, ontologies-based semantic analysis comes to play as the external knowledge bases are allowed to use and build more accurate user profiles (Lops et al., 2011; Magnini & Strapparava, 2001). ITem Recommender (ITR) proposed by Degemmis et al. (2007) has become a classic example as it is capable of providing recommendations through several domains (e.g. music, movies, books). Word Domain Disambiguation (Magnini & Strapparava, 2000), Word Sense Disambiguation (Semeraro et al., 2009), and WordNet-based (Degemmis et al., 2007) are some key techniques that have been widely used in this approach.

With the increased availability of different types of objectives in Heterogeneous Information Networks (HIN), the meta-path-based algorithms have been inspired and used as a powerful way to improve cold start problems (Sun et al., 2011; Yu et al., 2014; Kula, 2015; Vasile et al., 2016). Shi et al. (2015) suggested a weighted HIN and weighted meta path algorithms for movie recommendations. They also tested their semantic path-based personalised recommendation method with the open public dataset Douban and Yelp. This produced satisfactory results demonstrating the usefulness of their methods in many domains, e.g. movie, business, customers.

In recent years, Deep Learning (DL) approaches have provided a more flexible framework to explore existing structures in item and user data, to model temporal and sequential aspects, and even generate rich explanations from user reviews (Musto et al., 2018; Deldjoo et al., 2018; Hidasi et al., 2015; Song et al., 2016; Lu et al., 2018). Zhang et al. (2017) designed a top-N recommendation system based on a joint representation learning (JRL), which includes heterogeneous information sources such as textual and multimedia features, external knowledge and user context and interaction data. Seo et al. (2017)'s inspiring research improved the interpretability of deep learning models and achieved transparent and explainable recommendations.

However, as content-based approach requires in-depth understanding and knowledge of the recommended items in the profile, its domain-specific nature may present significant limitations. These algorithms often struggle to break beyond filter bubbles and lead to overspecialisation (Pariser, 2011; Isinkaye et al., 2015). As such algorithms selectively predict the information or items a user likely to prefer based on their historical behaviour, this narrow focus leads to a reduction in the diversity of content presented. Consequently, users are frequently exposed to ideas, opinions, and information that reinforce their existing beliefs, potentially isolating them from contrasting viewpoints (Zhang & Iyengar, 2002; Min & Han, 2005). This challenge is particularly pronounced in the realms of news and book recommendations. The propensity for users to remain within their comfort zones exacerbates the challenge of mitigating the widespread dissemination of misinformation. By limiting engagement with contrasting viewpoints, individuals may become more susceptible to erroneous information, thereby undermining the integrity of public discourse (Diaz, 2008; Fernández et al., 2021). A similar issue occurs in the food domain. For instance, if a user prefers fast food, such as pizza or high-calorie items, it becomes difficult for the algorithm to suggest healthier alternatives based on the user's preferred recipe characteristics. This can further reinforce unhealthy eating patterns and habits (Min et al., 2019a).

2.2.1.2 Content-Based Approaches in Food Recommendation

CBF approaches are particularly popular for understanding users' individual preferences and tailoring recommendations accordingly. Freyne & Berkovsky (2010) creatively broke-up recipe ratings into ratings for individual ingredients and score them according to the ingredients contained times in the positively rated recipes. More specifically, if potatoes are contained in a recipe that the user has reported as liking, then the rest of the recipes containing potatoes would be considered as also liked by the certain user. Their results show that the CBF approach achieved improved MAE performance compared to the CF approach. Later work by Harvey et al. (2013) has extended this research. The study adopted a variation of Term Frequency - Inverse Document Frequency (TF-IDF) weighting while assessing the similarity between users and recipes. More importantly, their method not only considered positive ingredient biases but also integrated negative feedback from users. In this context, ingredients present in recipes can be interpreted as items users dislike. This research is particularly insightful because many existing studies on RS tend to focus predominantly on positive feedback, often neglecting the significance of user discontent (Min et al., 2019b).

Such TF-IDF and Vector Space-based approaches have exerted considerable influence and have been widely employed in FRS. Maheshwari & Chourey (2019) employed TF-IDF and Word2Vec models to vectorize recipe ingredients, subsequently identifying similar ingredients using cosine similarity. Building on this work, Chhipa et al. (2022) proposed a content-based recommendation model that also utilised TF-IDF and cosine similarity to suggest Indian recipes based on available ingredients. Their system further enhances user experience by allowing users to filter recipes according to dietary preferences.

Notably, Teng et al. (2012) proposed complement and substitute ingredient networks for recipe recommendations. The complement network captures the ingredients often occurring in the same recipes, while the substitute network is derived from user-generated suggestions for modifications. The experiment results showed that the use of these ingredient networks can significantly improve the performance compared to individual ingredient lists features. Lin et al. (2014) proposed a content-driven recipe RS incorporating 6 different types of content information with novel time-dependent features. Their model represents a significant improvement over the existing MF, CBMF (Forbes & Zhu, 2011) methods. Most recently, Vairale & Shukla (2021) proposed a content-based healthy food recommendation framework for thyroid patients. Their model is based on domain knowledge and is capable of analysing unique food characteristics and generating diet recommendation lists.

As food decisions are often visually driven, various image processing approaches have been applied in food recommendations (Milosavljevic et al., 2012; Yang et al., 2017; Elsweiler et al., 2017; Zhang et al., 2020a). Yang et al. (2017) designed FoodDist using Convolutional Neural Networks (CNN) based multitask learning, which provides a powerful improvement at the model generalisation level. Their research further demonstrates that an online learning algorithm is capable of efficiently learning food preferences. Elsweiler et al. (2017) provided evidence that automatically extracted low-level image features—such as brightness, colourfulness, and sharpness—can be valuable for predicting user food preferences. However, Zhang et al. (2020a) found that these low-level features do not significantly enhance the performance of deep neural networks. This indicates that while basic image characteristics may provide some insight into user preferences, more advanced features and representations are often necessary for effective modelling in FRS.

2.2.2 Collaborative Filtering Recommender Systems

Instead of relying on content information, the CF approach utilises ratings for items provided by a collection of users. The fundamental principle behind CF is to identify similar users or items based on user interactions and their historical preferences, then recommending items that the target user may potentially enjoy. (Desrosiers & Karypis, 2010; Elahi et al., 2016). To illustrate the collaborative filtering approach, consider a scenario in food shopping once more. Suppose a target user has purchased spaghetti, cake, and bread. There is another user who has also purchased spaghetti and cake, indicating a similarity between the two users. If this similar user also buys pizza, there is a likelihood that the target user may also be interested in pizza. Conversely, items preferred by different users—such as tea, oranges, and bananas—would not be recommended to the target user, as their purchase history does not align with those items. Figure 2.3 presents the detailed workflow of the CF approach, illustrating how user similarities and interactions inform item recommendations.

CF overcomes several limitations of CBF approaches. For example, when explicit features of items are difficult to obtain, perhaps due to the complexity of items or due to ethical, or privacy concerns. CF relying on user interaction data, both explicit and implicit (e.g., ratings, purchases, clicks, bookmarks). This approach enables recommendations to be generated without the need for detailed item characteristics, which can be particularly useful in scenarios where such information is scarce or sensitive (Adomavicius & Tuzhilin, 2005; Ricci et al., 2015). Moreover, CBF systems may inadvertently limit recommendations to items that very closely match a user's past preferences, potentially leading to a lack of diversity in suggestions. CF, on the other hand, can introduce a broader range of items by identifying similarities across different users, which can lead to more varied recommendations (Koren et al., 2021). However, since the CF approach heavily relies on user interaction data, it can suffer from the cold start problem. For new users, there is often insufficient interaction data to generate meaningful recommendations, while new items lack user ratings and feedback, making it difficult for the system to identify potential preferences (Adomavicius & Tuzhilin, 2005). The following discussion will critically examine the development of the CF approach and then focus on its implementation in food recommendation systems.

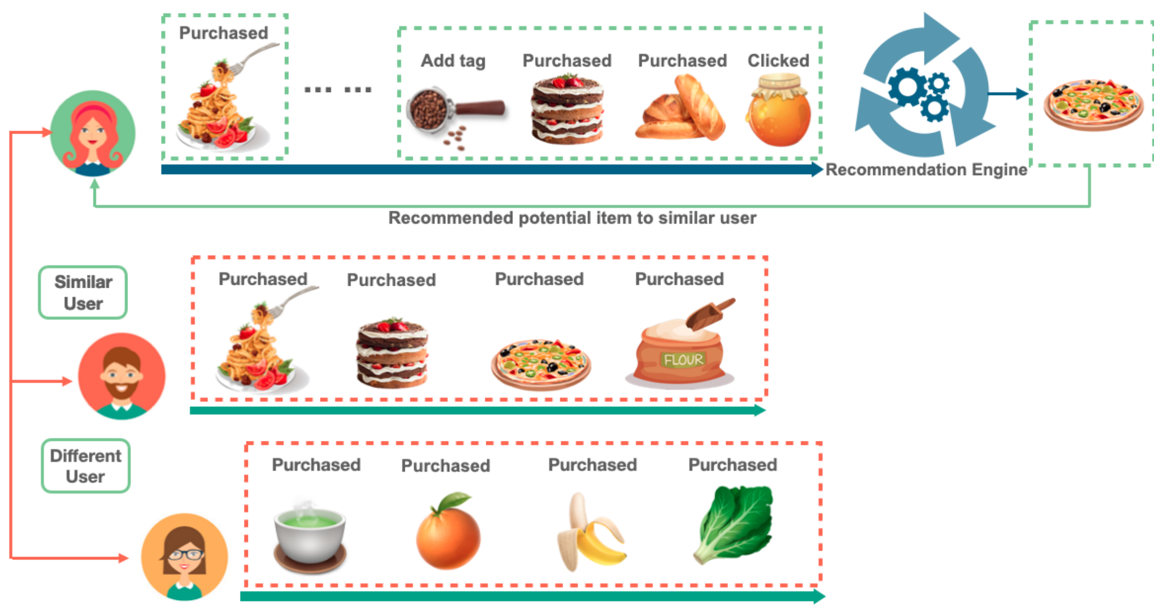


Figure 2.3: *General framework of collaborative filtering recommender system*

2.2.2.1 Neighbourhood-Based and Model-based Methods

Many famous collaborative systems have been developed both in academia and the industry. For example, GroupLens (Koren et al., 2021), Ringo (Shardanand & Maes, 1995), Amazon book RS, and PHOAKS a web source information RS (Terveen et al., 1997). According to Adomavicius & Tuzhilin (2005) and Breese et al. (2013), algorithms for a CF approach can be classified into two popular methods: neighbourhood-based and model-based methods. Neighbourhood-based methods (also called memory-based) are heuristic in nature, and are either user-based or item-based recommendation (Desrosiers & Karypis, 2010). User-based systems, such as GroupLens (Konstan et al., 1997) and Ringo (Shardanand & Maes, 1995), evaluate the user preference based on finding similar user, for example, when predict a users' preference for a certain item, it use this item's ratings from other users who have similar rating patterns or most associated to the given user's ratings (Elahi et al., 2016). Item-based approaches (Sarwar et al., 2001; Deshpande & Karypis, 2004), on the other hand, measure the similarity based on items, these approaches predict the user rating for a potential item based on the user ratings for similar items.

Therefore, what cannot be ignored here is finding the proper way to define the similarity. Various algorithms have been proposed to measure similarity between both users and items, arguably, the two most popular approaches are correlation and cosine similarity (Sarwar et al., 2001; Lee et al., 2007; Mu et al., 2019). These methods are widely recognised for their effectiveness in measuring relationships within user-item interactions, providing foundational frameworks that have led to further advancements in recommendation systems. Novel similarity measurement approaches, such as vector-based similarity calculations, statistical Bayesian methods and heuristic similarity measures have been further developed (Breese et al., 2013; Patra et al., 2015; Ahn, 2008).

Shang et al. (2010) later proposed a diffusion-based similarity measure for graph data. The Sparse Linear Method (SLIM) (Ning & Karypis, 2011) and the Factored Item Similarity Model (FISM) (Kabbur et al., 2013) have advanced item similarity learning through data-driven approaches. SLIM improves top-N recommendation accuracy by predicting scores for a new item based on an aggregation of other items' interactions. In contrast, FISM leverages the product of two low-dimensional latent factor matrices to model item similarities directly. Subsequently, a Global and Local SLIM (GLSLIM), introduced by Christakopoulou & Karypis (2016), extended SLIM to address its limitations in capturing users' preferences in different user group. GLSLIM enhances the top-N recommendations performance by providing a separate local item-item model for each user subset. Around the same time, Patra et al. (2015) found that the Bhattacharyya coefficient for neighbourhood-based CF outperforms existing measures even in sparse data. Recently, Mu et al. (2019) proposed the Common Person Correlation Coefficient (COPC) as a way to improve the similarity measures as well as introducing Hellinger Distance (Hg) as global similarity when lacking co-rated items.

Different from neighbourhood-based systems, which directly adopt the stored rating to compute the similarity and make predictions, model-based systems apply these ratings to learn a predictive model. Potential patterns and special characteristics of users and items will be learned by a set of model parameters from training data, and then the model will be used to predict new ratings. Based on a survey conducted by Su & Khoshgoftaar (2009) and Desrosiers & Karypis (2010), numerous ML techniques have been adopted in model-based CF and include Bayesian algorithm (Miyahara & Pazzani, 2000; Robles et al., 2003), Clustering (Ungar & Foster, 1998; King, 2015; Jiawei & Micheline, 2006; Su et al., 2005), Latent Semantic Model (Hofmann, 2001, 2004; Blei et al., 2003; Agarwal & Chen, 2010; Kim & Shim, 2014), Markov Decision Process (MDPs) (Shani et al., 2005), association rule-based (Neysiani et al., 2019), and dimensionality reduction techniques (Zarzour et al., 2018).

Miyahara & Pazzani (2000) proposed a simple Bayesian classifier to predict both negative and positive ratings from users. Their model has been tested on datasets for public movies and joke recommendations with the results indicating that the Bayesian approach significantly outperformed a correlation-based CF system. This was further improved by Robles et al. (2003) who proposed a parameters optimisation approach of Naive Bayes through calculating the confidence interval of its parameters, which is capable of finding the best classifier through heuristic search. This new approach outperforms the simple Naive Bayes conducted by Miyahara & Pazzani (2000). Clustering methods as a further solution of CF has the advantages of relative efficiency and easy implementation (Ungar & Foster, 1998). For instance, k-means, density-based methods, and hierarchical methods (Su et al., 2005; AL-Bakri & Hashim, 2019). A Latent Dirichlet Allocation (LDA) as a probabilistic model can be suitable for generating and understanding the text topic as it introduces latent class variables in a mixture model set to discover user preference and interest. Compared to standard memory-based methods, this technique has higher scalability (Hofmann, 2001; Blei et al., 2003; Hofmann, 2004). Agarwal & Chen (2010) proposed an upgraded LDA model named fLDA for solving the cold start of web-based recommendations. fLDA exact model fits through a Monte Carlo EM algorithm and does not rely on variational approximations. Kim & Shim (2014) generated a TWILITE system based on LDA model, which captured the user behaviour in Twitter and recommended Top-k most interesting tweets for each user. They tested their model with a real-life dataset and confirmed its effectiveness and accuracy. An interesting way of generating recommendations is through the MDP model, which treats recommendations as a sequential optimisation task instead of a prediction problem (Shani et al., 2005), paving the way for reinforcement learning RS models.

2.2.2.2 Collaborative Filtering in Food Recommendation

The implementation of CF approaches in the FRS domain has not been extensively researched compared to content-based filtering methods. Early work by Freyne & Berkovsky (2010) applied Pearson correlation within a nearest neighbour framework using a users' rating matrix. Their findings indicated that the CF model performed worse than the aforementioned CBF model, which subsequently inspired further development of CBF techniques in the FRS area. Later on, Ge et al. (2015) highlighted that their tagging and latent factor matrix factorisation model outperformed Freyne & Berkovsky (2010) proposed state-of-the-art content-based approach. They developed an Android-based FRS that allows users to add specific tags, such as "spicy" or "comforting," providing more precise references to understand user preferences. However, their study primarily focused on improving recommendation accuracy and did not incorporate nutritional or health information.

More recently, Trattner & Elsweiler (2017b) researched the nutritional information on the large online recipe dataset taken from popular US online recipe portal allrecipes.com. They tested 9 prominent recommender models using the LibRec3 framework, and the LDA and Weighted Matrix Factorisation (WRMF) approaches performed the best. Ornab et al. (2019) conducted an empirical study of collaborative filtering algorithms for meal set recommendation. They find out that model-based approach is more reliable than the memory-based approach when making meal set recommendations, which inspired this research to consider adopting a model-based approach. Fakhri et al. (2019) proposed a restaurant recommendation system for the Bandung Raya region of Indonesia that employs a user-based collaborative filtering approach. This system integrates both user rating similarity and user attribute similarity (e.g., gender, age) in its algorithm. Their findings demonstrate that the user-based CF method outperforms the user attribute-based approach in generating recommendations. However, their research treats the user-rating matrix and user attributes matrix separately, without investigating whether combining these two matrices would enhance model performance. Exploring the synergistic effects of user attributes and rating data might lead to significant advancements in recommendation system effectiveness. This concept aligns with the principles of context-aware recommendation systems, which aim to incorporate contextual information to improve the relevance of recommendations. Therefore, further in-depth research in this area is warranted (Adomavicius & Tuzhilin, 2011).

2.2.3 Hybrid-Based Recommender Systems

Hybrid RS overcome certain limitations of individual content-based and CF approaches by adopting the combination of multiple types of algorithms (Jain et al., 2015; Kouki et al., 2015). For instance, the CF approach struggles to recommend items without prior rating from the user, or new users who have not have interacted with the system, which commonly referred to new-user and new-item problems. In contrast, CBF approaches are not restricted by this limitation, as the prediction is generated based on the characteristics and features of the item themselves, enabling recommendations even in the absence of user ratings.

To enhance the performance of RS, knowledge-based systems are also crucial for knowledge extraction and acquisition, particularly as many artificial intelligence applications encounter well-documented bottlenecks (Adomavicius & Tuzhilin, 2005). Several strategies have been proposed to combine basic RS techniques, creating more robust hybrid predictions and recommendations (Adomavicius & Tuzhilin, 2005; Agarwal et al., 2011). For example, combining CF with demographic approaches can uniquely identify cross-genre recommendations, encouraging users to explore unfamiliar areas (Burke, 2007). The following section will review the current research on hybrid RS.

2.2.3.1 Hybrid-Based Approaches

Based on early research conducted by Burke (2007), there are seven different ways to manipulate basic RS (CF, CB, demographic-based, knowledge-based) and build hybrid RS include weighted, switching, mixed, feature combination, feature augmentation, cascade, and meta-level. Adomavicius & Tuzhilin (2005) summarised the building of hybrid RS into three ways: 1) combining separate recommenders, 2) incorporating one individual approach into another one), and 3) constructing a unified model that integrates both basic RS characteristics. The general workflow has shown in Figure 2.4 below.

- *I. Combining Separate Recommenders*

  In this approach, both CF and CBF model will be maintained initially, The outputs of each algorithm can then be combined into a single recommendation using either a linear combination or voting scheme. Alternatively, the individual recommenders can still be used imdependently at any given time, based on some recommendation "quality" metric (Claypool et al., 1999). For example, the DailyLearner system built by Billsus & Pazzani (2000) uses the Nearest Neighbour algorithm to model user's short-term interests, and a Naive Bayesian classifier to model user's long-term preferences. Subsequently, both

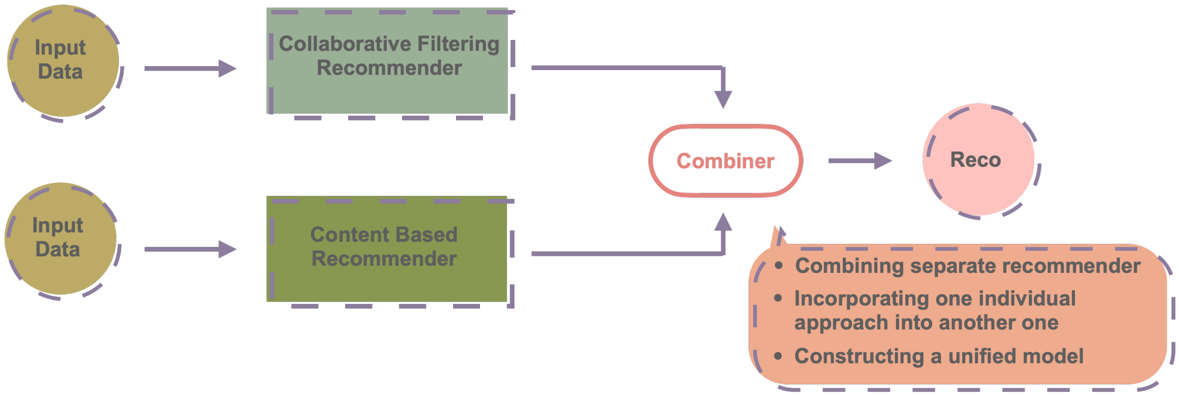


Figure 2.4: *General framework of hybird recommender system*

models are incorporated into a hybrid system that selects the RS that can recommend news with a higher level of confidence. Tran & Cohen (2000) proposed a hybrid system for electronic commerce that incorporates both CF and knowledge-based approaches, leading to a system capable of providing the recommendation most consistent with the user's past ratings and preferences.

- *II. Incorporating one individual approach into another one*

  Several hybrid approaches have been developed in this way. The Fab system, designed by Balabanović & Shoham (1997), is highly representative. It sought users with similar website preferences based on the content of their profiles. This approach helped overcome some of the sparsity-related issues present in individual CF systems (Pazzani, 1999). Similarly, Melville et al. (2002) proposed a Content-Boosted CF approach where the user's preferences and ratings were predicted through a content-based classifier. Recommendations were then made using a CF approach by computing the Pearson correlation coefficient to identify similar users. Their novel approach achieved an improvement of 9.2% and 4.0% in MAE compared to the pure CBF and pure CF approaches, respectively. While, Soboroff & Nicholas (1999) used Latent Semantic Indexing (LSI) to create a collaborative view of user profiles. With the profiles represented by term vectors, their results showed a better performance than the individual content-based approach.

- *III. Unified Recommendation Model*

  This approach has received increasing attention. Probabilistic Latent Semantic Analysis, a popular approach, has been proposed and employed to combine CF and CBF recommendations in several studies (Hofmann, 1999; Schein et al., 2002; Popescul et al., 2013). Another insightful approach, proposed by Condliff et al. (1999), used Bayesian mixed-effects regression models and Markov chain Monte Carlo methods for parameter estimation and prediction. Gunawardana & Meek (2009) later proposed Boltzmann machines (the probabilistic models) that first integrated both collaborative and content information as features, then learned the weights to reflect how well each feature predicts user actions. Yao et al. (2014) proposed a probabilistic generative model that considers both rating data and semantic content data to predict user web service preferences.

> Recently, Xiong et al. (2018) proposed a novel deep learning-based hybrid approach for Web service recommendations by combining CF and textual content. Moreover, several studies conducted an empirical comparison of the performance between a hybrid approach with individual CF and CBF algorithms. The results show that the hybrid approach could provide more robust and efficient recommendations (Bellogín et al., 2013; Koohi & Kiani, 2016; Xiong et al., 2018).

Although hybrid recommendations can avoid or remedy the weaknesses of individual recommendation techniques and improve overall performance, there is a potential risk of overfitting. Since such systems combine several techniques, they may become overly tailored to the training data, which may reduce the model's generalisation capability in new situations (Ricci et al., 2015). In addition, it could be challenging to pursue the optimal balance between different recommendation techniques in hybrid systems. Weighting the components incorrectly may result in one method dominating others, negating the potential benefits of using a hybrid approach (Burke, 2002).

2.2.3.2 Hybrid Recommender Systems in the Food Domain

Compared to other approaches, hybrid recommendations in the food domain have not been well researched. Sobecki et al. (2006) proposed a hybrid system and introduced fuzzy reasoning to recommend recipes. Freyne & Berkovsky (2010) broke the recipes down into individual ingredients and employed a unified hybrid approach to integrate three different recommendation approaches using a switching strategy. Another example is provided by Harvey et al. (2013) who conducted research to understand a user's recipe choice under different contextual environments and learn user preferences regarding both individual ingredients and a combination of ingredients in terms of their nutritional content. They compared their model with the state-of-the-art CF baselines method, and the results showed that their SVD model combined with weighted and biases significantly outperformed CF baseline. Recently, Chavan et al. (2021) introduced a weighted hybrid system that combines content-based and collaborative filtering approaches, The results showed that the hybrid model outperformed the pure CB and CF models in both recall and accuracy metrics. However, the study lacks detailed insights into how the healthy recommendations are generated or how the recipes are re-ranked. Instead, the researchers incorporated nutritional information as an additional feature during the model training and rating prediction process, without further elaboration on its impact on the recommendation process.

Based on the experimental results of existing research, hybrid algorithms have demonstrated promising performance, suggesting that hybrid-based recommendation systems in the food domain are worth exploring in greater depth. To further improve the performance of hybrid-based RS, it is essential to consider and incorporate contextual information. Doing so can make recommendations more diverse and better tailored to the user's current situation. A system that integrates contextual data would be capable of generating more personalised, intelligent, and situationally-relevant recommendations (Raza & Ding, 2019). The following section will provide a detailed discussion of the context-aware recommender systems approach.

2.3 Context-Aware Recommender Systems

Unlike CBF and CF approaches, which typically focus on two-dimensional user and item information, CARS additionally consider contextual information such as, location, time, mood, purchasing purpose. Each contextual factor can also consist of different elements; for example, time can be described in terms of seconds, minutes, days, or years. According to Adomavicius & Tuzhilin (2005), the CARS implement a three-dimensional model containing users, items and contextual information and comprising four different components: (1) user's preferences and interests (input), (2) current contextual information (input), (3) RS and (4) the set of recommendation results (output). Before developing a CARS, it is beneficial to understand the categories of context to ensure effective implementation. The following section will review the categories of contextual factors (Section 2.3.1) and the state-of-the-art algorithms for developing CARS (Section 2.3.2). Additionally, influential factors affecting people's food choices, identified from the field of psychology, will be reviewed and critically discussed (Section 2.4.3), with novel factors reflected upon to support the development of CARS in the food domain (Section2.3.3).

2.3.1 Categories of Context

As the context concept significantly evolved during the past few decades, several systems have been proposed, starting with the domain of tourism, leisure, and e-commerce because of the important impact of these industries on the world economy (TTCI, 2013). From RS technique perspective, it has been widely accepted that the diverse contextual information can be classified into three categories shown in Figure 2.5, including: physical context (e.g., temporal, spatial, environmental and equipment), personal context (e.g., demographic, social, Psychophysiological and cognitive) and technical context (e.g., hardware and data dimension) (Ferdousi et al., 2017; Colombo-Mendoza et al., 2015). Each contextual dimension also contains several levels, for example, temporal dimension refers to the time of the day (weekday, weekend, birthday, events), and spatial dimension may include GPS location or at work, at home normal classes. Among them, location and time as the most popular contextual features has been extensively researched in the domains of music, movies and tourist recommendation. Due to the difficulty of monitoring and collecting psychophysiological signals and cognitive data, these types of factors have rarely been incorporated into recommender systems, particularly in the food domain. However, such factors may exert a considerable influence on individuals' decision-making processes and fluctuating preferences, making them highly deserving of further in-depth investigation (Bavaresco et al., 2020; Tkalčič et al., 2013).

Depending on the system's knowledge about contextual factors such as what exactly the RS knows about what is being observed, the system can be classified into three categories: *fully observable*, *partially observable* and *unobservable*. In fully and partially observable situations, the model building is based on explicit user feedback. Otherwise, in unobservable situations the latent knowledge of contextual information is used to build a latent predictive model using such as hidden Markov models (Adomavicius & Kwon, 2011). Moreover, based on whether contextual factors change over time or not, they can be classified as *static* (the relevant contextual factors remain stable over time) and *dynamic* (when the contextual factors change over time). These two aspects combined with three knowledge classifications give rise to the

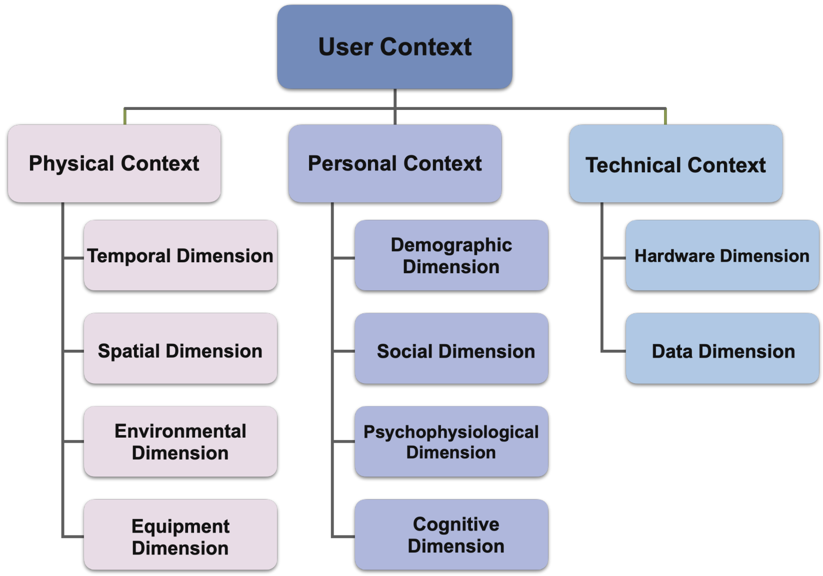


Figure 2.5: *Contextual features categorisation in CARS (Ferdousi et al., 2017)*

3*2 diagrams presented in Figure 2.6. Due to the frequent changes in dynamic factors over time, capturing and monitoring them often requires real-time data collection, such as GPS tracking, sensor data, or behavioural tracking, which can be technically complex. Additionally, since dynamic factors are constantly evolving, developing recommender systems that can adapt to these changes presents even greater challenges (Polignano et al., 2021).

2.3.2 Context-Aware Recommender Systems Algorithms

Adomavicius & Tuzhilin (2011) introduced three paradigms in CARS: *pre-filtering*, *post-filtering* and *contextual modelling*. In the contextual pre-filtering approaches, the contextual features "c" are used to filter and construct a relevant dataset, then the similarities can be computed as usual using traditional 2D algorithms. In contextual post-filtering approaches,

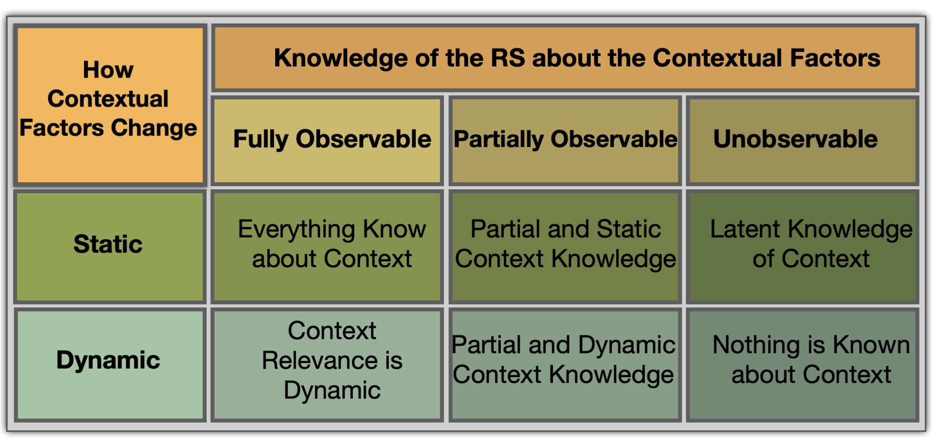


Figure 2.6: *Contextual information dimensions (Adomavicius & Tuzhilin, 2011)*

the context information is initially ignored, and the ratings are still predicted using a 2D recommender algorithm on the entire original dataset. Then the recommendation results are adjusted based on the contextual features. Finally, in contextual modelling approaches, the contextual information is directly incorporated into the similarities computation and rating estimation, requiring techniques capable of handling multi-dimensional models. The diagram of three approaches is shown in Figure 2.7 below.

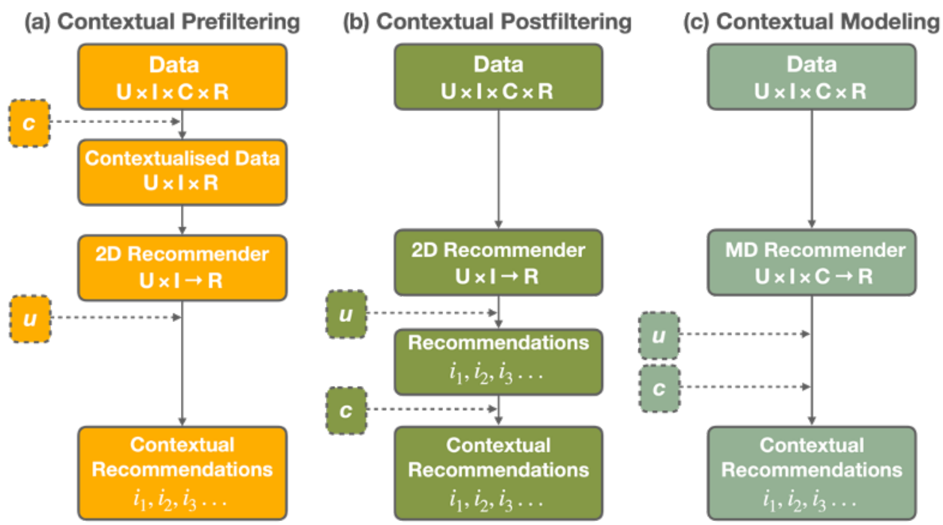


Figure 2.7: *Three paradigms for context-aware recommender system (Ricci et al., 2015)*

In the first CARS conference workshop, Lombardi et al. (2009) claimed the importance of incorporating context within the RS. They use the pre-filtering approach to evaluate the effectiveness of adding time contextual features to predict the users' online retail behaviour, and the performance represents a significant improvement compared with the baseline uncontextualised models. Interestingly, Baltrunas & Ricci (2009) came up with a somewhat different approach to contextual pre-filtering named item splitting, which split each item into several fictitious items based on different contextual variables to generate new items. In the same year, Baltrunas & Amatriain (2009) proposed a micro-profiling approach similar to the item splitting idea, but they focused on splitting the user profile into particular context instead of splitting the rated items compared with Baltrunas & Ricci (2009). Their proposed techniques work particularly well under the contextual pre-filtering paradigm, as it sufficiently reduces the multidimensional recommendation problem to a standard user*item (2D) matrix, which means the traditional recommendation techniques can be applied for rating prediction. Since then, the idea of generalised contextual pre-filtering has been adopted in various research studies. Insightfully, Zheng et al. (2013) proposed a novel approach for handling contextual features in collaborative filtering, called the differential context weighting (DCW) approach, which demonstrated better generalisation ability than their previous differential context relaxation (DCR) model. The DCW model was tested on two real-world datasets in the food and movie domains. Compared to baseline models (CF, pre-filtering, and DCR), DCW achieved the lowest RMSE, approximately 1.03 and 2.62, respectively. Codina et al. (2013) leveraged semantic similarities across different contextual situations to improve recommendations. Their proposed approach generates recommendations not only from a single contextual situation, but is also capable of predicting user ratings based on contextual situations that are semantically similar to the target context.

However, as pre-filtering approaches face challenges related to dataset splitting, the question of how to combine the predicted results generated from each profile to present a final prediction remains an area for future research (Baltrunas & Amatriain, 2009). One major advantage of the post-filtering method is it is easy to generalise, which allows any of the traditional recommendation techniques, such as CF and CBF to apply this framework effortlessly (Adomavicius & Tuzhilin, 2005). Panniello et al. (2009) conducted an inspiring and comprehensive study of comparing contextual pre-filtering method and two different post-filtering methods (Weight PoF and Filter PoF), as well as modelling and evaluating techniques under various conditions (e.g. find-all vs. top-k, accuracy vs. diversity). The weight PoF approach re-ranks the recommended items based on the rating probability of relevance in the given context, whereas the filter PoF approach filters out items that are not relevant in that context. Among their findings, the comparison results show that there is no 'universally' best approach under all circumstances. For example, the post-filtering may demonstrate better accuracy performance, but the pre-filtering approach often provides better diversity. Ramirez-Garcia & García-Valdez (2014) applied a post-filtering method to restaurant recommendations based on explicit user feedback of different locations and time. Their results show that incorporating the contextual information does not improve the performance significance. This is possibly due to the relatively small dataset, with only 50 users and 1,422 ratings of restaurant choice under different situations, and the unreasonable designed questionnaire questions may have imposed additional cognitive load on users when expressing their preference. In a recent study, Gong et al. (2020) developed a post-filtering re-ranking algorithm for music recommendations by incorporating physical activities (such as running, walking, and sleeping) and time-of-day features (morning, afternoon, evening) using Spotify's Audio Features. Their results revealing a strong correlation between audio features and specific contextual conditions, and underscore the potential benefits of including contextual factors in recommender systems. Additionally, they pointed out the personalised model delivered promising outcomes, which consistently outperforming the global model and enhancing the initial non-contextual recommendations.

Unlike the contextual pre-filtering and post-filtering approaches, which can both be achieved by adopting classic 2D recommendation algorithms, the contextual modelling approach introduces truly multidimensional recommendation methods (Adomavicius & Tuzhilin, 2011). Oku et al. (2006) proposed a restaurant CARS incorporating four contextual dimensions including time, companion, schedule, and weather. They employed Support Vector Machine (SVM) classification methods to analyze the set of liked and disliked items by users across various scenarios. Their context-aware SVM model significantly outperformed non-contextual SVM based recommendation algorithms. A more insightful example is given by Karatzoglou et al. (2010), who developed the Multiverse Recommendation based on Tensor Factorisation. This approach offers a straightforward way for integrating contextual information into the model, enabling it to handle n-dimensional data and incorporate various contexts by combining into a single model. A key milestone in the development of contextual modelling approaches occurred when Rendle et al. (2011) introduced Factorisation Machines

(FM), representing a significant advancement in this field. FM combines the strengths of Support Vector Machines (SVM) with factorisation models, allowing for the straightforward integration of contextual factors into predictive models. Its ease of use, even for those without expert knowledge in factorisation models, has contributed to its widespread adoption in the development of CARS. Hariri et al. (2013) employed the LDA model for the joint modelling of users, items, and the meta-data associated with contexts. Recently, Zheng & Zhu (2017) proposed a Preference Integration (PRIN) approach to conduct division learning and preference mining for CARS. In order to solve the sparsity and dimensionality challenges within CARS. Livne et al. (2019) suggested a new Sequential Latent Context Model (SLCM) trained by Long Short-Term Memory (LSTM) encoder-decoder network. The sequential latent contexts are generated from a compression of the contextual space. Their empirical results show that the SLCM model surpassed the state-of-the-art CARS models.

With the advancement of Machine Learning and Deep Learning models, there is a great opportunity to implement contextual modelling approaches in CARS. These approaches have not yet been widely tested in the domain of RS, presenting a valuable area for exploration and innovation (Santana & Domingues, 2020). Jeong & Kim (2021) proposed an autoencoder, non-supervised deep learning approach to integrate user, items and context. Their proposed model outperforms user KNN, SVD++ and Probabilistic Matrix Factorisation (PMF) in all three datasets (DePaulMovie, InCarMusic and Restaurant (Tijuana)).

Unfortunately, however, contextual datasets in the food domain are relatively rare, and the implementation of contextual modelling in the areas of food and healthy food recommender systems has not been extensively explored. Before developing a CARS model, it is crucial to investigate the potentially influential factors that impact people's food choices and decision-making processes. The following section will present a review of the literature from the fields of psychology and food science to guide the further development of food and healthy food CARS.

2.3.3 Context-Aware Recommender System in the Food Domain

Several contextual factors have been shown to be influential in food recommendations. One example is gender. Cavazza et al. (2015) carried out a study on gender-based stereotypes about food type, portion size and dish presentation. In short, their study confirmed that women are more interested in smaller and elegantly presented meals than the large and rough meals. Rokicki et al. (2016) generated 88 features to examine the gender differences in online cooking (e.g. with preference for dishes or the use of spices), they take gender into account for a simple CF approach and acquire the positive model performance. Time is also considered as an important factor. Kusmierczyk et al. (2015a) analysed the data from large-scale German online food platforms and found clear temporal (seasonal and weekly) patterns in online recipe producing and uploading behaviour. Trattner et al. (2017b) investigated the relationship between cooking interests (e.g. Mediterranean or Middle-Eastern cuisine), hobbies and nutritional values of online recipes. They also suggested that learning the patterns between a user's hobbies and eating preferences could provide the motivating goals for persuasive systems. Location, as the most popular contextual factor, has also been widely researched in the food domain (Cheng et al., 2017; De Choudhury et al., 2016; Zhu et al., 2013). Cheng et al. (2017) investigated the influence of city size on dietary preferences and found that incorporating city size into the RS can also improve the recommendation performance. Zhu et al. (2013) examined the similarity of regional cuisines based on geography and climate change. They argued that geographical instead of climate proximity is a crucial factor when determining cuisine similarity. Moreover, the availability of food as a factor has been researched by De Choudhury et al. (2016). They examined food preference and nutrition consumption via social media platforms in order to solve the 'food deserts' challenges. Harvey et al. (2012) investigated several key contextual factors contributing to how recipes are rated by users. The users not only provide rating data but also specify the reasons behind their rating, such as not liking a certain ingredient or that it contains too many calories. By analysing these factors with a regression model, the results show that both the availability of ingredients and temporal factors can influence the recommendations. Teng et al. (2012) conducted in-depth research using complement and substitution ingredient networks to understand the user's recipe preference and to predict which recipe will acquire the best ratings. Oh et al. (2010) proposed an application of context-aware FRS that can inform users of available foods by applying contextual information obtained from sensed profiles, physiological signals, and environment conditions. They did not use traditional CF or CBF based recommender techniques, instead using 5W1H (who, when, where, what, how, and why) to fuse all context information and only test with the four users.

All the aforementioned research employed relatively simple and classical statistical techniques to examine how recipes were rated or uploaded under different contextual scenarios, rather than focusing on providing food or healthy food recommendations. The challenge still remains about how to integrate this contextual information and understand which are the most valuable features, and how to best account for these algorithmically to provide personalised food and healthier recommendations (Trattner & Elsweiler, 2017a; Rokicki et al., 2017). Moreover, most proposed approaches are restricted to only one type of context (e.g. location, time, gender). This research builds on the success of previous studies by exploring whether novel contextual factors, such as emotional status and physical environment, influence eating behaviour, nutritional intake, and recipe ratings. Additionally, the study aims to examine how the combination of recipe content, user demographics, and contextual features can enhance the performance of recommender systems compared to traditional CF and CB baseline methods.

Another significant domain-specific challenge arises when balancing nutrition with people's food preferences, especially when both aspects are integrated into a food recommendation system (Elsweiler et al., 2015). Simply considering the accuracy of recommendation results isn't sufficient to build as ideal FRS, as the most popular online recipes tend to be unhealthy (Trattner et al., 2017a). Addressing trade-off challenges should be a top priority in order to develop more effective FRS. The following section will provide an overview of the current developments in the food recommendation domain, and discuss common approaches for incorporating healthiness into RS.

2.4 Anthropological, Ecological, and Psychological of Food Choice

Food choice is a complex and multidimensional process influenced by a wide range of contextual factors, including cultural traditions, social norms, environmental conditions, and psychological mechanisms. When developing a CARS for food and healthy food recommendations, identifying the most influential contextual factors should be a top priority. Given the overwhelming number of contextual factors that could potentially influence food preferences and decision-making, it is clearly impractical to measure them all. Therefore, understanding which elements have the greatest impact on human food intake behaviour is crucial for creating effective recommendations. The following section explores food choice through three key perspectives: **anthropology, ecology, and psychology**. Anthropology provides a broad cultural and societal perspective on dietary behaviour (Section 2.4.1), while ecological models emphasize the interactions between individuals and their environments (Section 2.4.2). Meanwhile, psychology examines cognitive, emotional, and behavioural processes that influence food choices (Section 2.4.3). Together, these disciplines offer a comprehensive understanding of the factors shaping dietary decisions.

2.4.1 Cultural, Historical and Social Contexts in Anthropological Food Choice

The study of food anthropology provides critical insights into how cultural, historical and social contexts shape food preferences and dietary habits. Since food is more than sustenance; it is a symbolic, cultural, and social construct that reflects identity, tradition, and history (Mintz & Du Bois, 2002). Understanding these contextual influences is crucial for developing context-aware recommendation systems that promote healthy eating by aligning food suggestions with diverse populations' dietary habits, preferences, and evolving nutritional needs.

Tiu Wright et al. (2001) highlighted how geographical, historical, and economic contexts within a **culture** influence individuals' food preferences, shaping dietary traditions and consumption patterns over time. For example, China, with its vast territory and significant regional diversity, exhibits a wide range of distinct food traditions shaped by local climates, agricultural resources, and historical exchanges (Murphy et al., 2020; Beecroft, 2010). **Religion** also plays a pivotal role in shaping dietary habits. In East Asian cultures, fermented foods are deeply embedded in both religious practices and traditional culinary customs (Tiu Wright et al., 2001; Siddiqui et al., 2023). Furthermore, food symbolism reinforces cultural values and societal norms. Certain foods hold ritualistic significance and are central to festivals and celebrations, such as turkey being a staple of Thanksgiving in the United States and dumplings being traditionally consumed during the Chinese New Year (Counihan et al., 2013). **Historical** influences further contribute to food preferences. Mintz (1986) demonstrated how sugar consumption in Europe was shaped by colonial exploitation and the transatlantic slave trade, leading to the widespread integration of sugar into European diets. Similarly, globalization has facilitated culinary exchanges, fostering the development of fusion cuisines. The spread of spices from India to Europe, the introduction of tomatoes to Italy from the Americas, and the global popularity of fast food influenced by American culture all exemplify how food preferences evolve through historical interactions (Goody, 1982). These examples illustrate that food preferences are not static but are continually shaped by cultural, and historical factors. As globalisation accelerates, culinary traditions continue to evolve, blending influences from multiple regions while maintaining cultural significance.

The anthropological perspective on food preference also highlights that broader **social class** and **economic status** play a crucial role in determining access to food, thereby influencing dietary habits across different socioeconomic groups (Islam et al., 2019; Cai, 2024; Monterrosa et al., 2020). Bourdieu (2018) argued that food consumption patterns serve as markers of social distinction, with elite classes favouring gourmet and organic foods, while working-class diets often depend on affordable, mass-produced options. Furthermore, family and peer influences significantly shape food preferences from an early age. Research has shown that parental eating habits and communal dining experiences contribute to lifelong dietary choices (Rozin, 1996). In addition, contemporary social media and global food trends increasingly influence younger generations, introducing new dietary patterns such as veganism, organic eating, and fusion cuisine (Johnston & Baumann, 2014). These social and economic factors demonstrate how food choice is not merely an individual preference but a culturally and structurally embedded practice shaped by economic access, social influence, and evolving food trends. To further explore the contextual interactions among individual, societal, and environmental influences on food behaviour, researchers in interdisciplinary fields such as public health, nutrition, and the behavioural and social sciences have proposed ecological frameworks to explain these complex relationships (Story et al., 2008; Penney et al., 2014; Monterrosa et al., 2020). These factors and frameworks will be discussed in detail in the following subsection.

2.4.2 Ecological Framework: Interactions Between Individual, Societal, and Environmental Factors in Shaping Dietary Behavior

From an ecological standpoint, food choice is shaped by dynamic interactions between individual, societal, and environmental factors (Story et al., 2008; Bronfenbrenner, 1979). This framework is rooted in Ecological Systems Theory proposed by Bronfenbrenner (1979), which posits that human behaviour is shaped by multiple layers of influence, ranging from personal choices to broader social and environmental contexts. Within the domain of public health and nutrition, researchers have employed ecological models to explore how these interrelated factors shape food choices, eating patterns, and adherence to healthy dietary habits (Story et al., 2008; Monterrosa et al., 2020).

The Social-Ecological Model (SEM) proposed by Story et al. (2008) is one of the most widely used frameworks for understanding the multi-level determinants of healthy eating behaviour. This model illustrates how food choices are influenced by four interconnected levels, shown in Figure 2.8: **individual**, **social**, **physical**, and **macro-level environments**. Individual factors, including cognition, behaviour, and demographics, shape eating habits through motivations, self-efficacy, and behavioural capability. The social environment, encompassing interactions with family, friends, and peers, influences food choices through role modelling, social support, and social norms. The physical environment, which includes settings such as homes, workplaces, schools, and supermarkets, determines food availability and creates barriers or opportunities for healthy eating. At a broader scale, macro-level factors, such as food marketing, agricultural policies, economic conditions, and societal norms, exert a powerful but indirect influence on dietary behaviours. These four levels interact dynamically, collectively shaping what people eat. Story et al. (2008) support the idea of this research on a broader scale, emphasising that individual behaviour change in improving dietary and lifestyle patterns is extremely difficult to achieve without addressing the context in which people make decisions. To support healthier choices, it is essential to take significant initial steps to ensure that nutritious food options are widely accessible, clearly identifiable, and affordable. This approach should be inclusive of people from all racial and socioeconomic backgrounds and consider diverse geographic settings, including urban, suburban, and rural areas.

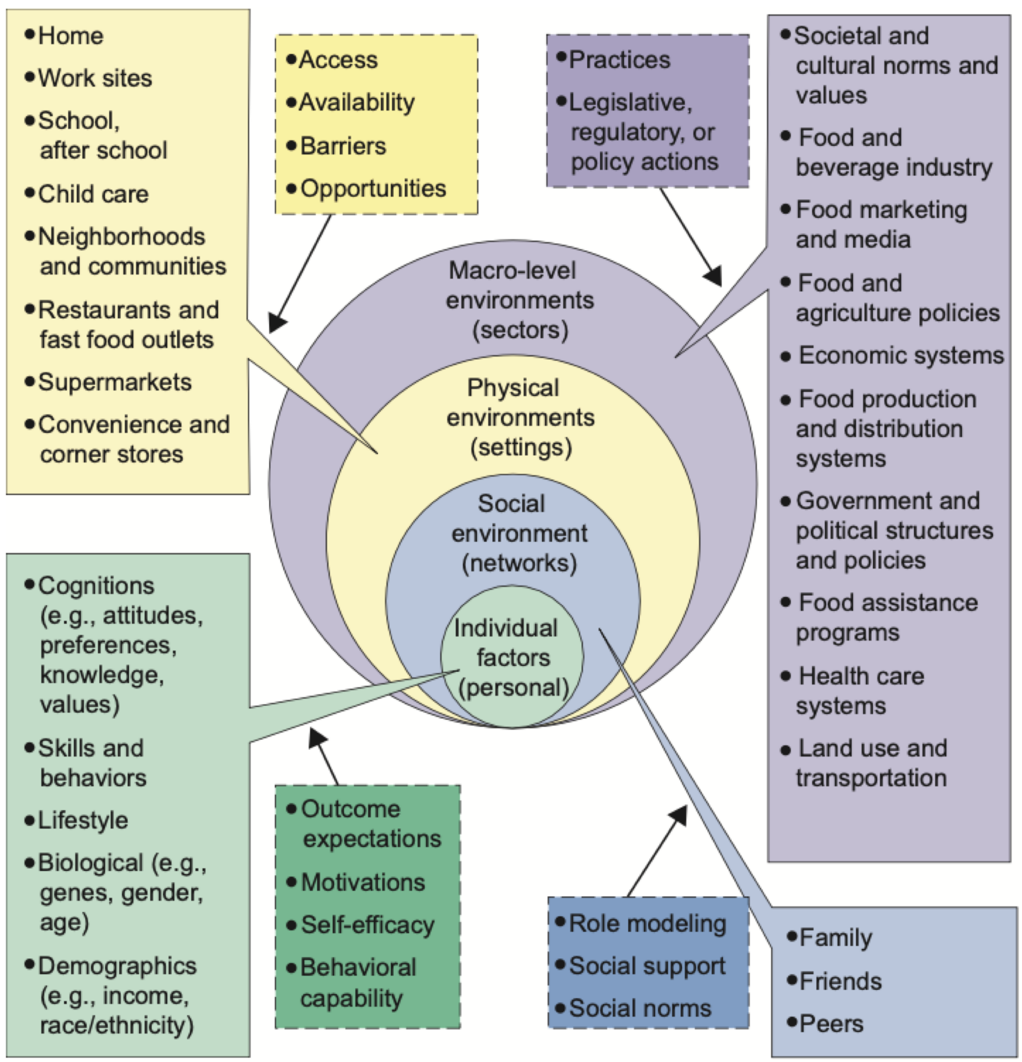


Figure 2.8: *An ecological framework depicting the multiple influences on what people eat ((Story et al., 2008)*

Later on, Contento & Koch (2020) further explore and expand on individual biological factors influencing food choices and dietary behaviours, suggesting that humans have inherent, biologically determined behavioural predispositions. These predispositions contribute to sensory-specific satiety mechanisms, which, in turn, shape food choices and diet-related behaviours (as shown in Figure 2.9). This effect may be particularly pronounced in children. Such mechanisms had adaptive value by ensuring that people consume a variety of foods and obtain the necessary nutrients for survival. However, in today's food environment, Monterrosa et al. (2020) point out that the extensive variety of foods available in modern meals may contribute to overweight and obesity. The positive aspect is that the biological factors controlling food intake can be influenced by learning and experience or altered by disease states.

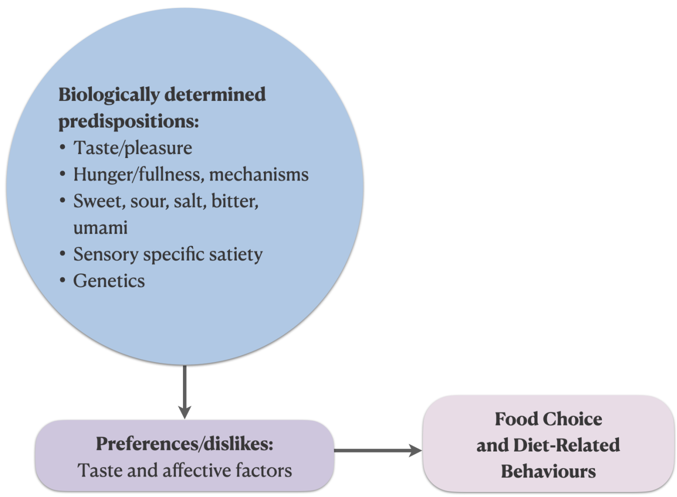


Figure 2.9: *Our biologically determined behavioural predispositions that influence food choices and dietary behaviours ((Contento & Koch, 2020)*

2.4.3 Contextual Features in the Psychology of Food Choice

In the fields of psychology and food science, various factors have been discussed and potentially identified as influential on people's food choice and nutritional intake behaviour, providing valuable insights for the development of context-aware food recommender systems (Cameron et al., 2015; Shepherd & Raats, 2006; Devine, 2005). However, little research has focused on capturing and controlling both dynamic and static contextual factors to gain a better understanding of people's food preference, and further exploring how incorporating these factors could improve recommendation performance. This research aims to bridge this gap. Shepherd & Raats (2006) proposed a food choice process model, emphasising the fact that eating preferences are fluid and change throughout time, and that people's food choices should be reconstructed as their lives progress and change. ***Trajectories***, ***Transitions***, ***Timing***, ***Ideals***, ***Personal variables***, ***Resources***, ***Social factors***, and ***Environments*** are among the eight factors that may impact human food choice.

Trajectories refers to people developing their own food and culinary preferences under certain settings and historical contexts, then later become tenacious, momentum and continuity, which demonstrate their own characteristics (Devine et al., 1999). For example, if a person grew up with a family tradition of eating salad at dinner, they may insist on continuing on this trajectory for the rest of their lives. ***Transitions*** means an important life status change which leads to modification of behaviours, these changes may include entering school, changing employment, and moving to a new environment. ***Timing*** refers to a transition or turning point in an individual's life that may effect their dietary choices. For example, mothers will choose to adopt healthier eating habits during pregnancy and child-rearing. ***Ideals*** are learnt culturally through families and other institutions, and they represent eating objectives and expectations. Individuals consider cultural and sub-cultural norms in food selection when determining which foods are accredited and desirable for consumption among broader cultural and ethnic groups. ***Personal variables*** are more related to the individual itself. These personal factors include physiological factors (sensory, endocrinological, genetic, etc.), psychological and emotional characteristics (preferences, personalities, moods, phobias, etc.) and relational factors (identities, self-concept, etc.) (Bove et al., 2003). Reid & Hammersley (1999) pointed out that mood could influence food choices by altering appetite or influencing other behaviours that limit or alter food availability. Cameron et al. (2015) found out the positive mood may point to healthier food choices. Macht (2008) suggested that emotions may regulate eating and eating may regulate emotions. ***Resources*** related to assets are available for people making food decisions and can be both tangible (money, equipment) and intangible (time, knowledge, and cooking skills). ***Social factors*** refer to the relationships which surround people, influence food choice and play a role in social groups such as family, friends, clients (Sobal & Nelson, 2003). ***Environments*** represent the border contexts in which lifecycle changes take place, such as social structure, economic situations, historical eras, and shifting physical settings (Devine, 2005). Most environmental changes, such as food supply seasonality or the historical background of mass media marketing, advertising as a backdrop for food information, cause reconstruction of people's dietary choices (Avery et al., 1997).

In terms of personal food choice systems, how people make food choices not only depends on the taste of food, but also related factors such as convenience, cost and health considerations. Additionally, food choices may be shaped by self-identity and perceived needs (Connors et al., 2001). However, how to effectively incorporate contextual information into RS to make more intelligent and personalised recommendation, remains a topic that warrants in-depth research. Based on previous reviews in the food science domain, most research has focused on measuring a single type of factor—whether personal, emotional, temporal, or environmental. No studies have yet considered bringing multiple contextual factors, particularly dynamic contextual factors, together to identify the most influential ones on people's food preferences and nutritional intake behaviour. This gap highlights the need for a more comprehensive approach that considers multiple contextual factors simultaneously in research design, to optimise the understanding of contextual features from different perspectives and enhance the capabilities of FRS.

Relatively few dynamic contextual factors have been identified as statistically significant in predicting people's food preferences or recommending personalised nutritional recipes. This research intends to explore various contextual features (both static and dynamic) and how they affect people's food choice and nutritional intake behaviour, and further integrate these contextual features into recommender systems models aim to improve the model performance of making healthy food recommendations. Among the previously discussed factors, personal variables—particularly psychological and emotional factors—are more closely related to the individual. However, due to the difficulty in capturing these factors, they remain relatively novel in the FRS domain. This research will highlight these factors, emphasising their potential significance in enhancing personalised food recommendations. The following section discusses the development of context-aware FRS.

2.5 Food Recommendation and Incorporating Healthiness

While modelling human behaviour is challenging, influencing human behaviour poses even more of a challenge (Gupta et al., 2024; Maier & Cash, 2023). As of 2021, 1.1 billion people have suffered from hypertension, 2 billion people are overweight, 370 million people suffer from diabetes, and 350 million have pre-diabetes (World Health Organisation, 2021). Some of these diseases can be prevented and even reversed by regulating and balancing eating behaviours and lifestyles (Trattner & Elsweiler, 2017a), and this is where healthy food recommendations come into play (Freyne & Berkovsky, 2010; Freyne et al., 2011; Harvey et al., 2012). Ricci et al. (2010) claim that online RS are a useful technique to overcome the problem of information overload in diverse situations, as well as assisting users' decision making process and even changing their behaviour.

There are various challenges in the food domain which make the recommendation even more challenging than in other areas. Firstly, predicting a user's food preference is complex because it is multi-faceted, heavily influenced by culture, and even genetically determined, not to mention context-dependent (Min et al., 2019b). For example, different interests or behaviours such as reading a book at home, doing exercise, being with friends or with family may lead to various different food choices. Moreover, additional constraints such as allergies or lifestyle preferences makes FRS unique compared to other domains (Min et al., 2019a). Other challenges may be related to the usage of food terms (i.e. the same item may be expressed using very different terms), and the fact that ingredients can be prepared in different shapes (for example, cut in square or triangle) and using different preparation methods. Unlike the products or media recommendation domain, it is not always clear whether the recommended item can be prepared, consumed, or liked by users due to the lack of ingredients and cooking knowledge (Trattner & Elsweiler, 2017a).

From the user perspective, an ideal FRS should aim to solve the trade-off between personalised food preference and nutrition or health requirement (Elsweiler et al., 2015; Min et al., 2019a). Based on the research of Trattner, Elsweiler et al. (2017), they made an astonishing discovery that the internet recipes sourced from allrecipes.com (a famous online recipe website in US) tend to be less healthy (i.e., higher in protein, fat, saturated fat and sodium) compared to recipes from leading chefs and ready meals from leading UK supermarkets. This highlights one domain-overarching problem, which is the users tend to stick with popular recipes, but the popular recipes tend to be unhealthy (Rashid et al., 2002; Ekstrand & Willemsen, 2016; Trattner et al., 2018; Starke et al., 2020). FRS typically struggle to find healthier alternatives that align with users' preferences, especially when relying solely on traditional techniques (e.g., CF and CBF) for generating recommendations (Starke, 2019).

Trang Tran et al. (2018) showed that there are three types of healthy food recommender systems, namely a focus on user preference, a focus on a user's nutrition needs and a balance between both aspects of user preference and nutritional needs. The Table 2.5 below summarizes the state-of-the-art recommendation algorithms in the healthy food domain according to these three categories. CF and CBF have emerged as the dominant algorithms in the early development of healthy food recommender systems. In recent years, knowledge-based approaches and deep learning techniques have gained increasing attention. However, there is insufficient evidence to indicate that previous research has adequately addressed the incorporation of contextual knowledge into recommender system algorithms. A detailed discussion of relevant studies in the healthy food recommendation domain is presented below.

Table 2.2: *Summary of state-of-the-art healthy food recommender systems*

RS Types	Reference	RS Approach	Algorithms	Novelty and Functionality	Limitations
Considering user preferences	(El-Dosuky et al., 2012)	Knowledge-based recommendation	IF-IDF and cosine similarity	Incorporate healthy heuristic standard food database into knowledge-base to make recommendation	Does not consider personal nutritional needs, so the health of recommended recipes cannot be guaranteed
	(Freyne and Berkovsky, 2010)	CF, CBF, and Hybrid recommendations	Similarity computing and logistical decision tree algorithm	Break down into recipe into ingredients when making recommendations	
	(Freyne et al., 2011)	CF, CB, Machine learning techniques	Similarity computing and logistical decision tree algorithm	Using machine learning techniques to understand user reasoning when making recommendations	
	(Elahi et al., 2015)	Active Learning	Matrix factorization	Capable of learning both short-term and long-term user preferences	
	(Gao et al., 2019)	Deep Learning	Hierarchical attention network	For visually aware food recommendation	
Considering users' nutritional needs	(Ueta et al., 2011)	Goal-recipe recommendation	Create co-occurrence database listing 45 Common nutrients	Enables users to retrieve easily for nutrition recipes with natural language to improve specific health conditions	Does not consider past search history of user
	(Aberg, 2006)	Hybrid (CF and CBF) and constrain-based recommendation	Parameter-based and item-based algorithm	Acting for elderly people to improve their nutrition food consuming	Model working well in small dataset
	(Wang et al., 2021)	Health-aware recommendation	Item2vector and Deep learning model	Retrieval of the recipe based on user health profile, also able to monitor the ingredients availability in the market	Does not consider past search history of user
Balanced between both aspects	(Harvey and Elsweiler, 2015)	CF, CBF	Weighted model, SVD	Presenting personalised healthy meal plan, using Harris-Benedict equation to estimate individual calorie requirement	This is a demo presentation, shows the possibility of combining nutritional guidance into healthy meal plan RS

Continued on next page

RS Types	Reference	RS Approach	Algorithms	Novelty and Functionality	Limitations
	(Elsweiler et al., 2015)	CF, CBF	Weighted model, SVD	Proposing three ways to incorporate nutritional information into recommender systems	Proposed three ideas, still need experiment to find the best performing algorithm
	(Chen et al., 2020)	Knowledge-based recommendation	QE (Query Expansion) module and KA (Knowledge graph Augmentation) module	Presenting a question-answering over knowledge-based (KBQA) FRS that regards personal dietary preferences and healthy guidelines	The model performance is around 61% MAP and 60% MAR, there is still a lot of room for improvement
	Pecune et al. (2020)	CF	Alternating Least Squares (ALS), Bayesian personalised Ranking (BPR) and Logistic Matrix Factorization (LMF)	Investigate whether introducing healthy tags when providing recipe recommendations would have an influence on people's decision making	Not enough diversity in recommendations

2.5.1 Considering user preferences

Learning users' dietary preferences is a crucial step to recommending dishes that they may like. Freyne & Berkovsky (2010) conducted seminal research by breaking down recipes into ingredients, and compared the performance of CF, CBF and hybrid models. In their study, a content-based model overperformed other two approaches when generating recommendations in ingredient level. A year later, they improved their research by cooperating user reasoning patterns into content-based algorithms when providing recommendations (Freyne et al., 2011). Later on, El-Dosuky et al. (2012) used the Term Frequency-Inverse Document Frequency (TF-IDF) and cosine similarity to identify similar recipes and user profiles. Moreover, the novelty of their research is the incorporation of healthy and standard food databases extracted from the United States Department of Agriculture (USDA) to a knowledge-based system which subsequently modifies the recommendations. In another study, Elahi et al. (2016) applied active learning and matrix factorisation into a FRS, which is capable of learning long-term user preferences. Gao et al. (2019) have proposed a Hierarchical attention network for visually aware food recommendations. While their model has outperformed state-of-the-art RS algorithm and achieved an Area Under Curve (AUC) score of around 64%, the nutritional value of the recommended recipes is not disclosed, making it difficult to determine whether the suggested recipes are healthier alternatives. Although considerable progress has been made in providing food recommendations aligned with users' past preferences, health-related issues have rarely been addressed in previous research (Trang Tran et al., 2018).

2.5.2 Considering user nutritional needs

As unhealthy lifestyles and dietary habits could be the major factors causing high levels of obesity (Chen et al., 2022), FRS are capable of being an intelligent nutrition consultation system for preventing or even improving these circumstances. Several FRS prioritize delivering personalised healthy recommendations. Ueta et al. (2011) proposed a goal-oriented recipe RS to provide dishes that contain the right type of nutrition to treat users' health problems. They created a co-occurrence database incorporated into the system that listed 45 common nutrients for common ailments such as colds, bone or acne. This system is able to seek appropriate recipes that meet users' specific requests, for example, the prevention of diabetes or the relief of fatigue. Aberg (2006) demonstrated a parameter-based healthy FRS for elderly people by considering user's specific requirements and environmental information such as nutritional values, dietary restrictions, preparation time and the availability of ingredients. Similarly, and based on the study of Aberg (2006), Trang Tran et al. (2018) improved a constraint knowledge model by incorporating more features such as time, cost, energy, protein, allergies, and disease. Most recently, Wang et al. (2021) proposed a health-aware FRS named Market2Dish, which is capable of mapping the ingredients in the market, then to creating healthy dishes eaten at home, and user's health profiles which are taken from social networks. The main obstacles for such pure health-aware recommender systems are tracking whether users consistently follow the system's guidance and whether they actually enjoy the recommendations. Since users' specific health-related needs often outweigh personal preferences, it is a significant challenge for these systems to provide healthy eating guidance that users not only adhere to but also enjoy (Trang Tran et al., 2018).

2.5.3 Balancing between preferences and nutritional needs of users

If the system only considers the user's preference, then for those users who have unhealthy dietary patterns, their unhealthy behaviour will be constantly encouraged. For example, if a user fancies eating high-calorie or fatty food, then the system will recommend other high-calorie or fatty foods, thus likely making the problem worse. On the other hand, if the system focuses entirely on the user's nutritional needs and neglects his past preferences, the recommended recipes may not be at all appealing to the given user, as so they may still insist on eating the recipes they like that may not be so healthy, which will not improve their living habits. Thus, considering and balancing both aspects seems to provide the best solution. Relatively few studies focus on this type of healthy FRS. Harvey & Elsweiler (2015) presented a healthy FRS by figuring out the trade-off and providing recipes not only liked by the users, but also ones that fit their daily nutrition needs. They collect the user demographic information including age, gender, height, weight, daily activity level and goals namely weight loss, weight gain or weight maintenance. They integrated an updated version of the Harris Benedict equation (Roza & Shizgal, 1984) to calculate the daily nutrition requirements of each user, then adjusted the preference recommendation result. Another highlight of this research is the recommendation of complete meals (breakfast, lunch, dinner) rather than individual meals. In the meantime, they also proposed three ideas of incorporating nutritional information into the recommender system, but still need further experiment to find the best performing algorithm (Elsweiler et al., 2015). Trang Tran et al. (2018) discussed a similar but simpler RS, which generates users' occupation, physical activity and health problems and then estimates the daily energy requirements filtering the recommended recipes based on the aforementioned information. Recently, Chen et al. (2020) proposed a knowledge-based personalised FRS with consideration of users' unique health requirements and their preferences. However, their model has not performed well, as all the models have Mean Absolute Precision (MAP) results lower than 61%. Similarity, Pecune et al. (2020) utilised classic CF approach to build all three types of healthy and personalised recipe recommendation systems (preference-based, healthy recommender and hybrid recommender) and investigated whether adding healthy tags will impact people's decision-making through a between-subject study. The result shows that the people who cared about the healthy tag were more likely to choose hybrid RS. However, there is not enough diversity in their recommendations. They suggested that adding the know-based post-filtering after the classic CF-based algorithm may be able to solve this problem, which highlighted the necessity for developing context-aware healthy food recommender systems. Although the afore-mentioned approach provided valuable insight for addressing the trade-off between user's preference and nutrition needs, the suitability of combining recipe content (e.g. ingredients, cooking methods, cooking complexity) and complex contextual factors into a recommender system are still worth being considered into more detail in order to make a more appealing recipe (Elsweiler & Harvey, 2015).

In summary, there is a significant gap in linking healthy recommendation with contextual factors. Individuals may not consistently maintain an unhealthy diet under all circumstances, and making blanket health recommendations could potentially overwhelm or frustrate users. Indeed, specific situations may trigger or encourage unhealthy eating habits. Therefore, it is crucial to examine whether users exhibit varying dietary and nutritional intake patterns under different contextual situations. Additionally, it is important to investigate when it is most necessary to provide healthy recommendations and how to integrate nutritional information to reduce the likelihood of recommendation rejection.

Prior research has placed insufficient focus on healthy FRS that can simultaneously align with users' past preferences while optimising for healthier choices. No research has yet integrated both dynamic and static contextual factors with complex multimodal features to achieve optimal rating prediction models. Furthermore, the trade-off between users' preferences and nutritional needs within specific contexts remains unaddressed, leaving an opportunity to develop context-aware healthy recommendations that may be more easily accepted by users. Such a system would also be capable of providing diverse recommendations and, ideally, offer partially-explainable results to build trust and encourage adoption by potential users. Existing literature on context-aware healthy recommendations often emphasises conceptual ideas or prototypes rather than developing robust contextual modelling algorithms. This underscores the need for more comprehensive research to identify which contextual and food-related factors most significantly influence recipe rating predictions and the effectiveness of healthy recommendations.

2.6 Recommender System Evaluation

Since user preferences may fluctuate, evaluating RS algorithms is not always simple or straightforward. The evaluation process would benefit from frequent user engagement and participation, as this dynamic nature adds complexity to accurately assessing the system's performance over time. Besides, there often exists a trade-off between what users like and what they actually need. In this case, the evaluation should be multidimensional and comprehensive (Ricci et al., 2015). During the past decade, a vast amount of recommendation algorithms have been designed, but how to select the best performing algorithm under certain circumstances has become crucial. Initially, the ranking power of recommendation results has been treated as common evaluation regulation. However, it can now be argued that accurate predictions, recommendation diversity, recommendation efficiency, and protection of user privacy are all important, and a single regulation would be insufficient (Pu et al., 2012). According to Gunawardana et al. (2012), there are three experimental ways to evaluate RS: *offline tests*, *user studies* and *online experiments*.

The attraction of an offline experiment is its lower cost, as it can be carried out with pre-collected datasets from user selection or ratings without the need for real user interaction (Ricci, 2022). This means there is no engineering risk of online development and a waste of valuable online traffic resources. When this experiment is used, the user's integration with the system will be simulated and guide us to making reliable decisions, as well as allowing multiple sets of parallel testing to be carried out together. Among them, the main methods of offline evaluation include holdout inspection, cross-test, leave one verification, and self-help method. The evaluation index mainly includes user satisfaction, prediction accuracy, recall rate, coverage, diversity, novelty, popularity, Root Mean Squared Error (RMSE), AUC, and ROC curve (Shani & Gunawardana, 2011). However, the main drawback of offline texts is that they can only answer a narrow set of questions, particularly about the prediction power of an algorithm. Also, the whole evaluation of how RS influence user behaviour are based on prior well modelled by RS algorithm otherwise, the evaluation cannot be directly measured.

As RS naturally rely on user interaction, sometimes it is difficult to simulate user interactions with the offline system, and this is where there is a need for user study. Moreover, even when offline tests are possible, studying with real users can still bring valuable information about system performance. A user study normally starts with the recruitment of a group of users and asking them to fulfil some interaction task, as well as observing their actions and behaviour. However, user studies are relatively expensive to conduct (Kohavi et al., 2013). In order to avoid failure of experiment, a reasonable experimental design and real interaction simulation would be worth further study.

The real effect of RS on user behaviour is related to various aspects, online experiments provided the strongest evidence of how the system performed as the system is used by real users and perform real tasks. However, there are still some considerations, for example, if a text system suggests irrelevant recommendations and thereby heavily affecting users' experience, they may discontinue its use. Furthermore, this would be unacceptable in an e-commercial application. Wu et al. (2012) classified evaluation criteria from a system and user aspect based on the recommender algorithm. From the system angle, the confidence, scalability and adaptivity of recommendations should be important factors, meaning that if the RS is to be trusted, if the content has better explained, as well as if the designed system can be scaled up and operate in real life dataset. From a user perspective, if the user is satisfied and trusts the system, if their privacy information has been protected? This is essential for RS to measure. In recent years, there has been an increased interest in user-centred evaluation metrics of RS, as mentioned by Díaz et al. (2008). There is evidence that being able to explain the recommendation can significantly increase user trust (Abdollahi & Nasraoui, 2018; Balog et al., 2019; Fu et al., 2020).

2.7 Discussions and Summary

Based on the above discussion, various researchers have made contributions towards developing a more intelligent RS. However, the challenges still remain, particularly on solving context-awareness, cold start, diversity and explainability problems (Park et al., 2012; Felfernig et al., 2013; Sharma & Mann, 2013; Sharma & Gera, 2013b; Tintarev & Masthoff, 2015; Khusro et al., 2016; Kunaver & Požrl, 2017; Mohamed et al., 2019). In food recommendation domain, solely focusing on the accuracy of recommendations results may not sufficient for creating an ideal FRS. Doing so may inadvertently encourage unhealthy eating behaviours, as not all individuals naturally enjoy healthy eating habits. An ideal FRS should strike a balance between healthiness and taste, being able to identifying healthier alternatives that align with users' expectations and preferences across various contextual situations (Starke, 2019). Addressing this challenge is a crucial area for further research (Elsweiler et al., 2015; Trang Tran et al., 2018).

Among healthy FRS, balancing between user preferences and nutritional needs shows a great opportunity to develop an intelligent healthy FRS that users are likely to use as well as follows its guidance (Harvey & Elsweiler, 2015). Relatively few studies have focus on development of this type of healthy FRS. This may be due to the inherent challenges associated with integrating health and nutritional information, which may often lead to a reduction in prediction accuracy and recommendation performance. Moreover, the task of identifying healthy alternatives that users may potentially be willing to accept and consistently follow has complicated the entire design process and poses an even more significant challenge.

Although CF and CBF are among the most classic and traditional RS algorithms and have been dominant for years (Lops et al., 2011; Aggarwal et al., 2016; Schafer et al., 2007), they lack the ability to deliver highly personalised and context-specific recommendations that can adapt to various situations. These algorithms generate recommendations based on hidden latent factors, which often lack transparency and the ability to provide explainable results, making it difficult for users to understand why certain recommendations are made (Ricci, 2022). Integrating relevant contextual factors and making recommender systems context-aware has become an inevitable trend (Khusro et al., 2016).

Due to the complexity and high cost (in terms of time and labour) involved in investigating the influential contextual factors on people's preference and decision making process, relatively less research has focused on this area. Particularly in FRS, current research has primarily restricted to measuring a single dimension of contextual factors (Trattner & Elsweiler, 2017a; Rokicki et al., 2017), for example, either gender, location or time (Cavazza et al., 2015; Rokicki et al., 2016; Kusmierczyk et al., 2015a). No research has systematically explored how multiple dynamic factors affect people's eating and recipe rating behaviour, and identify the most influential contextual factors. Rare contextual factors have been provided as statistically significant factors when predicting food preferences or recommending personalised food plans.

In the meantime, the difficulty of capturing and measuring contextual factors has resulted in a lack of contextual datasets in FRS domain, further limiting the potential to developing context-aware FRS. Although contextual factors are believed to have much influence on individuals' food choice in the psychology field (Connors et al., 2001; Shepherd & Raats, 2006), the scarcity of contextual datasets presents a significant challenge for researchers aiming to develop context-aware models (Trattner & Elsweiler, 2017a). Besides, the most influential contextual factors at the algorithmic level remain unclear (Min & Han, 2005). To address this gap, it is insufficient to simply build a contextual recommendation model by randomly selecting and incorporating contextual features. There is a strong need for research that is designed to gain a deep understanding of individuals' intuitive behaviours related to contextual food choices.

In addition, there is lack of research on FRS that focus on modelling and integrating complex and multimodel features (e.g. dynamic and static contextual factors, recipe ingredients, cooking directions, recipe images, user demographic information) to achieve optimal rating prediction accuracy for users' food preferences and rating behaviours. Various machine learning and deep learning algorithms has not been widely implemented in FRS to develop contextual modelling approach. Such approaches would be capable of gaining a deeper understanding of the patterns and factors that drive different user behaviours, enabling improvements in relevance, accuracy, and personalisation of recommendations (Adomavicius & Tuzhilin, 2011).

More importantly, significant challenges remain in providing healthy recommendations that effectively balance users' historical preferences with their nutritional needs; only a few studies have attempted to address this trade-off problem (Harvey et al., 2013; Elsweiler et al., 2015; Min et al., 2019a). There is a notable gap in linking healthy recommendations with contextual factors to optimise rating prediction and deliver seamless healthy recommendations across various contextual situations, which may be more readily accepted by users. After reviewing the literature in RS, FRS and food science domain, several research gaps have been identified:

- Little work focuses on systematically measuring the relationship between contextual factors and people's food choice, nutritional intake, and recipe rating behaviours in the FRS area.
- Most studies on CARS tend to focus on integrating only one dimension of dynamic contextual factors, such as gender, time, or location.
- Currently, the field of FRS lacks a standardised contextual dataset that includes multimodal contextual factors, recipe content, and user demographic information. Such a dataset is crucial for developing context-aware healthy FRS, which may enabling more accurate and personalised recommendations.
- Little research has focused on balancing food preferences and nutritional needs to address the trade-off problem in food recommender systems under varied contexts. This gap highlights the challenge of delivering adaptive and context-sensitive recommendations that balance users' taste preferences with the promotion of healthier choices.
- Prior research has primarily focused on utilising recipe content information, often breaking it down to the ingredient level. However, no studies have focused on modelling a multimodal feature set for user rating prediction or identifying the most effective feature combinations that lead to optimal model performance.

The next chapter will outline the methods and strategies employed to address these research gaps, with a detailed focus on the research design at each stage of the study.

Chapter 3

Methodology

This chapter details the methodology employed in designing this research, and elaborates how the research questions are addressed using the gathered dataset. The research design follows an ***exploratory approach*** to initially investigate the potential influential contextual factors of individual food choice and nutritional eating behaviour within real-life experiences. This was achieved through qualitative semi-structured interviews. Based on the initial findings, a quantitative experimental study was conducted to collect contextual recipe rating preference data and to examine whether individuals' recipe ratings and implied nutritional behaviours varied under different simulated contextual scenarios. Subsequently, contextual modelling approaches and a weighted contextual healthy recommendation approach were proposed to develop a context-aware healthy food recommender system.

This chapter begins by outlining the philosophical assumptions and research approaches in Section 3.1, followed by an explanation of the research structure and overview of methodology used in this research in Section 3.2. The discussion and justification of the chosen methods and approaches for each phase are detailed in Sections 3.3 to 3.5. Ethical considerations are addressed in Section 3.6.

3.1 Philosophical Assumptions and Research Approaches

3.1.1 Philosophical Assumptions

Philosophical assumptions refer to the underlying beliefs and principles that guide researchers' approach to their study and their *"worldview"*. Addressing philosophical issues is crucial to conducting robust research (Robson, 2002). Creswell et al. (2009, p.36) highlighted four widely recognized philosophical assumptions including *postpositivism*, *constructivism*, *transformative* and *pragmatism*. The appropriate paradigm for mixed methods research is often debated. Creswell & Clark (2017) suggested that *pragmatism* is the most suitable paradigm for implementing mixed methods, as it allows flexibility between qualitative and quantitative approaches based on the research needs. Morgan (2014) supported Creswell & Clark (2017)'s view, suggesting that the mixed methods paradigm should grant researchers the freedom to make decisions and handle complex procedures effectively. As Creswell et al.

(2009, p.28) noted, the pragmatism approach "is not committed to any one system of philosophy and reality", enabling it to be either subjective or objective depending on the research questions.

Given these insights, this research was conducted within a pragmatic paradigm, leveraging its inherent flexibility to apply diverse research methods and data collection processes effectively (Johnson & Onwuegbuzie, 2004). As this research aims to develop a context-aware healthy food recommender system which capable of balancing individuals' food/recipe preference and nutritional needs under varying contextual conditions. Beginning with scenarios that realistically reflect users' dietary and healthy eating needs, and identifying the most influential factors among various contextual elements, provides valuable insights. Subsequently, integrating relevant features into the RS model based on these insights may potentially enhance the overall validity and rationality of the research. Therefore, employing both qualitative and quantitative methods to ensure a comprehensive analysis of user behaviours and perceptions is essential. Pragmatism's flexibility allows for the adaptation of approaches based on the specific needs of the study, whether qualitative or quantitative (Creswell et al., 2009). By focusing on practical outcomes, pragmatism supports the integration of various data sources and methodologies, leading to a more holistic understanding of how contextual factors influence eating, recipe rating and nutritional intake behaviours. This adaptability allows the research to remain applicable in addressing the dynamic and multifaceted nature of real-world scenarios.

3.1.2 Deductive and Inductive Approaches

From philosophy to sociology, two broad methods, namely the deductive and inductive approaches, stand out as being useful in representing the nature of the relationship between theory and social research (Bryman, 2016). In the deductive approach, assumptions and hypotheses are initially generated from an existing theory and then refined into specific research methods designed to collect data and test these hypotheses. This method of reasoning, prevalent in the natural sciences (Collis & Hussey, 2021), can be informally described as a 'top-down' approach, beginning with a comprehensive review of the literature related to the phenomena in question. Based on these findings, a related theory is generalised and subsequently narrowed down into a list of hypotheses to be tested. Observations are then made to collect data, which is analysed to assess its validity and reliability. Finally, the hypotheses are tested against the empirical evidence, leading to their confirmation or rejection (Trochim & Donnelly, 2001). However, the inductive approach begins with exploring and understanding what is happening within the research area. This initial step allows for greater flexibility, as the emphasis of the study can be adjusted based on the insights gained from this exploration (Saunders et al., 2009). Inductive approaches can be classified as 'bottom up', as they start with observations of the phenomena under investigation and generate research questions from the reviewed existing literature as well as personal knowledge and experience (Matthews & Ross, 2010). Following the observation phase, patterns are identified and hypotheses are formulated, which are then developed into a broad and general theory (Trochim & Donnelly, 2001).

This study leverages existing recommender system algorithms, machine learning, and deep learning theories, to develop a context-aware FRS. While deductive reasoning aligns closely with this research, given the structured nature of hypothesis testing in algorithm development, the lack of integration of contextual information in current food recommendation research necessitates an initial inductive approach. Considering the vast amount of contextual information available and the practical challenges in collecting all possible data, it is essential to start with an inductive approach. Semi-structured interviews were conducted to gain new insights from users about the contextual factors influencing their food choices. This qualitative method was used to identify key contextual variables that are currently underexplored. In the subsequent stages, the research transitioned to a deductive approach.

3.1.3 Mixed Methods

After determining the philosophical assumptions and logical reasoning, it is crucial to consider the research approach in the next step (Bryman, 2016). A well-defined research approach enhances the validity and reliability of findings by establishing clear methods for data collection, analysis, and interpretation, thereby ensuring credible results (Creswell et al., 2009). For this research, the choice of research approaches were made among three typical approaches: quantitative, qualitative, and mixed methods (Bryman, 2016). Qualitative research is more exploratory in nature, focusing on constructing meaning and concerning a phenomenon, and thereby revealing clearer understanding of social situations. It emphasises the qualitative 'softer' meaning in data collection and the analysis process (Creswell & Poth, 2016). Despite this, qualitative research typically involves a relatively small sample size, and the interpretation of research findings may be heavily influenced by the researcher's perspective, potentially leading to bias. This, in turn, can limit the generalisability of the research findings (Creswell & Poth, 2016). Conversely, quantitative research is inherently confirmatory in nature. It emphasizes standardised measurement, deductive reasoning, and hypotheses testing using quantitative numerical data. This approach enhances the validity and reliability of findings by providing a structured and systematic framework for data collection and analysis (Creswell et al., 2009; Bryman, 2016). However, behind many sophisticated models, there is often a trade-off with explainability. Additionally, quantitative research often misses the full complexity of human experiences and social phenomena (Denzin & Lincoln, 2011).

As research topics grow in complexity, it is necessary to create in depth understanding and break the boundaries of qualitative and quantitative methods, which has led to the emergence of the mixed methods (Klassen et al., 2012). The mixed method draws on the potential strengths of both qualitative and quantitative methods, allowing one to explore diverse perspectives and uncovering the many relationships that exist within complex situations, thereby producing a clear and rich picture (Shorten & Smith, 2017). Creswell & Clark (2017) presented six experimental design procedures including: *convergent design*; *explanatory design*; *exploratory design*; *embedded design*; *transformative design*; and *multiphase design*. Among the various research designs, an exploratory design begins with the collection of qualitative data to gain an in-depth understanding of the research area. Subsequently, quantitative methods can be employed to confirm, enrich, extend, and generalize the qualitative findings (Creswell & Clark, 2017). This approach is valuable for generating new ideas, frameworks, and directions for future research, opening up new avenues for exploration and fostering innovative solutions (Stebbins, 2001). Which aligns well with the objectives of this research. By starting from the user's perspective and gathering information on how contextual factors influence their food choices, nutritional intake, and online recipe search habits, this method enriches the metadata required for further algorithm design. It allows for the development of a recommender system grounded in a comprehensive understanding of user needs and behaviours. The exploratory design is particularly effective at capturing the complexity and context of social phenomena, enabling researchers to understand the broader environment and nuanced factors influencing the subject. Such insights are essential for building comprehensive, context-sensitive systems (Yin, 2009). As discussed above, a mixed-methods exploratory design is well-suited to this research.

3.2 Overview of Methodology

To summarise the previous discussion and justification, this research adopts a pragmatic philosophical stance, primarily following deductive reasoning. However, the research design incorporates inductive reasoning in its initial stages to identify novel influential contextual factors affecting food choice and nutritional intake behaviour, based on users' daily experiences. The overall research design follows a mixed-methods exploratory structure, carried out in three stages. The phases of the research design are illustrated in Figure 3.1. Research stage one and stage two are focused on addressing **RQ1** *(what contextual factors affect people's (online) food choices?)* and **RQ2** *(What impact do these same factors have on people's nutritional intake?)* through both qualitative and quantitive approaches. The preliminary development of a pre-filtering context-aware recommender system took place in stage two to address **RQ3** *(Can integrate these contextual factors enhance the performance of recommendation systems?)*, while the third stage continued to focus on further addressing **RQ3** and **RQ4** *(How to combine this knowledge of contextual factors to recommend people healthy recipes that they will enjoy?).*

T

The first stage of the research aims to gain new insights about the contextual factors that may potentially impact people's food choices and nutritional intake behaviours. Semi-structured interviews are particularly advantageous for this research, as they allow participants the freedom to express their thoughts and experiences comprehensively (Creswell & Poth, 2016). This approach enables the collection of more in-depth and nuanced data regarding subconscious and habitual dietary choices in various contextual situations, as well as attitudes towards nutritional information and cooking and recipe searching habits. Insights gained from this process would help identify potential influential contextual factors, guiding the subsequent stages of the research. The flexibility inherent in semi-structured interviews also allows for further probing based on participant responses, uncovering insights into the reasons behind changes in food choices under different contexts combined with their daily life and even childhood experiences. This is crucial for informing the subsequent experimental design and for enabling personalisation and providing explanations during the algorithm development phase.

In the second stage, a quantitive experimental study was conducted aims to generalise the findings from the first stage and statistically explore how people's recipe rating behaviour varies under different contextual scenarios. To facilitate large-scale studies (n=397) and ensure the recruitment of participants with diverse demographic backgrounds, participants

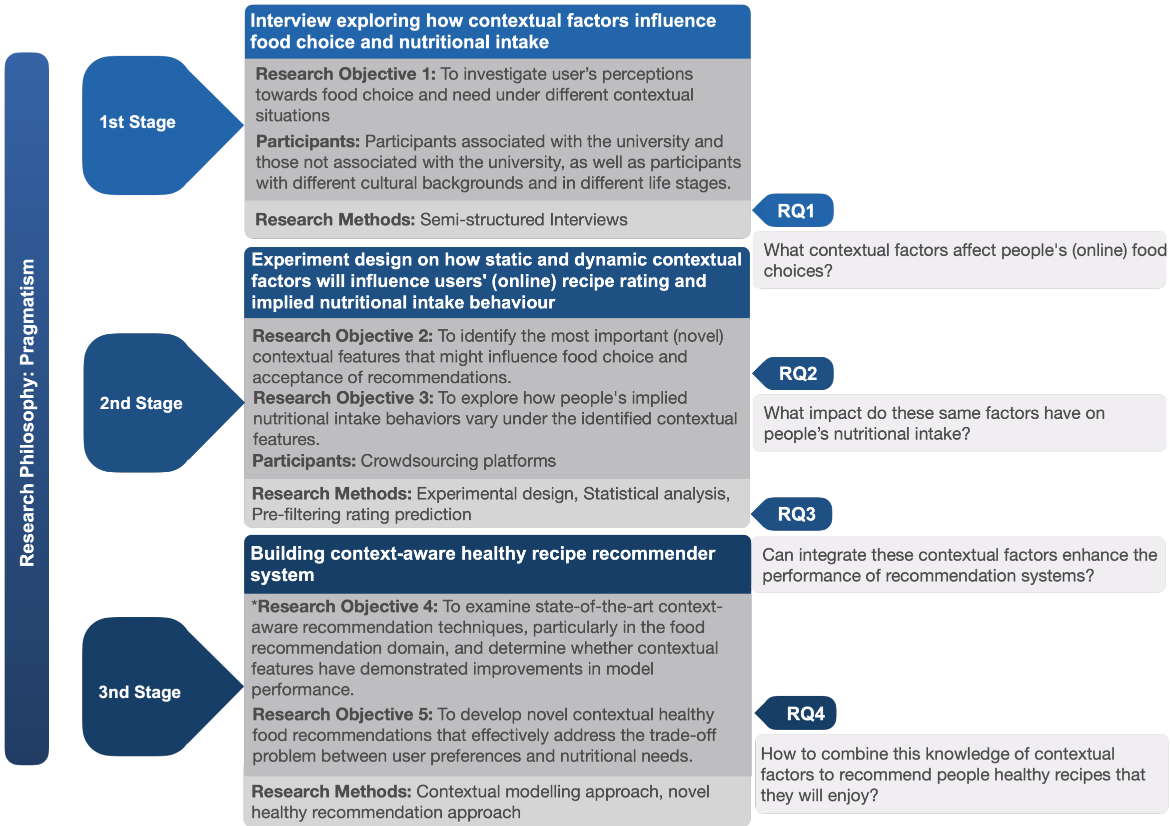


Figure 3.1: *Three stage research design demonstration*

were considered for recruitment through the popular crowdsourcing platform Prolific. The survey was designed using Qualtrics. Each participant was randomly assigned to one of seven contextual scenario groups (selected based on the findings from the semi-structured interviews, which presents in Chapter 4) or a control group with no specific context. Each participant rated 30 recipes from a pool of 75 using a Likert scale from 1 (strongly dislike) to 5 (strongly like). Detailed information regarding the selection of the 75 recipes and the survey workflow design is discussed in Section 3.4. By creating simulated contextual scenarios and collecting participant's explicit contextual rating data, which enable to statistically analysis whether individual's food choices, and their subsequent nutritional intake behaviour varied under different contextual situations. Additionally, it provides a unique dataset for further studying and developing context-aware recommender systems and healthy recommendation algorithms.

In the third stage, the investigation focused on whether integrating contextual features would enhance recommender system performance, and identifying the optimal feature combinations for developing a context-aware rating prediction model, then addressing the trade-off challenges in providing healthy recommendations. Popular embedding methods were employed to extract multimodal feature sets in an effort to optimize model performance, which included TF-IDF embedding, BERT embedding, GloVe embedding, recipe cooking methods matching, and image feature extraction. These features encompassed user demographic features (e.g., age, gender, home country, physical activity levels), recipe content features (e.g., recipe category, nutritional value, embedded ingredients, and cooking directions), and recipe image features. To further enhance model performance, a novel feature engineering approach was implemented to create contextual mediator features. Given that participants' explicit ratings were collected during the experimental study, the rating prediction model was framed as a regression problem. Three commonly used machine learning algorithms and a deep learning algorithm were employed to develop the rating prediction model: XGBoost, Ridge Regression, Support Vector Regression (SVR), and a Multilayer Perceptron (MLP). To achieve optimal predictive accuracy, a two-stage model was developed, integrating decomposed features from Singular Value Decomposition (SVD) and Non-negative Matrix Factorization (NMF) to replace the label-encoded user and item IDs. To address trade-off challenges and enhance the provision of healthy recommendations, a weighted contextual healthy recommendation equation was introduced, accompanied by health evaluation metrics.

The following section details the research methods employed for each phase of the study, along with justifications for why they are most appropriate for this research.

3.3 Research Methods Phase 1: Semi-Structured Interviews

Interviews are widely employed in qualitative research due to its ability to provide in-depth and richer data, and facilitating a deeper understanding of the participants' experiences and perspectives (Kvale & Brinkmann, 2009). There are typically three forms of interviews, including *structured*, *semi-structured* and *unstructured interview* (Bryman, 2016). In *structured interviews*, the interviewer maintains full control over the process, with interviewees answering a predetermined set of questions. This type of interview often employed in quantitative research, as it enables the collection of numerical data that can be systematically organized and presented in tables for analysis (Bryman, 2016; Patton, 2014). *Semi-structured interviews* have a more flexible structure, allowing greater space for the interviewees to participate and contribute. While there is typically a set of prepared questions, referred to as an "interview guide", the interviewer is not required to follow them in a strict order. This flexibility enables the interviewer to adapt the flow of the conversation, ask additional questions based on the interviewee's responses, and explore emerging issues naturally during the discussion (Kvale & Brinkmann, 2009). According to Creswell & Poth (2016), semi-structured interviews are particularly useful for generating new insights, especially when the research aims to explore complex behaviours, experiences, or new phenomena that may not yet be fully understood. *Unstructured interviews* have the most flexible form, also known as an "informal interview," allows the interviewee, often considered an expert in their domain, to guide the conversation. This format enables the interviewer to gain new insights and knowledge from the interviewee (Bryman, 2016). However, unstructured interviews can be challenging to analyse due to their variability and lack of consistency. Additionally, they also require highly skilled interviewers to ensure the conversation focused and relevant to the research objectives (Patton, 2014).

Semi-structured interviews were selected as the principal research method due to several advantages. Firstly, this approach provided a balance between structure and flexibility, enabling the exploration of predetermined contextual factors affecting eating and nutritional intake behaviour while allowing for adaptation to novel insights that emerged during the conversation (Kvale & Brinkmann, 2009). Secondly, it encouraged a more natural and relaxed interaction, prompting participants to share their experiences and perspectives more openly and thoroughly (Patton, 2014). Given that the impact of contextual factors is often overlooked, semi-structured interviews offered a flexible approach that supported participants in recalling details from their daily lives. The use of an interview guide also ensured consistency across different interviews. Lastly, in comparison to diary studies and observation, semi-structured interviews were more feasible, being less costly and time-consuming. This method also allows for the collection of more detailed and realistic information, thereby facilitating a comprehensive understanding of how eating habits are influenced by various factors and enabling the identification of the most influential ones.

3.3.1 Participant Recruitment and Sampling Strategy

Participant recruitment for this study was carried out through the University mailing list and Twitter advertisements. The objective was to enlist participants from various groups without imposing any specific demographic restrictions while still avoiding a homogenous participant pool where all individuals originate from the same countries, regions, or share identical occupations. Since the study is a preliminary step toward developing context-aware healthy food recommender systems, it is crucial to include individuals from various demographic backgrounds. This approach allows for the examination of whether people from different demographic groups are jointly influenced by contextual information on their food choices, eating habits, and nutritional intake behaviours (Kearney, 2010). Consequently, restrictions based on age, occupation, or education were deemed unnecessary for this study. Since all the interviews were conducted in English, participants were required to possess basic English communication skills.

Target participants received an invitation email included a brief introduction to the interview structure, a participant information sheet, a consent form, and a pre-interview questionnaire (see Appendix C). To ensure diversity among participants, the aim was to have no more than five participants with homogeneous characteristics. The pre-interview questionnaire was used not only to collect demographic information, food preferences, cooking and recipe searching experience but also to provide an overview of participants' basic information, such as gender and ethnic origin. This preliminary data facilitated the filtering and selection process, ensuring a diverse participant pool. Once participants expressed their willingness to participate in the research, their demographic information (particularly gender, occupation, and ethnic origin) was reviewed. Ultimately, all prospective participants were included after this screening process, as none of the participants were rejected based on their demographic information.

Regarding the number of participants, 12 to 15 were considered, according to Guest et al. (2006) and Francis et al. (2010), for most qualitative studies, the data saturation often occurs within the first 12 to 15 interviews. In this study, data saturation strategy is employed to determine when enough data has been collected to adequately explore the research questions and themes (Mason et al., 2010). Recruitment of participants continued until data saturation was achieved, as indicated by the repetition of information and the emergence of consistent themes across the semi-structured interviews. This approach not only enhances the reliability of the findings but also ensures that the diverse perspectives of the participants are thoroughly represented. Monitoring for data saturation involved regular review and analysis of the interview transcripts to identify when no new significant information was being uncovered (Guest et al., 2006).

Since there were no specific restrictions for participants, no particular sampling strategy was employed in this study. Instead, snowball sampling was utilised, leveraging the social networks of existing participants to identify and recruit additional participants, thereby streamlining the recruitment process. This method proved effective for this study by fostering a trusting, rapport-building, and relaxed atmosphere between the interviewer and interviewee, which stimulated and encouraged participants to share details of their daily lives. Recommendations from known individuals may further enhance participation (Biernacki & Waldorf, 1981).

To create a more participant-centred approach and enhance the quality and inclusiveness of the research (Bauman, 2015), participants were given the option to attend interviews either face-to-face or online. Ten interviews were conducted face-to-face, while four were conducted online via Google Meet. Only audio was recorded throughout the interviews. The primary recording was done using MacBook recording software due to its ease of use, with iPhone recording software serving as a backup to ensure data integrity in case of interruptions. Due to the semi-structured interview process, varying speaking speeds among participants, and differing levels of detail in responses, the duration of the interviews varied between 32 minutes and 1 hour and 8 minutes. There were no noticeable differences in duration between face-to-face and online interviews, indicating that both methods provided comparable quality and effectiveness.

3.3.2 Interview Piloting

Before the formal data collection, conducting pilot interviews is crucial for testing and refining the interview guide. This process helps identify potential issues with the flow or structure of the interview, allowing for necessary adjustments before the main study commences (Turner III, 2010). The design of the interview followed the Interview Protocol Refinement (IPR) framework proposed by Castillo-Montoya (2016). This framework is designed to enhance the reliability and validity of interview protocols by ensuring that questions are aligned with the research questions, logically ordered, and capable of eliciting rich, detailed responses. The IPR framework consists of four phases:

- *Ensuring interview questions align with research questions.* The design of the interview questions was informed by a comprehensive review of literature on recommender systems and the psychology of food science. The necessity of the study was confirmed. As developing context-aware recommender systems that balance users' previous diary preferences and nutritional needs requires identifying potential influential contextual factors from the user perspective. The interview questions were designed to explore these relative and novel contextual factors and address the research questions.

- *Constructing an inquiry-based conversation.* The interview questions were carefully developed and validated, deliberately articulated in everyday language to avoid the use of technical terminology that might confuse participants. Clarifications were available for any terms that participants found unclear during the interview process. The questions were designed to be open-ended to encourage participants to elaborate on their personal experiences. Additionally, a detailed script was prepared to guide the interviewer and facilitate a natural conversational style. For key interview questions, a series of follow-up questions and prompts were also devised to ensure comprehensive data collection.

- *Receiving feedback on interview protocols.* The interview questions were continuously refined through discussions with the supervision team and consultations with an experienced qualitative research fellow. Additionally, the interview questions were further improved based on feedback obtained from pilot interviews conducted with friends and research fellows.
- *Piloting the interview protocol.* After completing the previous three phases, pilot interviews were conducted to assess the effectiveness of the interview questions. The aim was to determine whether the questions were easy to understand and capable of encouraging participants to share more detailed experiences rather than providing simple yes or no answers. During the pilot interviews, the researcher took notes on potential improvements to enhance the acquisition of information relevant to answering the research questions.

Before confirming the preliminary interview questions for the pilot interviews, the questions were revised and improved three times through continuous discussions with the supervision team to ensure that each section of the interview corresponded to the respective research questions. In the final version, four sections of the interview process were established, as detailed in the following section. Six pilot interviews were conducted to test this version. The basic demographic information is shown in Table 3.1. Overall, the participants in the pilot interview provided positive feedback on their interview experience. Based on their input, warm-up questions were added, such as "Can you recall what you ate for your most recent meal?" Additionally, the order of the interview questions was slightly modified. If any participants were confused about the meaning of contextual factors, explanations were provided immediately during the interview, with the researcher briefly listing a few contextual factors to help participants recall relevant past experiences. The final stimulated scenario section was also revised and more carefully designed. After pilot interviews, the researcher provided a summary of the findings and continued to discuss refinements with the supervision team, until the finalised interview guide established.

3.3.3 Formal Interview Process

After receiving feedback from the supervision team and pilot interview participants, four sections of the interview process were confirmed, and the detailed interview guide for each section was finalised. The complete interview guide can be found in Appendix D. To enhance the robustness of the research, the interviews commenced by addressing real-world eating behaviours, thereby ensuring a grounded foundation. Subsequently, a more focused exploration of participants' online eating and recipe-searching habits was conducted.

Participants were given the opportunity to express their experiences regarding food decision-making and their general attitudes toward nutritional information. More importantly, contextual scenarios were provided to enhance participants' immersion, helping them better recall or imagine their experiences under different situations. The study received ethical approval from the University of Sheffield, the approved document has been attached in Appendix A.

Before starting the interview, an overview of the interview's aim was provided, along with details on the number of sections and the topics to be discussed in each section. Additionally, participants were reminded that the interview would be audio recorded. All participants had signed the consent form and completed the demographic questionnaire prior to the formal interview.

In the initial phase of the interview, questions were designed to be broad and open-ended. Participants were asked to recall their recent meals and share their favourite foods, thereby creating a conducive atmosphere for the interview, facilitating participant engagement, and making participants comfortable discussing the subsequent questions. Then, the key questions related to contextual eating behaviours began with participants discussing their general food decision-making processes and whether and how their family's eating behaviours impact their current eating habits. They were encouraged to express their opinions, drawing from their daily life experiences, on what contextual factors might influence their eating patterns. If participants struggled to identify these factors initially, reminders based on contextual factors studied in previous research, such as seasons and temperatures, were provided. Additionally, participants were prompted to explore whether any novel factors played a role in shaping their food preferences based on their individual experiences.

In the second part of the interview, participants' attitudes regarding nutritional information were elicited. Specifically, they were asked to express whether they felt they received sufficient guidance on nutritional information in their daily lives and to what extent they cared about this information. More importantly, participants were asked whether they had ever declined food choices based on their nutritional profiles and how factors such as nutritional content (e.g., fat, sugar, calories) and food-related aspects (e.g., ingredients, cooking methods, preparation time) impacted their food decision-making process.

In the third part of the interview, participants' online recipe searching and selection behaviours were investigated. This included examining when and where they typically search for recipes online, as well as how online recipes influence their cooking experiences. Addition-

Table 3.1: *Pilot interview participant demographic information*

Code	Gender	Occupation
PP1	Male	PhD student
PP2	Female	Postdoc
PP3	Female	PhD student
PP4	Female	PhD student
PP5	Female	PhD student
PP6	Male	Lecturer

ally, participants were asked about their expectations for recommended online recipes.

Through this interview study, there is a broader intention to guide the experimental design for the second stage, particularly in collecting real user ratings under various contextual scenarios. This is why during the final section of the interview, participants were invited to engage in simulated contextual scenario activities. The aim of these activities was to assess whether participants could recall or imagine detailed food expectations and determine if they knew what they wanted to eat under specific contextual scenarios. To accomplish this and mitigate biases (Orne, 2017), efforts were made to ensure that participants were unaware of the study's expectations, thereby preventing them from answering the questions deliberately. Participants were randomly and unknowingly divided into two groups (Group A and Group B) without prior knowledge of the scenarios. Each group was presented with four simulated contextual scenarios, with the scenarios between the two groups deliberately designed to be contrasting and distinct. In each scenario, participants were asked to recall or imagine a similar past experience and express their willingness to cook and order food.

This design ensured the quality of data collection and minimise potential bias. In Group A, participants were presented with simulated scenarios related to a hot summer's day, a busy workday, being in a joyful mood, and post-physical exercise. On the other hand, participants in Group B encountered simulated scenarios related to a cold winter's day, a leisurely weekend, feelings of sadness, and relaxed moments, see Table 3.2). Before the end of the interview, participants were also given the chance to talk about their feelings, comments and suggestions related to the conducted interview. Interviews varied from 32 minutes to 1 hour and 8 minutes; however, the interview structure was essentially identical for all participants.

Table 3.2: *Simulated scenarios content designed for each group*

No.	Group A	Group B
1	During a hot summer's day	During a cold winter's day
2	A busy and stressful day at work	A leisurely weekend
3	Very happy and in celebratory mood	Received bad news and feeling sad
4	After completing tiring physical activities	Relaxing at home

Fourteen participants were invited to take part in semi-structured interviews; their demographic information is detailed in Table 3.3. The participants exhibited diversity in terms of age, occupation, and ethnic origin. The gender distribution among the participants was nearly equal, with six female and eight male participants. The participants' ages ranged from 25 to 51 years. Their occupations spanned a wide range, including individuals affiliated with the university, such as undergraduate, master's, and PhD students, as well as researchers and professors. Participants working outside the university were also included, with professions such as librarian, receptionist, and business strategist represented. The majority of participants were of Asian ethnic origin, but the group also included individuals from North African, White British, Caucasian, and Latin backgrounds. The physical activity (PA) levels collected during the interview were based on participants' self-reported assessments, ranging from 1 (very inactive) to 5 (very active). No specific guidelines were provided to participants

Table 3.3: *Demonstration of participant demographic information*

Code	Age	Gender	Ethnic origin	Home country	Occupation	PA level	SCS group
P1	36	Male	North African	Tunisia	Postdoctoral researcher	3	A
P2	26	Male	White British	UK	PhD student	5	A
P3	27	Male	Caucasian	Germany	PhD student	4	B
P4	25	Male	Chinese	China	PhD student	4	B
P5	25	Female	Chinese	China	Master student	2	A
P6	29	Male	Caucasian	Spain	Postdoctoral researcher	3	B
P7	53	Male	White	USA	Postdoctoral researcher	4	B
P8	30	Female	Asian	China	PhD student	4	A
P9	25	Female	Chinese	China	Undergraduate student	3	A
P10	29	Female	Islamic	Qatar	Librarian	4	B
P11	51	Male	European/White	Canada	Professor	4	A
P12	30	Male	White mixed	Yemen	Receptionist	4	B
P13	35	Female	White British	UK	Librarian	4	A
P14	42	Female	Latin	Mexico	Business strategist	5	B

Note: "PA level" refers to Physical Activity level, while "SCS group" stands for Simulated Contextual Scenarios group.

for these ratings. It was observed that levels 3 and 4 were the most commonly reported, with only 2 participants rating themselves as level 5, and just 1 participant reporting a level 2. Additionally, data on participants' hobbies, and the frequency of using cookbooks or searching for recipes online were also collected for subsequent analysis. All participants had prior experience searching for recipes online, either through websites or mobile applications.

3.3.4 Interview Transcription

Transcribing interviews is an essential step in qualitative research, as it converts audio recordings into written text, facilitating detailed analysis and interpretation. This process ensures that no valuable information is lost and allows researchers to systematically code and categorize data to identify patterns and themes (Bryman, 2016; Mack, 2005). Transcription also enhances the accuracy and reliability of the analysis by providing a consistent, verbatim record of participants' responses. Additionally, it enables researchers to engage more deeply with the data, promoting familiarity with the content and supporting the development of insightful conclusions (Lapadat & Lindsay, 1999).

In this study, all interviews were transcribed manually. Although transcription is time-consuming, it enables the researcher to become thoroughly familiar with the interview content. The transcription process involved listening to each recording and transcribing the content, followed by a second review of the recording to ensure the transcription's accuracy and clarity. Halcomb & Davidson (2006) emphasize that this meticulous approach helps ensure that transcriptions are both clear and correct, providing a reliable foundation for subsequent data analysis.

The researcher enriched the transcriptions by annotating the text with various notes and conventions, such as exclamation marks (!), question marks (?), bold text and red colour text, to highlight significant points. Additionally, the researcher noted speech fillers used by participants, including phrases like "Yeah," "While,", "haha" and "Uhm," to capture the natural flow and nuances of the conversation. This detailed annotation process enhanced the subsequent analysis, providing deeper insights into the data and contributing to a more meaningful interpretation and presentation of the findings.

Following the completion of several pilot studies and iterative refinements to the interview process, it was determined that the protocol was sufficiently robust. To ensure consistency and comparability across all formal interviews, no further modifications were made to the interview process. Transcriptions were completed after each interview to facilitate later analysis (Gibbs, 2018).

3.3.5 Interview Data Analysis

In this study, the interview data were analysed thematically to discover and explore new insights. The primary focus was to understand how contextual factors influence individuals' perceptions, experiences, and behaviours regarding their eating and nutritional intake, as well as identified which are the most influential ones. Thematic analysis, as argued by Braun & Clarke (2006), is particularly well-suited for exploratory research as it allows for the examination of complex phenomena without the need for a predefined hypothesis. It is "a method for identifying, analysing and reporting patterns (themes) within data" (Braun & Clarke, 2006). This approach is effective in identifying and describing the subjective experiences of individuals, facilitating a deeper understanding of the factors that shape their behaviours (Braun & Clarke, 2012; Nowell et al., 2017). Moreover, thematic analysis is relatively straightforward to implement, and emphasises participant voices by highlighting the themes that emerge directly from the data (Guest et al., 2012).

In thematic analysis, codes can be generated using either an inductive or deductive approach. An inductive approach involves developing codes directly from the data, without preconceived notions or hypotheses. This method allows themes to emerge organically, capturing novel insights and patterns that reflect participants' experiences and perspectives. It is particularly useful in exploratory studies, where the goal is to uncover unexpected themes and develop new insights (Braun & Clarke, 2006). Conversely, a deductive approach involves creating codes based on existing theories or prior research. This method applies a specific theoretical framework or set of concepts to analyse the data, ensuring that the analysis is guided by pre-existing knowledge (Fereday & Muir-Cochrane, 2006). An inductive coding approach was adopted in this study, which allowing themes to emerge organically without preconceived notions, enabling an iterative process of reading through the data to identify influential contextual and food-based factors.

The collected interview data were analysed using the six-step thematic analysis procedure proposed by Braun & Clarke (2012), which includes *familiarization with the data*, *generating initial codes*, *searching for themes*, *reviewing themes*, *defining and naming themes*, and *producing the final report*. It is worth acknowledging that the analysis steps were applied in a non-linear manner, allowing for movement back and forth between these steps to facilitate a more flexible analysis of the data.

- ***Familiarization with the data***

  In this study, interview transcription was conducted by carefully listening to the interview recordings. After completing the initial transcription, the recordings were reviewed twice to enhance accuracy and ensure the reliability of the transcription. The transcription process was carried out using Microsoft Word on a MacBook Pro. Throughout this process, familiarity with the interview content increased significantly. A thorough read-through of the transcriptions further deepened engagement with the data. As the transcription was completed, important content was highlighted and marked with different annotations, identifying potential codes and themes for subsequent analysis.

- ***Generating initial codes***

  Once complete the transcription, and decided to extract code inductively. The initial codes were start generated using NVivo software, which offers seamless integration with Microsoft Word. It also makes it easy to combine and organize codes, allowing for the efficient generation of themes and providing a streamlined approach to data analysis. At the early stage, some initial codes were iterative, as continuing to analysis, the new codes emerge. During the last few interviews, the interview start to reach the data saturation as no new codes were added. The data coding process commenced with the development of an initial set of codes derived from the research and interview questions. For example, childhood experiences, social factors, and emotional factors were extracted as initial codes related to how contextual factors influence eating behaviours, based on participants' expressions of their past habits and experiences. Preferences for ingredients, cooking complexity, and the availability of cooking equipment were also extracted as initial codes regarding their influence on recipe choices. Within each section of the data, more granular and detailed codes were identified to enrich the initial coding framework and uncover new insights from the participants. The initial codes and subcodes were developed through meticulous and repeated examination of the interview transcripts. An example of the coding and data extraction process is presented in Table 3.4. Each participant was assigned a unique, identifiable colour in NVivo Home tab, facilitating easy reference to their demographic profile and enabling the exploration of potential demographic contextual factors.

- ***Searching for themes***

  The initially generated codes underwent continuous review. Codes reflecting similar content across different participants were combined to streamline the analysis process. Following this combination, the relationships among the codes were carefully examined to organize them into distinct themes. Any codes that did not fit within the established themes were temporarily placed in a "miscellaneous" category (Braun & Clarke, 2006), labelled "other codes group", to be revisited later.

- ***Reviewing themes***

  To ensure the dependability of the analysis, several meetings have been arranged with supervision team to discuss regarding the initial codes and themes. Whether the generated themes are reasonable and cover the individual codes, whether some codes need to move to other themes. An iterative process was employed to refine the codes, with each new transcript being cross-referenced to assess whether additional codes were needed. The final set of revised codes was organized into four primary themes, each accompanied

Table 3.4: *Data extraction and coding examples*

Code	Sub Code	Transcript
Strong preference for home cuisine	Variety in food selection	Absolutely my home homeland, because I think the food in XXX is like, as various from different regions to different like cities. The food is quite different and I can try like different food and select the one I preferred.
	More comfortable with familiar foods	But maybe it's my problem. I prefer food that is similar to the food from my country. It's more suitable for me.
Willingness to explore food		Yeah! I like eating Indian curry and some European food like seafood, Paella, and Vietnamese pho. I like to try food from different countries.
Emotional factors	Happy - Craving for sweets (possibly hormonal effects)	I believe that happens not only to me. I think many people, including me, when in a bad mood, would love to eat something sweet. There's some chemical effect in our brain, I think, that stimulates happiness.
	Stressed - Increased intake of sweets and snacks	Maybe if I am stressed, I eat more sweets on the side, but it would not necessarily affect what I choose for dinner. It would just be more snacks, typically, but I don't think it would impact my choice of dinner.

by sub-themes that aligned with the research questions.

- ***Defining and naming themes***

  During this stage, the codes and themes were reviewed once more. For each individual theme, the researcher and the supervision team worked collaboratively to construct a comprehensive "narrative" that contained each theme and its sub-themes. Table 3.5 presents the themes and sub-themes, along with their definitions as identified in this study.

- ***Producing the final report***

  In the final stage, the complex relationships between participants' food preferences and nutritional intake behaviours in relation to contextual factors were thoroughly elaborated and explained in a valid and comprehensive narrative. The specific food-based factors, such as ingredient preferences and cooking methods, influencing the food and recipe decision-making processes were also identified. The analysis was guided by data-driven findings, including data extracts, to provide a coherent, non-repetitive, and logical account of these narratives (Braun & Clarke, 2006). Detailed findings from the interviews are presented in result Chapter 4.

Table 3.5: *Themes and Sub-themes*

Themes	**Sub-themes**	**Description**
Both static and dynamic contextual factors could be influential	Eating behavior is strongly linked to childhood experiences and cultural influences	People's eating habits are deeply rooted in their early life experiences, particularly the types of food typically eaten in their family and home country.
	Individuals with strong personal goals exhibit notably distinct eating behaviours	People with specific personal goals, such as fitness targets, weight management, or dietary restrictions (e.g., health-related goals), tend to make more deliberate and goal-oriented food choices.
	Emotions and busyness are the most influential dynamic factors shaping food choices	Emotional states (such as happiness, sadness, boredom, or fatigue) and busyness (such as during a stressful weekday or a relaxing weekend) are the most influential dynamic factors that may constantly affect individual food choices and eating behavior.
	Seasons, after physical activities, and sustainability are also influential	Contextual factors such as the seasons, physical activities (e.g., post-exercise hunger), and concerns about sustainability may also significantly influence food choices.
Nutritional information may be undervalued, and contextual factors may reshape nutritional intake behavior	Nutritional profile is often overlooked	Unless individuals have specific personal health goals, the selection or rejection of a food product is typically not based on its nutritional profile. Most individuals acknowledge that nutritional information is important but often choose not to focus on it.
	Contextual factors such as family habits, stress, emotional fluctuations, and social factors influence nutritional intake	Factors such as family eating habits, increased stress levels, emotional fluctuations, and social influences may often override nutritional considerations, further shaping individuals' nutritional intake.
Serendipitous recipe searches and the iterative nature of recipe selection and real-life cooking intention	Serendipitous recipe searches	Recipe searches may happen unexpectedly, with individuals often discovering new recipes through unplanned exploration. Serendipitous discoveries can be influenced by various factors, such as browsing social media or watching TV.
	Choose recipe-first, buy ingredient-second approach	In the decision-making process involved in meal preparation, individuals typically select a recipe first and then purchase the necessary ingredients to prepare the dish.
	Recipe (online recipe) rejection often based on presentation and combination of ingredients.	The primary reasons for rejecting a recipe are its presentation (whether the recipe appears visually appealing) and the combination of ingredients (whether the ingredients seem familiar and desirable).

Continued on next page

Themes	Sub-themes	Description
Essential for CARS development	Recognising and incorporating emotional and busyness awareness into healthy food recommender systems	Emotional states (such as happiness or sadness) and the user's level of busyness (such as during a busy weekday or a relaxing weekend) are potentially critical contextual factors that could significantly improve the relevance and personalisation of recommendations in healthy FRS.
	Combining contextual features with multi-modal features in existing FRS	Combining contextual features with multimodal features in existing FRS could substantially enhance the likelihood of recommendations being accepted.

Notes: The final findings from the semi-structured interviews are reported in the Results Chapter (Chapter 4).

3.4 Research Methods Phase 2: Experimental Study

In the first stage of the research, qualitative methods were used to gain a comprehensive understanding of how contextual factors influence food choices and nutritional intake behaviours. During this stage, key potential influential contextual factors were identified, while individuals' cooking practices and recipe-searching habits were also explored. Based on the findings from the first stage of the interview analysis, emotions, busyness, seasons, and physical activities have been identified as the most potentially influential dynamic factors on people's eating and nutritional intake behaviours. The detailed research findings can be found in Chapter 4. To generalise and validate these findings, further quantitative research is needed, to examine the statistical relationship between contextual factors and individuals' recipe ratings and nutritional intake behaviours. These analyses would enrich the previous finding and further contribute to the development of more personalised healthy food recommender systems.

The aim of this second-stage study is to quantitatively investigate the impact of previously identified contextual factors on people's online recipe ratings and nutritional intake behaviours. Additionally, it seeks to gather explicit user feedback on recipe ratings across various contextual situations, enabling further statistical analysis and preliminary development of a recommender system rating prediction models. A between-subjects experimental design was selected as the primary method to achieve the research objectives. This section outlines the research design structure, beginning with an introduction and justification of the selected research approach (Section 3.4.1), followed by the process of building the recipe database (Sections 3.4.2 to 3.4.3). Then address the stimulus and manipulation methods (Section 3.4.4), followed by a discussion on the rationale and process behind the creation of the recipe database used to gather participants' ratings (Section 3.4.3). The survey design workflow (Section 3.4.5), participant recruitment strategies, and data quality control measures (Section 3.4.6) are explained. This section concludes with a discussion and justification of the chosen statistical model (Section 3.4.8) and the preliminary rating prediction model (Section 3.4.9).

3.4.1 Experimental Design

The primary aim of this second stage research is to quantitively examine how seven dynamic contextual factors, including a hot summer's day, a cold winter's day, happy emotion, sad emotion, a busy and stressful weekday, a relaxing weekend, and after physical activities affect people's recipe rating, and implied nutritional consumption behaviour. To achieve this, a between subjected experimental study has been designed.

In psychological and behavioural research, two fundamental types of experimental designs are widely used, named *within-subject designs* and *between-subject designs* (Charness et al., 2012). In a *within-subject design*, participants are exposed to all experimental conditions, allowing researchers to directly compare each participant's responses across different treatments (Gravetter et al., 2009). However, this approach carries heightened risk of demand characteristics, where participants, having experienced all conditions, may infer the study's purpose and alter their behaviour accordingly, potentially introducing significant bias into the findings (Mummolo & Peterson, 2019). In addition, carryover effects may be difficult to eliminate, as participants can become fatigued or learn something from an earlier condition that affects their performance in later ones (Myers & Hansen, 2006; Montoya, 2023). This creates challenges in deciding the order of presented conditions (Gravetter et al., 2009). In relation to this research, continuously and frequently asking participants to recall or imagine vastly different scenarios (as this study aims to measure and stimulate recipe rating behaviours under seven extremely different contextual situations) may easily trigger demand characteristics and carryover effects, as previously discussed. This might introduce inconsistencies in participants' responses and potentially affect the reliability of the data.

In contrast, a *between-subject design* may effectively mitigate the limitations inherent in a within-subject design and therefore adopted as the key experimental approach in this research. Between-subject design assigns different groups of participants to different experimental conditions, with each participant experiencing only one condition, and comparisons are made between different groups (Erlebacher, 1977; Keren, 2014). This approach naturally eliminates the risk of carryover and order effects (Field, 2024), as participants were only exposed to a single condition. Compared to being exposed to all conditions, participants are less likely to experience fatigue or confusion. They do not need to engage with a range of contrasting contextual scenarios, such as imagining or recalling a hot summer's day immediately followed by a cold winter's day, then switching between different emotional states, levels of busyness, and post-physical activity situations. This would also prevent participants from altering their behaviour based on their perceptions of the study's purpose (Simkus, 2023). But, precisely because participants are restricted to a single condition, necessitating the comparison of different groups, between-subject designs often require a significantly larger sample size to achieve reliable and valid results (Maxwell et al., 2017; Wickens & Keppel, 2004). With the rise of the internet and digital technologies, crowdsourcing platforms have made it relatively easy to address this challenge by efficiently connecting researchers and organizations with a large, diverse pool of participants or contributors from around the world (Goodman et al., 2013).

Prolific crowdsourcing platform has been chosen to conduct data collection in this second stage research. According to Peer et al. (2017), prolific provide better-target recruitment and higher-quality data. Compared with Amazon Mechanical Turk, prolific offer more flexible setting regarding participants demographic information, such as, limited participants from certain culture background ((Palan & Schitter, 2018). Since the recipes are sourced from the popular U.S. website *Allrecipes.com*, the cultural impact on recipe selection, as demonstrated by Zhang et al. (2020a, 2023), is an important consideration. British and American food share more similarities than Asian and American food. Therefore, to minimize the influence of cultural factors, the plan is to limit participants to those from the United Kingdom or the United States. Additionally, most users on Prolific are based in the UK or the US, which further meets the requirements of this research. Furthermore, the built-in measures that Prolific offer could potentially enhance data quality, such as attention checks and manipulation check questions. It also provides tools for researchers to exclude participants who do not meet specific criteria or who have provided low-quality data in previous studies.

To ensure efficient use of allocated resources and to avoid unnecessarily large samples or under-resourced studies, ***power analysis*** was conducted to estimate the required sample size prior to participant recruitment. This analysis ensures that the study achieves an optimal balance, selecting a sample size sufficient to detect meaningful effects while conserving resources (Cohen, 2013; Button et al., 2013). G*Power was utilised to estimate an appropriate sample size for the study, according to Faul et al. (2007). Drawing on data from previous experimental studies on personalisation and user behaviour Kim & Gambino (2016); Liu et al. (2022), the effect size was estimated at 0.23. The analysis indicated that a sample of 360 participants would be required to meet a 5% significance level with 90% statistical power.

In summary, a large-scale between-subjects user study was conducted in this second stage of research. Participants were recruited through the Prolific crowdsourcing platform with the aim of gathering explicit user ratings across various contextual scenarios from a created pool of 75 recipes. The process of selecting and building the recipe database is discussed in the following Section 3.4.2 and Section 3.4.3. Participants were randomly assigned to one of eight groups (seven simulated contexts plus a baseline "context-free" condition), exposed to materials simulating the given context, and then asked to rate recipes within that scenario. Each participant rated 30 recipes using a Likert scale ranging from 1 (strongly dislike) to 5 (strongly like). The survey instrument was designed using Qualtrics, with details regarding the stimulus, manipulation materials, and survey design workflow presented in Section 3.4.4 and 3.4.5.

3.4.2 Determining Recipe Health Levels and Database Creation

From the perspectives of research feasibility and cost-effectiveness, 75 recipes were carefully selected for participants to rate. These recipes were chosen to ensure a roughly balanced distribution across various factors, including healthiness, seasonality (i.e., winter and summer), and recipe categories (i.e., main dish, soup, salad, and dessert/snack). Additionally, both vegetarian and vegan options were included to provide a comprehensive sample. The pool of 75 recipes was sourced from Allrecipes.com, one of the most popular online recipe websites, which has been utilised as a data source by numerous researchers in the food recommendation domain (e.g., (Zhang et al., 2020a; Harvey et al., 2013; Rokicki et al., 2018)). Notably, the website offers a free licence for use in research purposes (Dotdashmeredith, 2023).

Since the ultimate goal of this research is to develop a healthy food recommender system, accurately determining the health level of each recipe is crucial. To ensure robust and authoritative assessments, three well-established nutritional standards were applied: the World Health Organization (WHO) guidelines (Who & Consultation, 2003), the U.S. Food and Drug Administration (FDA) standards (Fda, 2022), and the United Kingdom Food Standards Agency (FSA) criteria (Fsa, 2016). By utilising multiple standards, the research ensures a more comprehensive and credible health evaluation. These international standards have been widely adopted in food recommendation domain (Trattner & Elsweiler, 2017b).

The health level calculations were based on the specific rules provided by each of these standards. Allrecipes.com, the source of the recipes, provides full nutritional labels per serving for each recipe, with nutrition facts presented as percent daily values. Essential nutritional information such as calories, total fat, cholesterol, sodium, and total sugars is readily available on the website. The following section provides a detailed introduction to each international standard, along with a comprehensive explanation of the health level calculations.

3.4.2.1 WHO Standard

By 2020, chronic diseases were projected to account for nearly three-quarters of all global deaths, a significant concern given that many chronic diseases are largely preventable Who & Consultation (2003). In response, the World Health Organization (WHO) proposed population nutrient intake goals aimed at preventing diet-related chronic diseases, as outlined in Table 3.6. As WHO emphasises the importance of alleviating the global burden of chronic diseases, evidence shows that the impact is even more pronounced in developing countries, where the burden is effectively doubled. This makes the WHO guidelines relatively fundamental and adaptable for various national and regional bodies. According to Howard et al. (2012), seven key nutritional components—Fat, Saturates, Cholesterol, Sodium, Carbohydrate, Sugars, and Protein—have been identified as crucial for assessing the healthiness of each recipe based on WHO standards.

Utilising the nutritional information provided by Allrecipes.com, the total grams of these components were determined by summing all available nutritional information. Next, the percentage of each nutritional component was calculated to assess whether the recipe's nutrient values fall with the standard ranges set by WHO guidelines. If a specific nutrient value fell within the recommended range, the recipe was awarded a point; if not, no points were given. Therefore, a higher score indicates a healthier recipe.

Table 3.6: *WHO standard ranges of population nutrient intake goals*

Dietary factor	Goal (% of total energy, unless otherwise stated)
Total fat	15-30%
Saturated fatty acids	<10%
Polyunsaturated fatty acids (PUFAs)	6–10%
n-6 Polyunsaturated fatty acids (PUFAs)	5–8%
n-3 Polyunsaturated fatty acids (PUFAs)	1–2%
Trans fatty acids	<1%
Monounsaturated fatty acids (MUFAs)	By differencea
Total carbohydrate	55–75%
Free sugars	<10%
Protein	10–15%
Cholesterol	<300 mg per day
Sodium chloride (sodium)e	<5 g per day (<2 g per day)
Fruits and vegetables	>=400 g per day
Total dietary fibre	From foods
Non-starch polysaccharides (NSP)	From foods

3.4.2.2 FDA Standard

Since Allrecipes.com is one of the largest American recipe websites, each recipe's nutritional information is provided by ESHA Research, a well-known technical, nutrition, and regulatory support company based in the US (ESHAresearch, 2023). Therefore, the original nutrition labels provided by Allrecipes.com directly adhere to FDA standards. The general guidelines of the FDA standard are presented in Table 3.7. The health score is determined by comparing each recipe's Daily Value (DV) with the standard goals. A "T" is assigned to nutrients that meet the standard, corresponding to a score of "1," while an "F" is assigned to those that do not, meaning no score is added. Consequently, the direction of change in health levels according to the FDA is aligned with that of the WHO, meaning that as the score increases, the recipe is considered healthier.

Table 3.7: *FDA general guide to DV*

Nutrient	DV	Goal
Saturated Fat	20g	Less than
Sodium	2,300mg	Less than
Dietary Fiber	28g	At least
Added Sugars	50g	Less than
Vitamin D	20mcg	At least
Calcium	1,300mg	At least
Iron	18mg	At least
Potassium	4,700mg	At least

3.4.2.3 FSA Standard

The FSA is an independent department of UK Government, responsible for food safety, and labelling policy. The FSA standard here stands for the well-known traffic light food labelling policy, which is also known as front of pack (FoP) nutrition label guide (Fsa, 2016). Unlike WHO and FDA standard mainly regulate ideal daily nutrient intake, FSA standard developed a relatively more accuracy guidance on single product. Due to the primary goal in this study is to measure the health level of each recipe, so the FSA standard can be treated as most appropriate criteria. However, the calculation of the FSA health score is more intricate, as it measure the nutrition content by 100g/ml as shown in Table 3.8. Since the nutritional information provided by the Allrecipes website is based on daily value. The total weight of each recipe was first calculated based on its ingredient list, and this total was then divided by the number of servings to determine the weight per portion. Subsequently, the values for fats, saturated fats, sugar, and salt per 100g were computed. These nutrient values were then categorized as green, amber, or red based on their respective ranges. A nutrient value in the green range was assigned one point, in the amber range two points, and in the red range three points. In this scoring system, a higher overall score for a recipe indicates a lower level of healthiness.

Table 3.8: *Criteria for 100g of food (whether or not it is sold by volume)*

Text	**LOW**	**MEDIUM**	**HIGH**	
Colour code	**Green**	**Amber**	**Red**	
			>25% of RIs	**>30% of RIs**
Fat	<= 3.0g/100g	>3.0g to <=17.5g/100g	>17.5g/100g	>21g/portion
Saturates	<= 1.5g/100g	>1.5g to <=5.0g/100g	>5.0g/100g	>6.0g/portion
(Total) Sugars	<= 5.0g/100g	>5.0g to <= 22.5g/100g	>22.5g/100g	>27g/portion
Salt	<= 0.3g/100g	>0.3g to <=1.5g/100g	>1.5g/100g	>1.8g/portion

3.4.3 Database Building Process

The development of the recipe database is an ongoing process of continuous updates and iterations. This database was initially established by identifying distinct popular recipes for each season. To accomplish this, experiments were conducted using a widely recognized secondary dataset, previously collected by Elsweiler et al. (2015). Recipe seasonality was determined based on the popularity of each recipe (i.e., total number of 4 or 5 ratings) during each season. For example, the recipes received most 4 and 5 ratings during the months of December, January and February are that recipe's winter popularity. Only the rating bigger than 3 were considered as favoured of the recipe. After group the user rating data by season, 23 recipes have been selected and treated as the potentially most popular recipe during winter and summer (in this research, only two seasonal factors have been considered). Next, the database has been enriched by browsing the most popular recipe among each category.

The categories are automatically provided by Allrecipes website, such as, salad, soup, dessert, etc. For the selection of recipes at each category, the number of user ratings and the rating level are treated as the most important evaluation indicators, and as for those recipes that have received more than 1000 ratings and have a rating level higher than 4 stars were prioritised. This process intend to control the likelihood of each recipe, and ensure each of them could be in similarly level. Although consider people's food preference can be varied, the recipes that no one favoured could be avoided, in these cases, the influence of contextual factors can be more prominent. All the chosen recipes has been classified as four categories: *Main dish, Soup, Salad, and Dessert/Snack*, each recipe only belongs to single category, no overlap categories.

Recipes were identified as vegetarian and/or vegan through a manual analysis of their ingredients. The cooking complexity was also manually assessed, taking into account preparation time, cooking time, and the number of required ingredients. Since the primary objective is to assess the health level of individual recipes; however, established standards like those from the WHO or FDA define ideal daily nutritional intake. Given that the typical dietary pattern in Western societies often consists of three meals a day (Symons, 1991; Lhuissier et al., 2013), in this study the standard WHO and FDA recommended daily nutritional intake was divided by three to create two new features, named WHO adjusted (WHO_adj) and FDA adjusted (FDA_adj) health levels. The idea is trying to acquire an approximate indicator of the nutritional quality of an individual recipe for further experiment.

After the above-mentioned steps, the primary database includes 92 recipes. In order to balance the recipe health level based on WHO, FDA and FSA standard, as well as considering the feasibility of the study based on the funding obtained. The 75 recipe are reminded for the final recipe database. The complete process of database creation can be seen in the figure 3.2. The final determined 75 recipe database is demonstrated in Appendix Table F.1. Very few recipes achieve four green label (most healthy recipe) or four red label (most unhealthy recipe) according to FSA standard, and most recipes have amber and red label. The extremely unhealthy and healthy recipe is also rare.

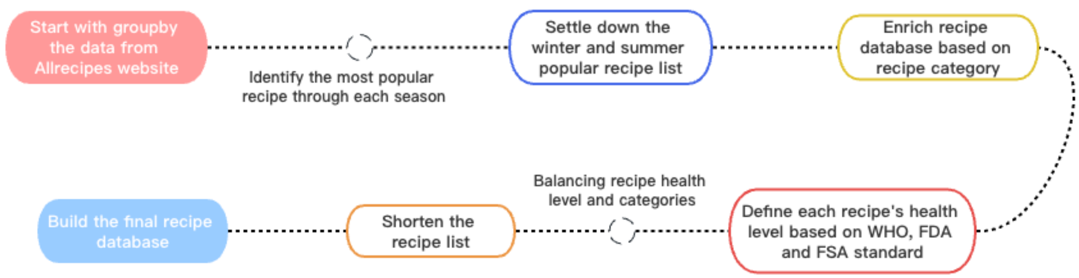


Figure 3.2: *Demonstration of database building workflow*

3.4.4 Stimuli

In the study, seven contextual situations were simulated, including hot summer's day (HSD), cold winter's day (CWD), happy emotion (H), sad emotion (S), busy weekday (B), relaxing weekend (R), after physical activities (APA), under which participants could rate recipes. To minimise the influence of real-life environmental factors and enhance the evocation of these simulated scenarios, inspiration was drawn from studies conducted by Imani & Montazer (2019) and van Strien et al. (2013). Choosing the appropriate methods for emotional and perceptual stimuli is critical.

There are various approaches to emotion elicitation, which can generally be categorized into *internal* and *external* methods based on the source of the stimuli (Hu et al., 2020). *Internal* emotion elicitation methods rely on participants recalling or imagining personal experiences under guided experimental instructions. However, due to individual differences in factors such as age, gender, habits, cultural background, childhood experiences, personality, and emotional perception, there are notable limitations to using internal stimuli for emotion induction (Hu et al., 2020). Specifically, Salas et al. (2012) highlighted that internal stimuli are more likely to evoke negative or mixed emotions, making it challenging to determine whether a specific emotion has been accurately triggered. Since this research aims to evoke two relatively extreme emotions, such as happiness and sadness, the external emotion stimuli method has been selected as the more appropriate approach (Salas et al., 2012; Hu et al., 2020). The *external* approach is particularly effective because it consistently evokes specific emotions and perceptions across different individuals. It primarily utilises source materials such as photos, music, or videos to elicit emotional responses. Among these options, videos are especially effective and immersive, particularly in controlled laboratory settings. Unlike images or music alone, videos provide a rich combination of visual and auditory stimuli, creating a multisensory experience that deeply engages participants, this abundance of sensory input makes videos the preferred choice (Hu et al., 2023; Somarathna et al., 2022; Hu et al., 2020). Furthermore, this method offers the advantage of standardising the emotional and perceptual experience across participants. By presenting the same stimuli to all individuals within a group, the emotional trigger remains consistent, minimizing variability that might arise from individual differences in interpreting or reacting to the stimuli. This allows for more reasonable comparisons between different experimental conditions.

Participants were presented with a 22-second video clip designed to engage their cognitive faculties and enhance their immersion in the assigned simulated contextual situation (Tennyson & Breuer, 2002). In total, seven videos were created by manually designing each one using the popular video platform Vimeo. Vimeo is an effective tool that provides high-quality video hosting, editing, and sharing capabilities. Its intuitive interface enables users to create and customise videos effortlessly, integrating advanced features such as text overlays, animations, and transitions (Walker & Boyer, 2018). Each video included prompts and was matched with corresponding elements of the contextual scenario. For instance, the video for a hot summer day featured a beautiful sunny day, a thermometer displaying high temperatures, and a person sweating, accompanied by prompts that stated:"*Imagine today is a hot summer's day, with the sun beating down, the temperature is very high, and you are sweating a lot. What would you like to eat? Please rate the recipes below.*" In contrast, the video for a cold winter day portrayed a wintry scene with heavy snowfall, roads blanketed in thick snow, and strong gusts of wind, accompanied by prompts that read:"*Walking into a winter wonderland... It is a very cold winter's day, the temperature is very low, and it's freezing outside. What would you like to eat? Please rate the recipes below.*" The displays and corresponding prompts of all seven videos are shown in Table 3.9. The background music for all videos was carefully selected to match the contextual scenarios. For example, as for happy contexts, music with an upbeat tempo, ascending melodic lines, and simple, consonant harmonies was chosen. This approach is supported by Gabrielsson & Lindström (2010), who highlights the significant contribution of musical structure to the perception of emotions. Conversely, for sad contexts, music in a minor key with a slow tempo and descending melodic lines was used. According to Hevner (1935), listeners consistently rate major key music as more joyful and minor key music as more sombre. For anyone interested in this research, all stimulus videos can be accessed through the following Google Drive link: Stimuli videos for second stage research link [1].

To further immerse participants in their assigned contextual scenario, recipe templates (see Figure 3.3) were designed to align with each scenario. The number of special elements, such as the sun for a hot summer day or weightlifting for after physical activities, was consistent across templates, and the sizes of these elements were nearly identical when inserted into the background. For reference, specific recipe examples can be found in Figure 3.3. Notably, participants in the "context-free" control group were not shown a video, and the recipe template remained basic, without any contextual elements stickers.

3.4.5 Survey Design

Upon the completion of the recipe database, the selection of an appropriate data collection platform becomes a critical consideration. Qualtrics, widely recognised as a leading web-based survey creation and data collection tool, is extensively cited in academic literature (Zikmund et al., 2013; Brunson, 2008; Boas et al., 2020; Molnar, 2019). Its flexible, user-centric design, coupled with robust advanced features, and customizable survey logic, provides substantial support for this research. Given the requirement for a substantial participant pool to facilitate the statistical analysis of recipe rating behaviours across various contextual scenarios, Prolific has been utilised for participant recruitment and research advertisement,

[1] Stimuli videos for second stage research link. Available at: `https://drive.google.com/drive/folders/1BjbuK90OZdQ2mDn58fRumaibwGeFC4Zj?usp=drive_link`

Figure 3.3: *Demonstration of recipe template under each contextual scenario (taking "Apple Pie By Grandma Ople" as an example recipe). Note that the example on the bottom-right is the baseline, context-free condition.*

Table 3.9: *Video display and corresponding prompts*

	Video display	Corresponding prompts
Hot summer's day	Featured a beautiful sunny day, a thermometer displaying high temperatures, and a person sweating.	Imagine today is a hot summer's day, with the sun beating down, the temperature is very high, and you are sweating a lot. What would you like to eat? Please rate the recipes below.
Cold winter's day	Portrayed a wintery scene with heavy snowfall, roads blanketed in thick snow, and strong gusts of wind	Walking into a winter wonderland... Imagine it is a very cold winter's day, the temperature is very low, and it's freezing outside. What would you like to eat? Please rate the recipes below.
Happy emotion	Featured an achievement, such as the growth of a plant, accompanied by cheer and an encouraging posture. The video also showcased present (gift-like) day scenery and a girl with a big smile, radiating joy and positivity.	Imagine you feel very happy and full of excitement because of something that has happened, you may have succeeded, you may have received a suprise present, you feel very cheerful. What would you like to eat? Please rate the recipes below.
Sad emotion	It featured a moment of walking with frustration, sitting on the floor with your head in your hands, deep in thought, accompanied by a sad expression, with rain and a gloomy atmosphere.	Imagine you feel sad about something, maybe you just received some bad news, or you recall a bad memory, you feel very gloomy! What would you like to eat? Please rate the recipes below.
Busy work day	It portrayed a busy working environment with a calendar packed with meetings, a long list of work tasks, and an individual with an overwhelmed and stressed expression.	Imagine you have a busy day at work, you have been quite overwhelemd, after a whole day of meetings, tomorrow you still need to work on a project report and the deadline is quite close. What would you like to eat? Please rate the recipes below.
Relaxing weekend	It showcased a relaxing time, staying at home with cats, lying on the bed to rest, or enjoying personal time, such as going camping.	Imagine you are having a relaxing and easy weekend, you have a light schedule, time to yourself. What would you like to eat? Please rate the recipes below.
After physical activities	It featured scenes of running, exercising, practicing yoga, and exuding confidence in front of a mirror.	Imagine you have just done some exhausting exercises, you burned a lot of calories, you are tired, but you feel great! What would you like to eat? Please rate the recipes below.

as it has been identified as the most suitable platform for data collection, as discussed in the previous section (see Section 3.4.1).

The comprehensive survey workflow is illustrated in Figure 3.4. When participants express interest in the survey on Prolific and click to participate, they are redirected to the Qualtrics platform. The Prolific ID of each participant is automatically populated. Before the survey begins, participants are asked to review the information sheet and complete a consent form. Following this, demographic information is collected, including essential details such as age, gender, and ethnic origin. This demographic data provides valuable insights into how participants' backgrounds might influence their eating habits and nutritional intake behaviours. Each demographic question is set as a mandatory response, with a "prefer not to say" option available to ensure participants do not feel pressured to disclose information they are uncomfortable sharing.

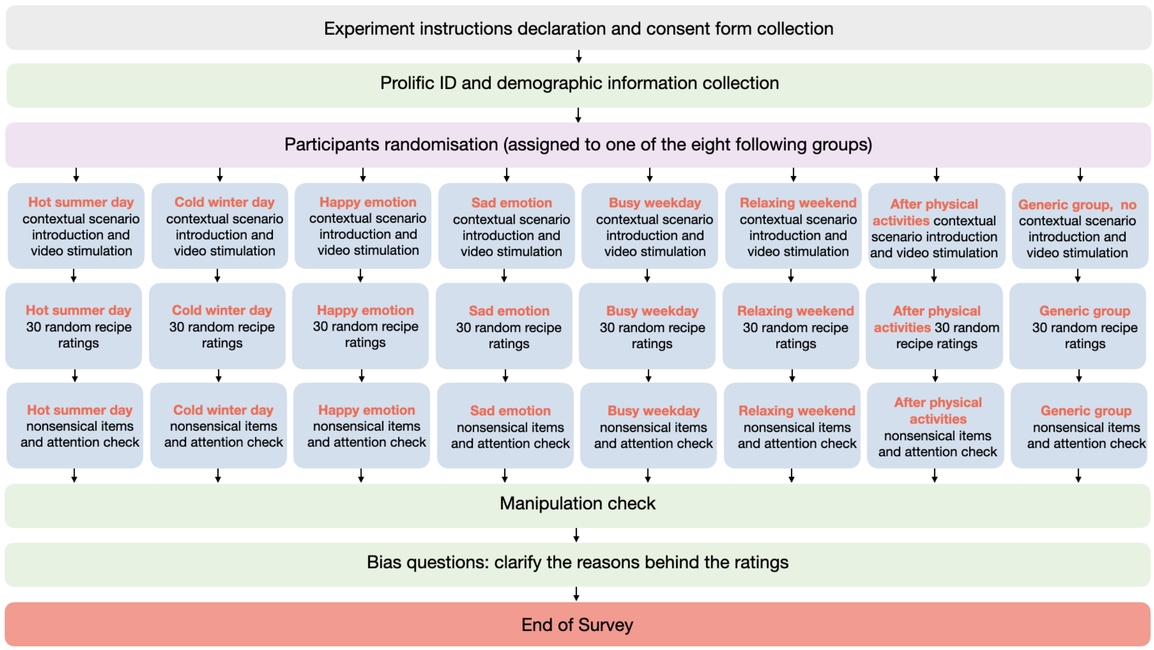


Figure 3.4: *Experiment workflow chart*

After completing the demographic questions block, participants encountered the recipe rating block. In this study, participants were randomly assigned to a certain group before providing their ratings. As discussed above, participants in the contextual scenario groups were instructed to watch a 22-second video clip designed to immerse them in that context. Each participant was tasked with rating 30 recipes within their assigned contextual situation block. Finally, participants were asked two questions regarding their reasons for assigning high or low ratings to each recipe.

To ensure the integrity and quality of the data collected, both manipulation and attention checks were implemented to confirm that participants were adhering to the prescribed procedures and maintaining appropriate levels of attention throughout the study. The manipulation check consisted of a question requiring participants to identify the theme of the video they viewed at the beginning of the questionnaire (e.g., "hot summer day"). This served to verify that participants were engaged with the content and had correctly processed the stimuli.

The attention checks were twofold. The first involved presenting a nonsensical recipe, which included ingredients such as "5 stones" and "3 cups of sand," designed to elicit a negative response from participants, thereby confirming their attentiveness. The second type of attention check involved specific recipes where participants were explicitly instructed to select a particular response, ensuring that they were carefully following instructions. These attention check questions were administered after participants had completed their ratings for all 30 recipes. While the textual content of the attention-check recipes remained constant, the background images were varied to align with the different contextual scenarios presented in the study, thereby maintaining visual consistency and reducing the risk of participants noticing the repetitive nature of the checks. An example of the attention check recipe is shown in Figure 3.5.

Figure 3.5: *Attention check example recipe (taking "hot summer day contextual scenario" as an example).*

3.4.6 Participant Recruitment and Data Quality Control

Participants were recruited exclusively through Prolific between May 2023 and July 2023. Each participant was compensated £1.82, in alignment with the current living wage standards. Ethical approval for this study has been granted by the University of Sheffield, as detailed in Appendix B. A total of 428 participants were initially recruited for this study. However, those who failed the manipulation check (19 participants, 4.4%) and the nonsensical item attention check (12 participants, 2.8%) were excluded from the analysis. After thorough data inspection and cleaning, valid data from 397 participants were retained for further analysis. The distribution of participants' demographic information is illustrated in Figure 3.6.

Among these participants, 212 (53.4%) identified as male and 177 (44.6%) as female, 5 (1.3%) were non-binary or gender diverse, and the remaining 3 (0.7%) preferred not to disclose their gender. The largest age group was 25-34, representing 30.7% of participants, followed by the 35-44 age group at 25.8%, the 45-54 age group at 19.4%, the 18-24 and 55-64 age group with a similar proportion at 10.8% and 10.3%, respectively. Participants aged 65-74 accounted for 3.3%, with only one participant aged 75 or above. The majority of participants were from the UK and the US. Regarding ethnic origin, White participants dominated with 329 (82.9%), followed by Asian or Asian British at 34 (8.6%). The remainder consisted of participants identifying as Black, Black British, Caribbean, or African (16, 4.0%), Mixed or Multiple Ethnic Groups (10, 2.5%), those who preferred not to disclose (5, 1.3%), and others (3, 0.7%). While participants' home countries were relatively diverse, the majority were currently residing in the UK (351, 88.4%) and the US (44, 11.1%). The detailed characterization of participant demographic features is provided in Appendix Table E.1.

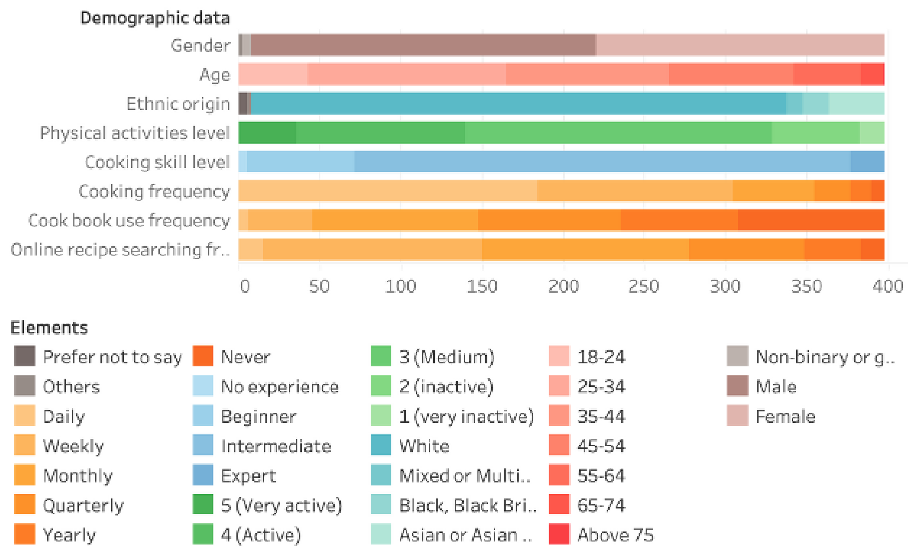


Figure 3.6: *Visualisation of participant demographics*

3.4.7 Data Cleaning and Preprocessing

The raw CSV data file was downloaded from Qualtrics and includes features such as *StartDate*, *EndDate*, *Participant Response ID*, *Questions*, and corresponding *answers* from participants. Qualtrics records the data in a horizontal format, where each row captures all responses from one participant. Since the survey questions were split into eight blocks, the data cannot be directly used for model building.

As data cleaning and preprocessing are the most critical steps in a machine learning project—rubbish in, rubbish out—it is essential to present the data handling strategy before discussing model building (Alpaydin, 2021). The cleaning and preprocessing steps were performed using Python Jupyter Notebook 6.3.0. A manipulation check was conducted manually, and participants who failed the check were removed from the study. Irrelevant records, such as "StartDate", "EndDate", "First_Click", and "Last _Click", were removed, while participants' demographic information was extracted and saved in a separate file. Participants' ratings for each recipe within the corresponding contextual block were extracted and transposed into vertical data records. The cleaned user*rating data matrix can be seen in Table 3.10. The entire dataset can be merged based on *recipe_name* and *user_id* for further model training.

After data cleaning and preprocessing, a total of 11,910 ratings were collected, distributed as follows: hot summer day (HSD) 1500 ratings, cold winter day (CWD) 1500 ratings, happy emotion (H) 1470 ratings, sad emotion (S) 1470 ratings, busy weekday (B) 1470 ratings, relaxing weekend (R) 1500 ratings, after physical activities (APA) 1500 ratings, and generic context-free group (G) 1500 ratings. The datasets for each experimental group were divided into training and testing sets using an 80:20 ratio. The target variable is the recipe rating, ranging from 1 to 5, maintaining consistency across each model.

A total of 29 features, which include contextual features, participants' demographic fea-

Table 3.10: *Example of cleaned user-rating matrix under hot summer day (HSD) contextual situation*

User_ID	Ratings	Contextual_scenario	Recipe_name
xxx01	1	HSD	Apple Pie by Grandma Ople
xxx02	1	HSD	Apple Pie by Grandma Ople
xxx03	2	HSD	Apple Pie by Grandma Ople
...	...	...	...
xxx50	3	HSD	Super Delicious Zuppa Toscana

tures (e.g. *age, gender, ethnic_orgin, home_country, current_living_country*), recipe content (e.g. *total_weight, recipe_category, cooking_complexity*) and nutritional features (e.g. *FSA_health_level, fat, saturates, sugar*), were utilised to develop rating prediction model. While the main focus of the study was on identifying novel dynamic contextual factors, the reason for including demographic features is that they have the potential to enhance model performance. As these features can provide the trained model with additional reference information.

In machine learning projects, the transformation of categorical data into numerical representations is crucial for algorithmic processing, and this is where label encoding and ordinal encoding play key roles (Udilă, 2023; Lopez-Arevalo et al., 2020; Singh & Singh, 2020; Poslavskaya & Korolev, 2023). Label encoding assigns a unique integer to each category, preserving the information but potentially introducing ordinal relationships where none exist (Lopez-Arevalo et al., 2020). Ordinal encoding, on the other hand, overcome the dimensionality challenge of one hot encoding, but is would particularly useful for features where the categories have a meaningful order, as it converting them into ordered integers (Udilă, 2023). In this study, label encoding (Low et al., 2022) was applied for categorical variable, such as *gender, recipe_name, user_id, ethnic_orgin, home_country, current_living_country, recipe_category*, and ordinal encoding (Choong & Lee, 2017) was applied for continuous variables, such as *age, cooking_complexity, physical_activities_level*.

3.4.8 Statistical Analysis

A one-way ANOVA (Analysis of Variance) test has been selected due to its widespread application in determining whether there are statistically significant differences between the means of three or more independent groups (Kim, 2015; Scalvedi et al., 2021; Bolek, 2020). This test is particularly well-suited to the research objective, which involves comparing the means of different contextual scenario groups to determine if one group's mean differs significantly from the overall mean. The one-way ANOVA evaluates the influence of a single independent variable (or factor) on a dependent variable by comparing the variance within each group to the variance between groups. If the variance between groups is significantly greater than the variance within groups, it indicates that the factor has a significant effect on the dependent variable (Dean & Voss, 1999).

In this study, the general preference of participants for each recipe (measured as the mean rating of each recipe) was compared across different contextual scenarios. The one-way

ANOVA is particularly advantageous because it offers greater control over the overall Type I error rate compared to conducting multiple t-tests. Performing multiple t-tests increases the likelihood of committing a Type I error, which is the probability of incorrectly rejecting the null hypothesis when it is actually true (Field, 2024).

Following the completion of an ANOVA test, if the results indicate a statistically significant difference among group means, it becomes necessary to identify precisely which groups differ from one another. While ANOVA can confirm the existence of differences among groups, it does not specify where these differences occur. To address this, post hoc tests, such as the Tukey Honestly Significant Difference (HSD) test, are employed.

The Tukey HSD test is particularly advantageous for this study, as it is capable of revealing which groups show significant differences in pairwise comparisons (Scheiner, 2020). Additionally, it is capable of controls the family-wise Type I error rate while conducting multiple pairwise comparisons between group means. The Tukey test systematically compares all possible pairs of group means, applying a correction that accounts for the number of comparisons being made, thereby ensuring the accuracy and reliability of the results (Nanda et al., 2021). By implementing the Tukey HSD test following an ANOVA, specific contextual groups that exhibit significant differences from others can be more clearly identified. This suggests that certain contextual scenarios may lead to varied recipe ratings and imply different eating behaviours.

3.4.9 User Rating Prediction Model Implementation

Two sets of experiment were conducted. The primary objective of the first experiment was to assess the importance of adding contextual features to the model. In this experiment, all 11,910 ratings were utilised. By systematically adding and removing contextual features within the model, the importance of each contextual feature can be determined. The second experiment was mainly to identify the most influential individual dynamic contextual factors at the model building level. To achieve this, the dataset was divided into eight groups based on contextual scenario groups to facilitate model performance comparisons.

According to recent systematic reviews on recommender systems conducted by Khanal et al. (2020) and Roy & Dutta (2022), tree-based models have been widely employed in model-based recommender systems. Therefore, we aimed to evaluate the performance of tree models on the collected dataset; XGBoost, a boosting tree model, was chosen for its efficiency and flexibility. Its objective function evaluates model performance based on a set of parameters, while a regularisation term controls model complexity to prevent overfitting (Lai et al., 2021). Notably, tree models offer more interpretable explanations than other models, which aligns with our goal of identifying the most influential features in the prediction task. In this study, the task of rating prediction was treated as a regression problem, rather than a classification or top-k relevance prediction problem, due to explicit feedback (recipe ratings) was obtained from participants. Consequently, evaluation metrics such as precision@K and nDCG would not be appropriate for evaluation. Instead, Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and R-squared (R2) were reported to evaluate the model test performance.

3.5 Research Methods Phase 3: Context-aware Healthy Food Recommender System Development and Evaluation

This section outlines the proposed methodologies for developing a context-aware rating prediction model and a healthy food recommendation system. In the previous stage of this research, contextual factors demonstrated significant potential to enhance model performance within pre-filtering and basic contextual modelling approaches. Building on these insights, the current stage seeks to further explore and optimise model performance by developing context-aware food recommender systems using advanced contextual modelling approaches. This approach integrates contextual information directly into the model learning process, alongside multi-level feature sets. The primary objective is to assess to what extent the incorporation of contextual factors enhances the effectiveness of the recommender system. Additionally, by systematically examining and evaluating the performance of different feature sets, the optimal combination of features with contextual factors that yields the best possible model performance was identified.

To counteract the inherent tendency of recommender system algorithms which potentially continue recommending unhealthy foods to users who enjoy high-calorie options, a novel weighted contextual healthy recommendation approach has been proposed. This approach is aligned with a new health evaluation metric and aims to balance the trade-off between user preferences and health-conscious recommendations, inspired by the work of Elsweiler & Harvey (2015). The goal is to ensure that the recommendations not only accommodate users' tastes, but also flexibly promote varying levels of healthier eating habits across diverse contextual situations.

This section begins with a description of the dataset in Section 3.5.1, followed by an explanation of the techniques used to extract and embed dynamic contextual features and multimodal feature sets in Section 3.5.2. Subsequently, a novel feature engineering approach is introduced in Section 3.5.3. The proposed one-stage and two-stage context-aware recommender system models are presented in Section 3.5.4, followed by a discussion of contextual healthy reranking methods in Section 3.5.5 and the evaluation metrics for the recommendation model in Section 3.5.6.

3.5.1 Data Description

The dataset used in this phase of the study was consistent with that collected during the second stage. However, the generic context-free group (G) was excluded from this analysis. This decision was made due to the ambiguity of the contextual situations when users provided ratings in this group, making it extremely difficult to accurately trace, control, or specify particular contextual situations for these ratings. Therefore, the removal of all data from the generic group was deemed reasonable. Moreover, the objective of this stage of the experiment was to identify the optimal feature combination with contextual factors that would enhance model performance. Including the generic group could potentially introduce confounding variables, thereby negatively impacting the model's performance.

As a result, the dataset now contains seven dynamic contextual factors, including hot summer's day (HSD), cold winter's day (CWD), happy emotion (H), sad emotion (S), busy and stressful weekday (B), relaxing weekend (R), and after physical activities (APA). The ratings from each participant were specifically provided under these contextual scenarios. After removing the genetic group, a total of 347 qualified participants and 75 unique recipes were included in this study. Each participant rated 30 recipes under a single assigned contextual scenario using a Likert scale from 1 (strongly dislike) to 5 (strongly like), resulting in a total of 10,410 ratings.

Notably, to mitigate bias during the second-stage experimental study, recipe images were not displayed to participants. As supported by (Zhang et al., 2020a, 2023), the visual appeal of recipe images may significantly influence choices; for instance, some images may appear brighter or more visually appealing. Additionally, since it was not feasible to select recipes created by the same chef or photographed by the same person, recipe images were excluded during the experimental step to maintain consistency.

However, during this third stage, the focus is on enriching the available features to enhance the RS model and identifying the optimal model performance. To support this effort, a recipe image database was created to facilitate the extraction of image features and their integration into model development. In this third stage study, recipe image features can be treated as hidden features (as participants have not directly seen these images while providing their ratings), to evaluate their potential for improving model performance and more accurately predicting rating preferences. This is also a common approach to integrating data to achieve advanced model performance in the domains of machine learning and recommender systems, as demonstrated by (Di Noia et al., 2012; Dang et al., 2021).

In total, 75 real-life taken recipe images corresponding to each selected recipe in this study were included in this stage of the research. All the images were manually downloaded from Allrecipes.com, and the researcher made every effort to ensure a consistent style when selecting each image, for example, the appealing level, the light and brightness among each recipe. The 75 recipe images can be seen in Google Drive: 75 selected recipe images [2]. Despite efforts to minimize bias, some may still persist. The selected recipe images may not be entirely uniform in style, and their level of visual appeal could vary. Additionally, consistency bias may arise, as the researcher's personal preferences for certain backgrounds, plating styles, or lighting conditions could reduce the dataset's diversity. Furthermore, the selected images may differ from how participants imagined the recipes, potentially influencing their perceptions. The selected images might not capture the full diversity of how the recipe is typically prepared, plated, or consumed across different cultures or regions. Future work could explore automated or randomised selection methods to reduce bias. For example, downloading multiple images for each recipe could help increase diversity in visual representations, ensuring a more balanced and representative dataset.

It is worth noting that the datasets used in this research, which including recipe content, nutritional information, image information, and user-recipe ratings under various contextual scenarios will be released for public access approximately 10 months from the time of writing, expected to be released on <6$^{\text{th}}$ January 2026>, and will be accessible through the following GitHub repository: Dataset for Context-Aware Healthy Food Recommender Systems

[2] 75 selected recipe images. Avilable at: `https://drive.google.com/drive/folders/1FrloqVCBAGUGJSEE_QF2ie2ck6irrVIV?usp=drive_link`

Development [3].

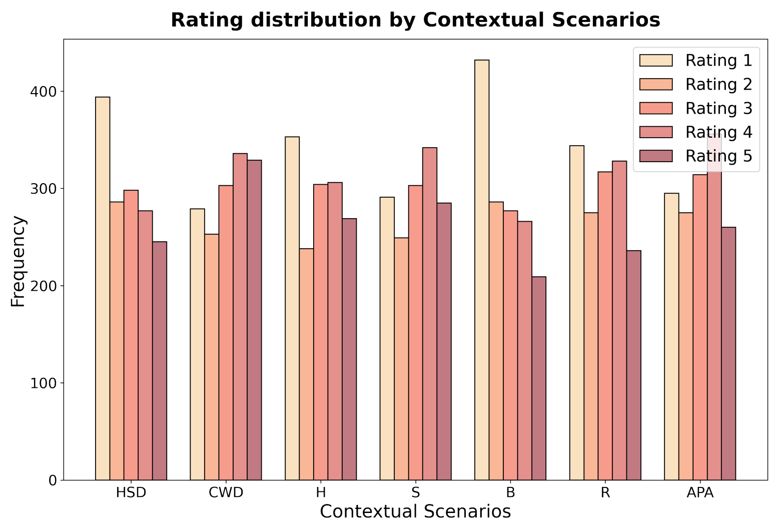


Figure 3.7: *Rating distribution among each contextual scenario*

The overall rating distribution can be seen in Figure 3.7, The variation in rating behaviour across different contextual situations highlights the necessity of this research in developing a contextual food recommendation model. Generally, the chosen recipes in this study are more popular during cold winter days and relaxing weekends. Conversely, these recipes are less preferred during busy and stressful weekdays, likely because they often require more preparation effort or time.

Following the data cleaning steps described in the second stage of the research (Section 3.4.7), five distinct parts of the data were separated, including user features, dynamic contextual features, recipe features, image features, and the user-rating feature matrix.

In this study, **dynamic contextual features** are key elements in the experiment and include seven factors. The original **user features** contain 13 features in total, include user ID, user demographic context features, which include *user_id, age, gender, ethnic origin, home country, current living country, current country living time, physical activities level, cooking frequency, cook book using frequency, online recipe searching frequency, cooking skill level.* The primary **recipe features** include basic recipe content features, such as, *recipe name, category, complexity, vegetarian and vegan label and recipe rating*, as well as recipe ingredients, cooking directions and images. As one of our aim is to achieve healthy recommendation, **nutritional information** also serve as key feature in this model, includes for instance, *sugars, fat, calcium, vitamin C, etc*, as well as three commonly used international standard (as mentioned above, FSA, WHO, FDA) to assigned health level for each recipe, the calculations

[3]Dataset for Context-Aware Healthy Food Recommender Systems Development. Avilable at: `https://github.com/Dreamiseast422/Dataset-for-Context-Aware-Healthy-Food-Recommender-Systems-Development.git`

were based on the detailed rules provided by each standard.

3.5.2 Multimodel Feature Extraction and Embedding

Various types of data are being processed in this study, including categorical, numerical, text, and image data. Feature-level fusion was utilised to incorporate all types of features into a unified representation (Ehatisham-Ul-Haq et al., 2019). Different strategies were employed for data preprocessing, feature extraction, and embedding, which are discussed in detail in the following section.

3.5.2.1 Encoding Categorical and Numerical Data

Unlike label encoding, which was discussed in Section 3.4.7, one-hot encoding represents features by creating binary columns for each category. This encoding method is independent of the target variable and ensures that each category is represented uniquely without implying any ordinal relationship. In this stage of the experiment, the *contextual scenario* feature was transformed using one-hot encoding, resulting in seven features: hot summer day (HSD), cold winter day (CWD), happy (H), sad (S), busy (B), relax (R), and after physical activities (APA). Recipe content feature, including, *recipe category, vegan vegetarian*; user demographic feature, including *ethnic origin, home country, current living country, and gender*, have all been processed with one hot encoding. The *cooking complexity, age, current country living time, and cooking skill level* have been encoded ordinarily, as these features represent levels of progression. However, due to one-hot encoding transforms categorical features into sparse vectors, increase the cost of dimensionality (Poslavskaya & Korolev, 2023). Features such as *user_id* and *recipe_id* are typically encoded using label encoding to avoid unnecessary increases in data dimensionality.

3.5.2.2 TF-IDF Encoding

Refer to previous work of Harvey et al. (2013) and El-Dosuky et al. (2012), the TF-IDF embedding has been utilised to handling recipe name, ingredients. TF-IDF effectively highlights the most distinctive terms within recipe descriptions by calculating the importance of words based on their frequency in a specific document relative to their occurrence across a larger corpus. This ensures that unique ingredients and key terms in a recipe are given greater weight, facilitating better differentiation between recipes (Ramos et al., 2003). Both recipe names and ingredient lists have been processed using basic text mining approaches first. Initially, any characters that are not letters (both uppercase and lowercase) or whitespace from the input text are removed. Then, all the characters are converted to lowercase. Next, the text is split into individual tokens, and a set of English stop words from the NLTK library is removed. Only the "-s" stemming has been considered for removal, to provide further clarification. Regarding the ingredient list specifically, a set of measurement words such as "cup," "pound," "tablespoon," "teaspoon," etc., has been removed from the tokens. The removed measurement list refers to wiki cookbook list [4]. After obtaining the cleaned tokenized list, TF-IDF vectoriser has been applied to both tokenized recipe names and ingredient lists.

[4]Wikibooks. *Cookbook:Units of measurement*. Available at: `https://en.wikibooks.org/wiki/Cookbook:Units_of_measurement`

After applying TF-IDF vectoriser, there are a total of 169 features in the recipe names and 322 features in the ingredient lists. As TF-IDF embedding also create a very sparse vector space, Principal Component Analysis (PCA) has been utilised to reduce the dimensionality of the features. According to James et al. (2013) and Wold et al. (1987), the determination of the number of components is based on explaining 95% of the dataset variance, representing the most dominant information. Consequently, PCA analysis resulted in 63 features from the recipe names and 62 features from the ingredient list.

3.5.2.3 BERT, Glove and Cooking Method-Matched Embedding

Previous work has seldom integrated cooking methods into recommender systems (Teng et al., 2012), despite the significant impact these methods have on a recipe's nutritional value. For example, a chicken salad prepared with deep-fried chicken fillets is likely to be less healthy than one made with grilled or boiled chicken. However, integrating cooking methods could be challenging, as the current state-of-the-art algorithms do not particularly work to separate cooking methods from cooking directions.

In this study, three methods were tested to handle cooking directions. After applying basic text mining approaches, similar to mentioned above, BERT Sentence Transformers (Reimers & Gurevych, 2019) and GloVe embedding (Pennington et al., 2014) were utilised to vectorize the cooking directions. By employing both models, text can be transformed into numerical data, making it easier to incorporate this cooking direction information into the rating prediction model. Particularly, as BERT Sentence Transformers have the ability to capture nuanced contextual relationships within text, considering both past and future context simultaneously. Additionally, the Sentence Transformers are fine-tuned on sentence pairs, which enhances their performance in semantic similarity tasks and makes them particularly effective for understanding and categorizing detailed and complex instructions found in cooking recipes (Devlin et al., 2018; Reimers & Gurevych, 2019).

As a comparison, GloVe offers robust word-level embeddings by analysing word co-occurrence statistics across a large corpus, capturing global statistical information about word occurrences (Pennington et al., 2014). This results in embeddings that effectively reflect the overall distributional properties of words in a language, ensuring a comprehensive understanding of their meanings in various contexts (Pennington et al., 2014). Since cooking methods are key information we would like to consider, they can also be treated as context-independent word representations. Therefore, it is worth investigating which embeddings would contribute to better model performance. This study initialise to use the GloVe pre-trained model "*glove.6B.100d.txt*".

Finally, in order to extract the cooking methods alone, a matching process for cooking methods was conducted. Using commonly used cooking techniques from Wikipedia [5] as part of analysis. A matching process was run for each token. If a token existed in the cooking techniques list, it was saved and output. Subsequently, TF-IDF metrics were created to represent the cooking methods of each recipe.

In order to reduce the feature dimensions, PCA based on explaining 95% of the dataset

[5] Wikipedia. *List of cooking techniques*. Available at: `https://en.wikipedia.org/wiki/List_of_cooking_techniques`

variance was again employed, resulting in 19 features for BERT sentence embedding and 30 features for GloVe embedding.

3.5.2.4 Image Feature Extraction

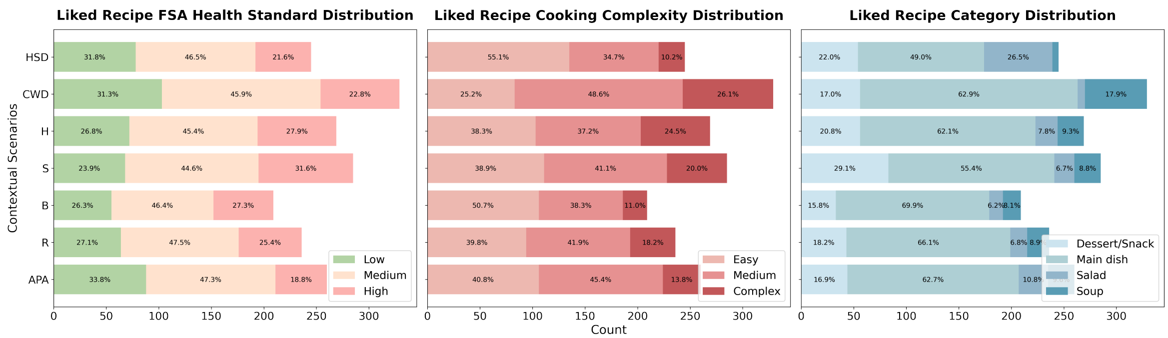
To further explore ways to enhance model performance, recipe image features were extracted and integrated into the model training process. Details about the utilised recipe image database can be found in 3.5.1. For the image feature set, five dimensions were derived to capture *sharpness, brightness, colourfulness, contrast, and texture (ASM)*. *Sharpness* typically reflects how clearly the details are defined in an image, such as the clarity of edges and fine details. *Brightness* refer to the overall lightness or darkness of an image. For example, an image of a salad with natural lighting might have higher brightness. *Colourfulness* measures how vivid or intense the colours in an image are. A fruit salad with vibrant reds, greens, and yellows may potentially have high colourfulness. *Contrast* normally refers to the difference between the light and dark areas of an image, it often helps to emphasise objects by creating a clear distinction between different elements. *Texture (ASM)* describes the surface quality or appearance of an object in an image, it captures the uniformity or smoothness of the texture (Jähne, 2005).

Incorporating these image features may potentially enhance the performance of the rating prediction model by providing additional information that helps the model capture and understand how visual appeal influences user preferences and rating behaviour. These features may also provide insights into the colour, texture, or shape of the ingredients, helping to better define similarities between recipes. The images features have been effectively utilised in previous studies to analyse the biases in food choices Elsweiler et al. (2017); Zhang et al. (2020a). Note that, to mitigate bias, recipe images were not displayed to participants during the data collection step (see survey design of the second stage study 3.4.5). The extracted recipe image features can be treated as hidden elements for integration into model training, enabling the evaluation of their potential to improve model performance and more accurately predict participants' rating preferences.

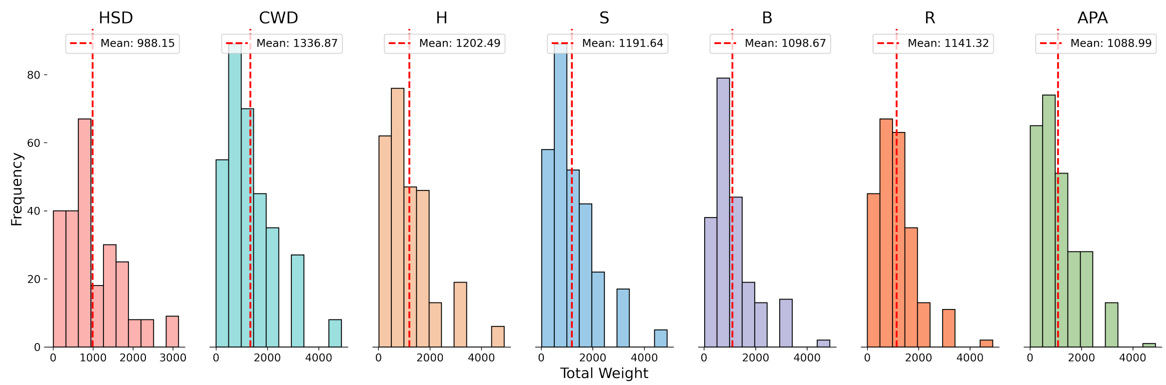
3.5.3 Feature Engineering

Based on the previous research results, which can be view in Chapter 4 and Chapter 5, individuals' recipe preferences vary among different contextual situations, which indicate the preferences in recipe content, such as recipe categories and recipe complexity, may differ, which in turn can lead to varying nutritional intake. To further investigate whether people's preferred recipes differed in terms of recipe content across various contextual scenarios, user ratings below 5 were filtered out. This allowed the analysis to focus on 1,833 highly favoured (5-star) recipes across seven contextual scenarios.

The impact of various food-based factors, such as FSA health level, recipe category, cooking complexity, and overall recipe weight, on preferred recipes across different contextual scenarios was visualized. As shown in 3.8, recipes categorised as unhealthy (FSA levels 10, 11, 12) were more preferred during cold winter days and periods of emotional change, such as happy and sad emotions. Interestingly, recipes favored during periods of sadness were predominantly of medium (44.6%) and unhealthy (30.6%) levels, according to the FSA health standard. Additionally, complex dishes and soups were preferred during cold winter days, while desserts and snacks were more popular during periods of sadness. Conversely, lighter dishes were particularly favoured during hot summer days.



(a) *FSA Health level, Cooking Complexity, and Category (categorical variables) distribution under each contextual scenario*



(b) *Total weight (continuous variable) distribution by contextual scenario*

Figure 3.8: *Preferred recipe FSA health level, cooking complexity, category, total_weight distribution among each contextual scenario*

Given the significant variation in user recipe preferences across different contexts, it is reasonable to hypothesise that integrating contextual information weights into basic recipe content features might potentially enhance model performance, at this approach may help to capture the hidden pattern in user's liked recipes.

Based on prior information, this study conduct further feature engineering experiment by adding the weight of contextual scenario to several recipe content feature, to explore whether this way of doing feature engineering could lead to improvement of model performance, the contextual mediator features are listed in Table 3.11. The weights were defined based on the values of label-encoded contextual scenarios.

3.5.4 Recommender Systems Algorithms

Given the availability of explicit user feedback (ratings on a scale from 1 to 5) acquired through the experimental study, the rating prediction task in this research is treated as a regression problem. The machine learning and deep learning techniques utilised in the

Table 3.11: *Contextual mediator features demonstration*

No.	Feature name	Explanation
1	Contextual_FSA	Weight of cs * FSA health level
2	Contextual_WHO	Weight of cs * WHO health level
3	Contextual_FDA	Weight of cs * FDA health level
4	Contextual_WHO_adj	Weight of cs * WHO_adj health level
5	Contextual_FDA_adj	Weight of cs * FDA_adj health level
6	Contextual_total_weight	Weight of cs * total_weight
7	Contextual_category	Weight of cs * category_label
8	Contextual_complexity	Weight of cs * cooking_complexity

Note: Weight of cs stands for weight of contextual scenario

development of the rating prediction model are then described. The proposed single-stage model workflow is depicted in Figure 3.9, enclosed in the pink frame. In the one-stage model, the original user*rating data was first split into an 80% training set and a 20% test set. The training set comprised 8,328 unique instances, while the test set contained 2,082 unique instances. The experiment was conducted using 5-fold cross-validation. These data were then merged with various embedded features, including contextual features, user demographic features, recipe content features, nutritional features, and image features. The merged training set (excluding user ratings) was used as the model input, with the predicted ratings as the output. Since the employed machine learning and deep learning approach does not incorporate the user's initial rating as a feature during the learning process, the proposed single-stage model may demonstrate a strong ability to address the cold-start problem.

To achieve optimal performance, a two-stage model is proposed, as shown in Figure 3.9 enclosed in the blue frame. The two-stage model incorporates decomposed SVD/NMF features to replace the label-encoded user and item features. By utilising participants' initial ratings from the training set, this approach is expected to improve model performance. It also demonstrates the potential of combining traditional recommender system techniques with machine learning and deep learning methods to develop a contextual modelling approach that offers better personalisation and explainability.

Further, the challenges associated with balancing trade-offs and providing healthy food recommendations are addressed by introducing the subtle weighted contextual healthy recommendation and novel health evaluation metrics. The details are discussed in Section 3.5.5 and 6.3.3.

According to the findings of the previous Literature Review Chapter 2. Support Vector Machine (SVM), tree-based models and deep learning model have been widely employed in model-based recommender systems (Khanal et al., 2020; Roy & Dutta, 2022). Therefore, three commonly used machine learning techniques and one deep learning technique were considered to determine which model would best address the challenge of rating prediction in this research. The models tested include ***XGBoost***, ***SVR***, ***MLP***, and ***Ridge regression***.

XGBoost, short for Extreme Gradient Boosting, is a powerful machine learning algorithm renowned for its effectiveness in various predictive modelling tasks. It operates by sequentially building a series of decision trees, with each subsequent tree attempting to correct the

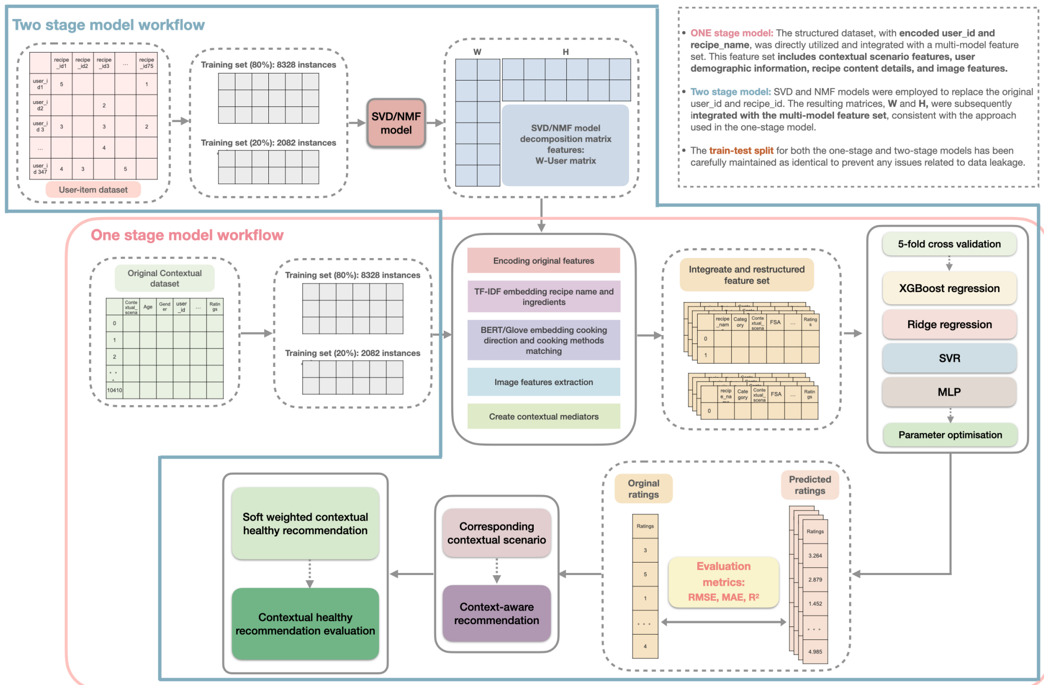


Figure 3.9: *One-stage and two-stage model building workflow*

errors of the previous ones. XGBoost uses a gradient boosting framework, where it minimizes a loss function by optimising the gradients of the loss function. This iterative process allows XGBoost to gradually improve its predictions over multiple rounds of boosting (Sagi & Rokach, 2021). As most of the features in this study are anticipated to have low correlation, XGBoost has its ability to handle complex, non-linear relationships within the data, making it well-suited for tasks such as rating prediction where the relationship between features and ratings may be intricate. Additionally, XGBoost incorporates regularisation techniques to prevent overfitting, which helps to improve its generalisation performance on unseen data. Its scalability and efficiency, coupled with the flexibility to handle various data types and formats, further confirming that it was the appropriate algorithm to examine in this study (Chen & Guestrin, 2016).

Support Vector Regression (SVR) is a variant of Support Vector Machines (SVM) that is specifically designed for regression tasks. SVR works by finding a hyperplane in a high-dimensional space that best fits the training data while also minimizing the error (Drucker et al., 1996). This model also have great ability to capture non-linear relationships between features and ratings through the use of kernel functions. By transforming the input features into a higher-dimensional space, SVR can effectively model complex relationships that may exist in the data. Additionally, SVR is robust to outliers, as it only considers data points within the margin of error, making it suitable for sparse data (Min & Han, 2005). However, compared to XGBoost, SVR training can be computationally intensive, and it may not perform well on large datasets with numerous variables.

Given that sequential data is not a consideration in the experiment, the primary objective is to identify a feature set that effectively represents user preferences, thereby enhancing the accuracy of rating predictions. In this regard, a Multilayer Perceptron (MLP) may be deemed a suitable model.

A ***Multilayer Perceptron (MLP)*** is a type of artificial neural network that consists of multiple layers of nodes, each connected to the next layer. The nodes in each layer apply a non-linear activation function to the weighted sum of their inputs, allowing the network to model complex relationships within the data. MLPs are trained using backpropagation, where the error between the predicted output and the true output is propagated backward through the network, and the weights are adjusted accordingly to minimize this error. MLPs are highly flexible and can learn complex patterns in the data, making them well-suited for tasks where the relationship between complex features and user ratings may be intricate or non-linear. Additionally, MLPs are capable of automatically extracting relevant features from the data, reducing the need for manual feature engineering (Naumov et al., 2019). However, MLP generally requires large amounts of data to perform well. With small training datasets or overly complex models, the generalization ability may decrease. It is worth examining the collected contextual rating dataset to determine whether MLP can achieve satisfactory performance.

In contrast to state-of-the-art algorithms, the efficiency of a simple ***ridge regression*** model is also worth examining in this study. Given the objective to identify the most influential feature set, the feature selection was not employed in the initial stages of the experiment. Ridge regression presents a compelling choice, especially when addressing multicollinearity, wherein predictor variables exhibit high correlation. This method is advantageous as it stabilizes the estimation process and enhances the model's robustness, a key attribute highlighted by (Hoerl & Kennard, 1970). Consequently, ridge regression emerges as a more suitable model for our specific case.

To further optimise prediction model performance, the traditional recommender system models, ***Singular Value Decomposition (SVD)*** and ***Non-negative Matrix Factorization (NMF)*** were incorporate into the contextual model-building process. This approach aims to explore whether these methods would lead to overall improvement in model performance. Both SVD and NMF approaches have been widely used in previous research in food recommendation (Harvey et al., 2012; Siddik & Wibowo, 2023). ***SVD*** is one of the most widely used matrix factorization algorithms due to its ability to reduce dimensionality and handle sparse data effectively (Koren et al., 2009). SVD decomposes user-item rating matrices into latent factors, which capture the underlying structure of the data. Given a matrix A of dimensions $m \times n$, the SVD of A is represented as:

$$A = U\Sigma V^T$$

where:

- U is an $m \times m$ orthogonal matrix. This could represent singular of user latent factors
- Σ is an $m \times n$ diagonal matrix. These values are known as the singular values of A.
- V is an $n \times n$ orthogonal matrix. This could represent singular of item latent factors.

After adopting the SVD model, the matrices U and V^T were extracted. These matrices were then matched and replaced the original *user_id* and *recipe_id*. Only the training data been used to generate this SVD feature representation, to avoid data leakage issue. This approach simplifies the integration of more complex contextual information and recipe content information, thereby may potentially enhance model performance. The model was trained from k=3 to 10.

The same procedure was followed for the NMF model to ensure comparability. ***NMF*** decomposes a matrix into two non-negative matrices (Lee & Seung, 2000), making it suitable for comparison with the SVD model. This comparison aimed to identify the most effective approach for integrating contextual factors into the rating prediction model.

3.5.5 Healthy Recommendation

According to Elsweiler et al. (2015), there are typically three ways to integrate nutritional information into RS. The first method involves creating a healthy recipe subset from the original dataset based on available nutrition information, and then making recommendations on this subset. For instance, a subset of recipes with relatively lower fat and sugar, higher protein and vitamin C content could be created. Although the model would be trained on the original dataset, recommendations would have drawn from the healthy subset, resulting in suggestions with lower fat and higher vitamin content. A second approach involves re-ranking the recommendation results based on the similarity of the "healthier" subset with the original gold-standard dataset.

Both approaches have the potential to recommend healthier alternative recipes. However, these methods may not appeal to users who have historically preferred richer, fattier dishes, as the healthier recommendations might feel too restrictive. Successful health recommendations employing these approaches depends not only on accurately predicting what users might like to eat, but also on effectively calculating recipe similarity. Without this, nutritional re-ranking results could deviate significantly from users' preferences. Additionally, finding the right balance is challenging—if the recommended substitutes are too healthy, users may lose interest, and if the selection is too small, maintaining recommendation diversity becomes difficult.

The third method is more complex but promising. This approach involves integrating user bias information, such as cooking time, recipe complexity (number of ingredients), and cooking methods (e.g., boiling, deep-frying, stir-frying), into the recommendation process. According to Harvey et al. (2013), these factors can significantly influence users' decision-making. By considering both contextual and constraint-based information that shape users' food preferences, this method has a higher likelihood of identifying recipes that users may be interested in. Consequently, the model is more likely to perform well and generate healthy recommendations that align closely with user preferences.

This third method closely aligns with the central idea of this research. To provide effective healthy recommendations, it is crucial to first identify the best way to offer suggestions that align with users' past preferences and then adjust these recommendations based on highly accurate predictions. In this study, prior to presenting healthy recommendations, hundreds of models have been evaluated to identify the optimal rating prediction model, integrating it with a comprehensive multimodel feature set that considers user demographics, recipe content, and contextual biases to determine the most effective combination. After identifying the best-performing model, a novel approach has been proposed to calculate the adjusted rating for contextual healthy recommendations. This approach dynamically adjusts the health level of recommendations based on specific contextual factors, enabling the re-ranking of recommendation lists to prioritise relatively healthier alternatives in each context. This method provides a more tailored and context-sensitive solution, potentially increasing the diversity and flexibility of the recommended options, as the re-ranking process does not solely rely on the health level of the recipes. Building on the work of Elsweiler & Harvey (2015), a new evaluation metric was introduced, specifically designed for healthy recommendations that balance with users' preference and nutritional needs. The detailed implementation is thoroughly discussed in the Chapter 6 Section 6.3.3.

3.5.6 Model Evaluation Metrics

Offline experiments serve as the primary evaluation method in this research. Various metrics were employed to assess the accuracy of rating predictions, including Average Root Mean Squared Error (Average RMSE) and Average Mean Absolute Error (Average MAE) and R2. RMSE and MAE are widely recognized as the most popular metrics for evaluating recommendation system prediction results, as suggested by Gunawardana et al. (2012). In the offline experiment, the system generates predicted ratings $\hat{r}_{ui}$ for the test set T of user-item pairs (u, i), which are then compared with the true rating r_{ui}. Typically, r_{ui} is known as the target feature but remains hidden during the offline experiment.

The RMSE between the predicted and actual ratings is calculated as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

Similarly, the Mean Absolute Error (MAE) is calculated as the average of the absolute differences between the predicted values and the actual values:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

Additionally, R^2, or the coefficient of determination, indicates how well the independent variables explain the variability of the dependent variable. It is a commonly used metric to evaluate the goodness of fit of a regression model and is calculated as:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

Here, n represents the number of observations, y_i is the observed value, $\hat{y}_i$ is the predicted value, and $\bar{y}$ is the mean of the observed values.

3.6 Ethical Considerations

The overall research design could be evaluated as low risk. In this research, semi-structured interviews and experimental study (survey design) were both acquired ethical approval by the Ethics Committee of the University of Sheffield, with reference numbers 045474 and 050555, respectively.

Before the start of any research activities, participants were kindly asked to read through the information sheet to familiarize themselves with the research aims and objectives, and to sign the consent form to indicate their understanding and agreement regarding the activities and responses that would be recorded, as well as how their data would be used.

During the experimental study, one of the contextual scenarios is designed to evoke feelings of sadness in participants, which might risk leading them into a negative emotional state. However, the video content only prompts participants to imagine or recall a moment, rather than inducing sadness in the present. Therefore, the overall research process carries little to no risk of physical harm to the participants. Even so, participants have been kindly informed that if at any point they encountered concerns or experience strong negative emotions, they are encouraged to contact their GP or the principal researcher and her supervisors. Additionally, the University offers mental health counselling services for timely support. Participants were informed of their right to withdraw from the study at any time, even after the survey had concluded (Kvale & Brinkmann, 2009) (The University of Sheffield Ethics Policy, 2018).

All user data collected through interviews and the experimental study were processed anonymously to protect participants' privacy. Throughout the project lifecycle, all research data has been stored in encrypted and password-protected environments. The data are primarily saved on the researcher's laptop, the University Google Drive, and the Prolific and Qualtrics platforms, all of which require complex passwords to access. Only the primary researcher and her supervisors have access to the data.

Chapter 4

Understanding General Food Preferences and Nutritional Intake Behaviour Under Different Contextual Situations: Insight From Interviews

4.1 Introduction

The chapter presents the findings from the semi-structured interview (the first stage of the work). The interview transcripts were analysed using thematic analysis, resulting in four primary themes that contributed to an overarching theme (see Figure 4.1). Each theme and sub-theme corresponds to the research questions addressed in the first stage of the study:

- RQ1.1: What static contextual factors affect people's (online) food choices?
- RQ1.2: What dynamic contextual factors affect people's (online) food choices?
- RQ1.3: What impact do these same factors have on people's nutritional intake?
- RQ1.4: How do individuals' online recipe searching behaviours match with real-life cooking intentions?
- RQ1.5: How can knowledge of the contextual factors benefit the development of healthy food recommender systems?

The results will be presented by discussing each theme and sub-theme in detail, supported by relevant quotes from the interviewees. The following sections will begin by explaining and demonstrating the most potentially influential static and dynamic contextual factors on food choices and decision-making processes, based on individuals' recalled daily eating experiences (Section 4.2 and Section 4.3). Next, Section 4.4 will focus on identifying individuals' nutritional attitudes and discussing how these contextual factors influence nutritional

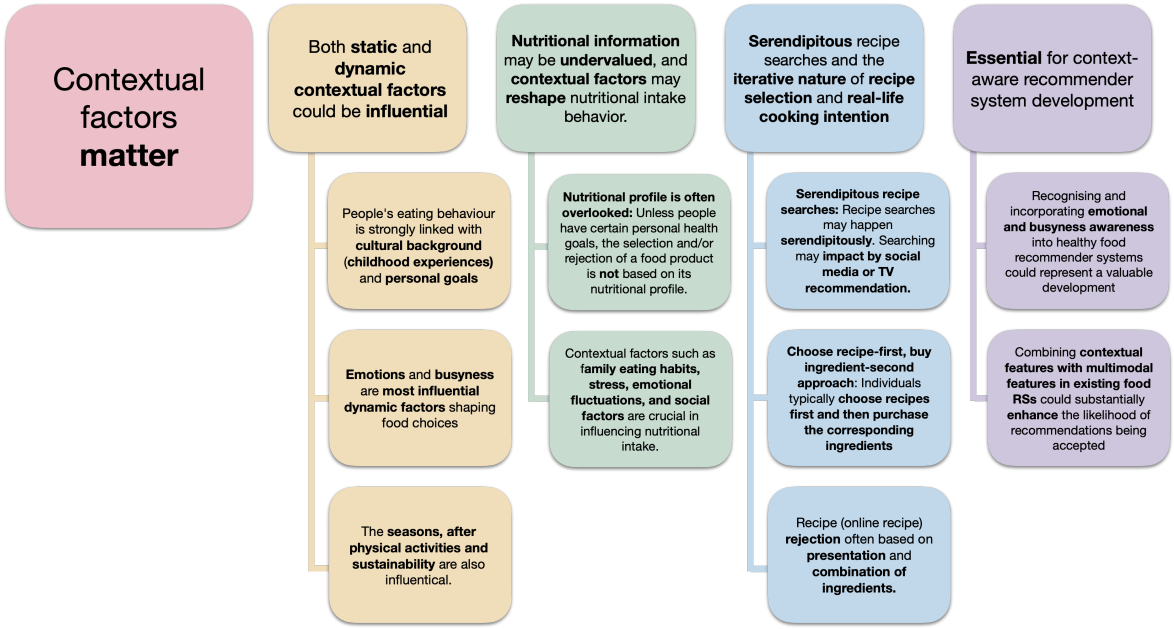


Figure 4.1: *Themes and sub-themes identified from analysis of interview data. (A detailed description and explanation can be found in Table 3.5 in the Methodology chapter)*

intake. Section 4.5 will analyse the alignment between individuals' online recipe searching, decision-making and their real-life cooking intention. Finally, Section 4.6 will discuss how these research findings can be applied to inform the development of the next generation of healthy food recommender systems.

4.2 Influential Static Contextual Factors of Food Choice

The most influential static contextual factor identified in this study including **cultural background**, particularly their childhood and family eating preference, and strong **personal goals**. People's eating habits are deeply rooted in their early life experiences, particularly the types of food typically eaten in their family and home country. In the meantime, people who have specific personal goals, such as fitness targets, weight management, or dietary restrictions (e.g. health-related goals), tend to make more deliberate and goal-oriented food choices.

4.2.1 Cultural Background

People's eating behaviour is strongly linked with childhood experiences, and the foods typically eaten in one's family and home country. 10 of the total 14 participants mentioned without hesitation that they prefer their home country's food. Participants commonly mentioned that, because they grew up there (P4, P5), they are simply used to their home country's food (P13), and it provides comfort.

- "*Personally I prefer food from my home country, definitely. I have a Chinese stom-*

ach..." (P4)).

- "*I eat my home country food since I was a little. So, year after years, I already get used to this food...*" (P5).
- "*I have a real love for stodgy classic British food... So I eat what my grandmother liked to eat... maybe I didn't grow up in a very multicultural place either...*" (P13).

Another reason was related to the level of knowledge participants felt they had about different cuisines, as they tended to know more or were more familiar with the food from their home country (P3, P9). Additionally, this may also be linked to the need for diversity in food choices (P4, P5, P12). One participant spontaneously mentioned that he believes his home country's food is healthier (P10).

- "*generally, I know more dishes from Germany [my home country] and I like them better...*" (P3).
- "*I do cook like pasta and steak and some pies sometimes at home. But mainly still Chinese food because I get used to its tastes also better. [Laugh]...*" (P9).
- "*The food are quite different, and I can try like different food. And to select the one I preferred...*" (P4)

Alternative responses were due to lack of recognised "standard" food in the participant's home country (P11), or a high level of similarity between the cuisine of the home country and the adopted country (P8). Moving to the adopted country at a young age can also bring about a dramatic change in eating preference (P7). Interestingly, being overly familiar with food from one's home country may paradoxically stimulate a preference for trying something new.

- "*We have a lot of kinds of food from a lot of different countries. So there is no standard home country food here... so, I would say that I prefer to eat a variety of different styles of food...*" (P11).
- "*Because both countries have got a lot of overlap in the food, in both countries have a lot of the same foreign foods, things like that. I don't really have a preference that way...*" (P8).
- "*But I've been living here for most of my life, so I know it very well. So I prefer to try new things...*" (P14).

The influence of one's current living environment could affect food choices, though it may be less impactful compared to cultural background and family eating habits. Most participants expressed an open-minded attitude toward trying foods from around the world, with foreign cuisine often seen as a source of discovery rather than a daily dietary choice.

4.2.2 Personal Goals

Individuals with strong personal goals exhibited notably distinct eating behaviours. While most participants expressed a desire for dietary diversity in their daily lives, those with well-defined personal objectives may lean toward a more uniform diet. For the latter group, the primary objective of their eating habits is to absorb essential nutrients, prioritising nutritional intake over the enjoyment of food, as exemplified by P2 and P14.

- Participant P2 explained,"*because I eat what I eat so regularly, if I deviate from that, it starts a pattern where I look forward to eating healthier, like after a while, it's good for a treat...*".
- Participant P14 stated: "*I think protein in general, is really tasty...not really care about eating diversity every day. My main goal is to achieve the protein intake...*".

Moreover, participants with strong personal goals often maintain a high level of physical activity and are inclined to meticulously track their food consumption to ensure they meet their nutritional requirements. Even though individuals with strong personal goals may occasionally have cravings for specific foods, they often refer to such occasions as "cheat meals", indulging once or twice a week before returning to their regular dietary patterns. This illustrates that strong personal orientation can, to some extent, override the natural human inclination to crave unhealthy foods.

4.3 Influential Dynamic Contextual Factors of Food Choice

4.3.1 Emotions

The findings in this study indicated that emotion is one of the most influential factor shaping an individual's food choice - 6 participants explicitly stated that they believe emotions to be the key factor that influences their food choices and eating behaviour. Female participants were typically more willing to admit that emotional change impacts their food choice to a large extent: "*yeah, definitely, I am definitely an emotional eater...*" (P13).

The most influential type of emotion mentioned was lethargy or tiredness, when participants would cook convenience food or not cook at all. Some participants mentioned they stop eating (P9), or want to eat something salty and pungent (P13) when they feel tired. Negative emotions tended to have more impact on eating patterns compared to positive emotions. For example, some participants noted if they feel sad or stressed that they tend to eat something quick and easy, or eat more sweets and snacks (P11 and P14), or they might stop eating at all (P9 and P12).

- "*If I don't feel like so well, and I'm sad... I might just not [be] eating all day...*"(P9)
- "*Maybe I eat less, maybe if I'm stressed and kind of working, I might forget to eat, or I will just have a really quick snack and back to work, rather than having a proper meal...*" (P11)
- "*When I'm stressed, when I'm, for example, if I am sad or got any emotional disturbances, I don't eat. Okay. I don't eat if I am [in] stress, I can stay really like two three days, only eating sweets and sugar, cocoa, lots of stuff...Sometimes I want to eat, but because I'm stressed, I can't, and I have food in front of me and I can't...*" (P12)

When people feel relaxed or happy, there may not be a big change in the types of food consumed, but the volume eaten may alter substantially. If they feel very happy or in a celebratory mood, they tend to go to a restaurant, and have a big meal to reward themselves.

- "*if I am happy, I eat a lot...*" (P5).
- "*If I got like first class, and I'm so happy, I will take my friends for Arabic food...*" (P12).

Notably, either when participants are happy or sad, sweet food is a popular choice. Participants frequently mentioned that during moments of happiness, they enjoy baking (P9) or consuming more sweet foods (P6). Also, when they feel sad or experience difficult situations, they also tend to have cravings for sweet foods (P14).

- "*Well, maybe just happy, or something happens, something good happens. I can eat some sweet food or something like a cake...*" (P6).
- "*Yeah, like I, if I had a happy day, or like, I would rather do some bakings, because I like sweet things, sweet things that will make me feel better...*" (P9).
- "*Like last week I had a very bad flu. So I was in bed like for three days and all I was thinking about was getting cake...*" (P14).

4.3.2 Busyness

Based on the frequency mentioned by participants, busyness was the second most influential contextual factor, and appeared to act more as a constraining factor, significantly limiting individuals' food choice-making processes. During busy or particularly stressful schedules, participants commonly mentioned they tend to eat food that is simple, easy and quick to cook, and doesn't require a long time to eat (P3, P11 and P14).

- "*I mean certainly. If I'm busy and have not much time in the evening, we need to cook something that doesn't require much time, or just get some food delivered. Yes, that would certainly have an impact...*" (P3)
- "*You know if I'm super busy with work, and I just don't have the time to put into, cooking a big meal. I will choose something to cook, that is quicker and takes less effort, and then so we do have some things like that, that we know that we can cook in, like, just maybe 5 or 10 or 15 minutes of prep time...*" (P11)

One participant expressed that they tend to rely on meal planning to maintain a balanced diet under a busy schedule (P8), but two participants spontaneously mentioned that it is difficult to consistently stick to meal plans (P13, P14).

- "*Yeah. I'm not, I don't sit down and like, I'm not one of these meal planners or crackers. I couldn't do that, because I, when I tried that in the past, I've kind of, but that's not what I feel like eating now. And I have to really feel like I want to eat that thing. And so, I can't decide too far in advance...*" (P13)
- "*When you're working, it's really hard for me to organise my food, my meals. So I end up having whatever I find on the street...I tried, but I really am bad at it. Like, no. I*

just, it makes me. I don't know. It's like a drug to me. It's like a task. Something that I find very boring ..." (P14).

In comparison, during a relaxing weekend, they tend to cook food that requires more effort and spend more time enjoying the process of planning for, cooking, and eating food.

4.3.3 Seasons

Participants commonly noted that their food preferences between two distinct seasons (summer and winter) are different. Some expressed strong differences between the seasons.

- "*it [the season] certainly does impact [my choices] a lot...So yeah, in the summer time we like to like do barbecue and smoke food and so... It's a lot of like, cooking outside on the barbecue, or on the smoker, and then in the winter, it's less nice out in the winter time...*" (P11).

Participants often mentioned they like to eat salads, light and cold foods in summer, or they like to cook outside and do more barbecues. Whereas in winter, they mentioned a preference for warm, large portions of thick soups or stews; one participant even called wintertime a "soup season".

- "*I'm so excited that we're coming back into a soup season [winter season] and I can make big vats of soup and eat soup every day...*" (P13).

However, few participants indicated that seasons don't impact their eating behaviour, except during special holidays, like Christmas. For example, participant 14 explained this as follows: *I couldn't say [that] seasons heavily influenced me; I would say like Christmas is, is the toughest one because it's when you socialise a lot, you tend to see a lot of people you probably drink more alcohol than before...*" (P14).

4.3.4 Physical Activity

It is clear from the interviews that most participants are aware of eating healthier after doing physical activity, for example, cooking something fresh (P2, P5), or trying to eat high-protein foods after doing physical activity (P13)

- "*if I've done physical activity, I'm gonna cook the meal in myself, like hands down every time...*" (P2).
- "*I would check my refrigerator to see if there is, especially some juicy food. And instead of dry food, to keep me refreshed... [I] wouldn't order food outside...*" (P5).
- "*I do quite often, in the situation of [sic] being going out for a run and ending somewhere for some food...and I'm always very happy if there's a nice big steak pie on the menu or like a meat, like a nice, big meat pie...*" (P13).

On the other hand, after engaging in sports, people may have specific dietary needs that differ from common understanding. For example, one participant, an amateur marathon enthusiast, mentioned that carbohydrates are more important than protein for long-distance runners. Participant P13 stated that "*People quite often think 'oh, you probably eat... really* *good stuff', but actually, quite often, it's lots of like white pasta, and white bread and Pizza. Pizza is my favourite thing to eat before we run...*" (P13).

Several participants expressed that the significant fatigue experienced after exercise makes them inclined to prepare something simple and quick (P9), or even eat leftovers (P11), rather than preparing a larger meal that requires a lot of effort. Interestingly, self-identified "foodies" (individuals with a strong passion for food) said they may eat fast food after exercise, because they treated themselves after exercise by eating delicious, high-calorie food. One of the participants, a University Professor, claimed that "*if you do a bunch of exercise, and then you have the fast food afterwards, then everything's balanced out...*" (P11).

4.3.5 Sustainability

Two participants mentioned that preventing or mitigating food waste is an important factor that influences their food decision-making (P7, P11). Their views affirm the importance of potentially incorporating sustainability in future food recommendation systems.

- "*Like these days, I pay attention to my food waste...because like if you cook something, it doesn't keep well, and you end up cook[ing] a lot...*" (P7).
- "*I don't like to waste things... the other types of things we cook are generally sized for what we want...if we're saying we're going to be travelling or in the next couple of days we might adjust what we're going to be cooking...*" (P11).

4.3.6 Pregnancy

It is worth noting that one female participant mentioned that her food preferences and appetite changed significantly during her pregnancy. However, since only one participant discussed their pregnancy experience, this factor is not considered representative in this study.

- "*since I'm pregnant..I start[ed] to hate pork, even [the] smell ... [can] make me feel sick*" (P9).

4.4 Nutritional Information May be Undervalued, and Contextual Factors May Reshape Nutritional Intake Behaviour

Participants' attitudes towards nutritional information varied. Their views broadly fell into three categories: *do not pay much attention, not overly concerned* (expressed by five participants), *important but not critical* (expressed by six participants), and *very important* (expressed by only three participants).

The data pointed to a pattern where participants who engage in higher levels of physical activities tended to care more about nutritional information. One participant, who is a professional athlete with clear goals for their body shape and composition, cared the most about nutritional information (P2). Who often reads nutritional labels and rejects food if he doesn't know exactly what is inside. However, rejection also happens very rarely, because he tends to choose foods which he already believes are healthy to eat.

- "*if I want to try something, and it didn't tell me what it had in it in terms of nutrition, I won't buy it...*" (P2).
- "*because basically like most of my recipes might come from something like the myprotein website and stuff like, where as the name suggest my protein, they're gonna be very careful in terms of the nutritional information that's on each of the recipes...*" (P2).

Participants who are less keen on exercise tended to care less about nutritional information, and rarely mentioned it during the interviews (P5). Unless they encounter specific problems with their health, for example, high levels of blood sugar or cholesterol (P1).

- "*I don't pay much attention [sic] on this [nutritional information]. I would say I feel guilty after [I eat] KFC or McDonalds and then tomorrow I would order a fast food again...*" (P5).
- "*The doctor told me that... I have not enough vitamin D, I have slightly high cholesterol... so I need to eat... in healthier way...*" (P1).

Other participants noted that they think nutritional information is important, but they do not care so much about it and their selection and/or rejection of a food product is not based on its nutritional profile.

- "*I think about it. But it doesn't really affect my choice...*" (P11).
- "*I want it to be important... I do try to think about it, but I don't want to get obsessed over it either...*" (P13).

Most of them believe that they already have or can acquire enough knowledge about nutrition, especially in the UK where nutritional information is easily accessible because of the traffic light health system (FSA standard (Fsa, 2016)). However, obtaining such nutritional information in Saudi Arabia and Qatar is more challenging. "*But in Qatar, we don't have this option yet [to acquire accurate nutritional information]...*" (P10). "*if I try to go and eat the Arabic foods, you're gonna get zero [nutritional information]...*" (P12).

Few participants mentioned they make efforts to be aware of and to lower their sugar intake. Although two participants (P4, P5) claimed not to care about nutritional information, but recently changed from Coke to Diet Coke, as they've become more conscious of body shape with age.

- "*I tend to try to keep sugar as low as I can. So I like to look for things that have got a low sugar content...*" (P13).
- "*Previously, I really like[d] drinking Coke, but now I just [drink] Diet Coke. It is a big change to me...*" (P4).
- "*I sometimes like I drink Coke a lot in recent years. And I clearly know that it's not good for my health. So I would choose like zero sugar, Coke or Diet Coke...*" (P5).

While people's attitudes towards nutritional information (NI) may be associated with their actual nutritional intake behaviour, this is not always the case. In this study, several situations were identified where participants' nutritional intake behaviour was significantly influenced by specific contextual factors. Attempts were made to provide explanations on how these changes occurred. These cases are discussed in detail below.

4.4.1 Unhealthy Eating Habits Often Originate from Family Eating Habits

Childhood influences on eating behaviour can, in some instances, outweigh later environmental factors. Many unhealthy eating habits tend to take root during childhood. For example, participants frequently mentioned that their family meals often lacked vegetables. Despite their attempts to incorporate more vegetables into their diets as adults, they often struggled due to their pre-existing and ingrained preference for meat (P1, P3, P14).

- "*We didn't have this habit that we need to include a salad for every meal...*" (P1).
- "*So one thing that comes to mind is [my] family rarely had vegetarian dishes...*" (P3).
- "*I don't tend to have as many [sic] veggies I think. I should. So I try to incorporate them but then it's kind of like forcing myself to, to eat them...*" (P14).

Some participants expressed that their family meal often consisted of very oily foods or lots of sweets (P10, P12, P14).

- "*I do like it [Arabic food]. But not all the time, because it just is too much really [too oily]...*" (P12).
- "*So they are cooking by their measurements which is oily and spicy and everything...this is really bad because, there are third generation, which is my nieces and my nephews, they are inheriting the same thing...*" (P10).
- "*I think my family tends to be very much of a sweet tooth. So we would normally have sweets or cakes...*" (P14).

As parents would often take charge of cooking, a few participants expressed that they don't have much freedom when it came to choosing their food (P9). Other participants claimed that their eating habits had completely reversed compared to their childhood (P9 and P10). But in certain cases, traces of childhood habits lingered. For instance, one participant unconsciously admitted that, when feeling very happy, she would order fast food delivery, even though she acknowledged she couldn't tolerate oily foods (P10).

- "*because when I was a little one, this they just forces [sic] me to eat too many fish. Till now. I don't want to [eat it], I still don't want to [eat it], even like they have like a very delicious delicious fish prepared for me, like put in my plate, probably, I just still don't want to [eat it]...*" (P9).
- "*whenever my husband travels, I stay at home, I order all the junk food, like punishing him. And I took a picture of him like, you're travelling, I would order [junk food]... But I don't eat when I'm sad...*" (P10).

4.4.2 Busyness May Lead to High Calorie and Unhealthy Food Intake

A noticeable change in nutritional intake behavior was observed during busy circumstances. Participants often mentioned that, when very busy, they have less time/mental space to cook and tend to eat instant noodles, pizza, fast food delivery, or leftovers. A balanced diet including fresh vegetables and fruits is less likely in such situations. Based on the life experiences mentioned by the participants, the complexity of cooking could be an essential factor to consider when providing recommendations for busy weekdays. If a recipe is overly complex to prepare or the ingredients are difficult to acquire, it may hinder accessibility to healthier options.

- "*There something happened to me to 2 or 3 weeks ago. I spend the whole week eating Subways, because I didn't have time to prepare my food...*" (P6).
- "*[If I have a busy day at work I will, I might just eat pizza, really. I just order wood-fired pizza...*" (P7).
- "*You know if I'm super busy with work and I just don't have the time to put into cooking a big meal, I will choose something to cook that is quicker and takes less effort...*" (P11).

4.4.3 Increases in Stress Level and Emotional Fluctuations May Lead to High Calorie and Sugar Intake, or Even Eating Disorders

This study identified that feelings of stress and sadness could trigger unhealthy eating behaviour. One participant mentioned that, when he is under a lot of pressure, he would stop eating normally and have a "sugar carnival", because he ceases to have sufficient appetite to eat full meals, and sugar helps to fill in the calorific void. Even though the participants are aware that this is an unhealthy behaviour, they find it challenging to control themselves. The prevalence of cases of eating difficulties under stress is much higher than anticipated, a problem that hasn't been addressed by current food recommender systems.

- "*I probably just, like I drink, I don't really eat. I don't really eat like when I am stressed...*" (P7).
- "*If I don't feel like so well and I'm sad or I'm angry, might just not eat all day...*" (P9).
- "*So, I will not eat if I am stressed, okay? I'll have two or three days and I will not eat, and will lose weight, but it's unhealthy weight...*" (P10).
- "*When I'm stressed, when I'm, for example, if I am sad or got any emotional disturbances, I don't eat. Okay. I don't eat if I am stressed, I can stay really like two three days, only, eating sweets and sugar, cocoa, lots of stuff...*" (P12).

4.4.4 Social Pressure Can Often Promote Unhealthy Eating

Ten out of fourteen participants mentioned that they often need to compromise when eating with friends or family and feel social pressure to eat. One participant admitted that personal choices within a social group are often restricted or limited (P3, P14). In the meantime, unhealthy eating behaviour can easily become contagious within a social group (P14).

- "*maybe sometimes I would prefer something healthy but [with friends] still end up at a fast [food restaurant] because it's convenient...*" (P3).
- "*[they think] you have to have breakfast, you have to eat, and then you feel forced to eat breakfast...*" (P14).
- "*if you have surrounded yourself with people who snack a lot, who eat lots of crisps for example, you end up eating the crisps as well because they are there, so they influence you too...*" (P14).

4.5 Serendipitous Recipe Searches and the Iterative Nature of Recipe Selection and Real-life Cooking Intention

A clear pattern has been observed in this study, indicating that people are increasingly turning to online search platforms to find recipes: seven of the participants claimed to search for recipes online daily or weekly, and six participants stated that they never use a cookbook. This was particularly so if they wanted to try new dishes or seek culinary inspiration. This also implies that people may be more vulnerable to the complexities of recipe information on the Internet, where there is potential for advertising or promotional messages to be personalised and direct. This further emphasises the importance of healthly recommendations. One participant, who is business strategist and self-described food app enthusiast, suggested that "*I think it's really difficult [...] we are surrounded by loads of trends, and and strange information, to be honest. I think if you...really don't understand nutrition, then you can actually become [sic] malnutritioned ...*" (P14).

In addition, nine participants expressed that the decision-making process of what to have for lunch or dinner is often a struggle. The reasons for their difficulty in deciding vary, ranging from a desire for dietary diversity to curiosity to try something new. Participants commonly reported that recipe searches may happen serendipitously, and could be easily influenced by the surrounding environmental context. For example, while watching TV or browsing social media, they might come across a certain recipe that attracts them. Subsequently, they might continue to search for information related to the recipe in order to save it for the near future.

Moreover, nine participants stated that they typically choose recipes first and then purchase the corresponding ingredients to cook. However, conversely, seven other participants indicated that they prefer to buy ingredients first and then select appropriate recipes to cook. Some participants mentioned that they might follow either approach depending on the situation. Traditional recommender systems work well when users select a recipe first and then buy the necessary ingredients. However, for those whose food decision-making process is the opposite, considering constraint-aware recommender systems, such as intelligent refrigerators that can monitor available ingredients and recommend recipes based on what's on hand, could be a more suitable approach.

Almost all participants emphasised the importance of the visual aspects of recipe photos when searching for recipes online and making food choices. - whether a recipe looks appetising significantly influences their decision to prepare or consume it. Additionally, the presence of ingredients that individuals dislike is the primary factor leading to the rejection of a recipe. This may emphasises the importance of integrating image features into recommender systems. If a multimodal feature matrix can be constructed, the system can assign weights to each feature, allowing it to prioritise and suggest options that are both healthy and visually appealing. This might substantially enhance the likelihood of a recommendation being accepted. Additionally, the presence of ingredients that individuals dislike is the primary factor leading to the rejection of a recipe. Surprisingly, the healthiness of a recipe does not seem to be a major concern for most participants. This is a positive sign, suggesting that the rejection of a recommendation is unlikely to occur due to the recipe being percieved to be too healthy.

Furthermore, a few participants emphasised that the quality or provenance of a recipe heavily influences their choice when browsing online recipes. One participant mentioned, '*It's more about where it's come from? It comes with the name of a particular chef or a specific cooking style attached to it, and there are certain people whose recipes I trust...*' (P13). Therefore, for researchers responsible for designing FRS, it is also their responsibility to ensure the reliability of the sources for the recipes. The nutritional information and the health level of recipes should be based on trustworthy sources.

4.6 Ideas and Implications for Context-aware Healthy Food Recommender System Development

As individuals' eating and nutritional intake behaviours change under specific contextual situations, such as emotional fluctuations and shifts in busyness status, these are the most influential factors, as supported by the previous analysis. Thus, the development of an emotion and busyness-aware persuasive healthy food recommender system could be highly beneficial. Particularly when people experiencing negative emotions and high levels of stress, disordered eating behaviours are more likely to emerge, highlighting the needs. Such a system could identify a user's emotional state through wearable devices, or based on user self-report. It could then provide context-aware recommendations aimed at promoting healthier eating habits, with emotional state operating as a moderator. For example, they could suggest alternative cooking methods that require lower effort or low-sugar dessert options to mitigate people's tendency to consume high-calorie foods, or to eat sweets as meal replacements, when stressed or sad.

Participants' expectations for the next generation of food recommendation systems mainly include three aspects: firstly, diverse recommendation results; Secondly, personalised recommendation results, and the ability to filter the results using a variety of facets; Lastly, two participants spontaneously mentioned the potential to recommend healthier alternatives.

- "*I would say keeping a history of what has been done in the past so you can recommend people the things that people like but also inserting some variety. So instead of always the same thing. I think that would be important...*" (P11).
- "*So yeah, if it takes into account, like all the different preferences, I think it's really hard to...There are people who are like intolerant to certain kinds of food and so on, so*

yes, I guess sometimes we do need inspiration..." (P14).

- "*I guess I suppose it's think, you know, like, is it really easy to filter out the stuff that you know, is not vegetarian or not vegan, filter out different ingredients. How long something takes to cook?...*" (P13).
- "*Obviously the taste better, it's like, I don't know, it's more for me. It's basically just what's in it? Is it healthy? Is it similar to this?...*" (P2)
- "*Also, some of them are healthy, the seafood. But also they have like the meat, the lamb, beef, the chicken and, you know, they have different types of rice dishes which is very oily, not very healthy I suppose...*" (P12)

This underscores a critical consideration in the development of the next generation of food and healthy food recommendation systems. These systems should not only account for individuals' past preferences but also strive to suggest options that extend beyond their comfort zones, thereby providing personalised recommendations that encompass more diversity and promote health. Furthermore, they should offer users more control over their recommendations. Such a system could also provide automatic reminders and explanations at the same time, rather than just passive recommendations. For instance, it could alert users when emotional fluctuations have been detected and send notifications such as: "*Based on your past eating behaviour, you would normally choose these recipes; however, we would recommend the following healthier alternative options....*" Furthermore, as users expect more control over the system (as mentioned by P13), the system should allow users to filter and re-rank recommendation results based on factors like sugar level, calorie content, or cooking complexity, among others. There is currently no such system in the field of food recommendation that combines emotional contextual factors with healthy recommendations. Nevertheless, it is reasonable to believe that such a recommendation system would have the potential to subtly influence and change habits over time. Ultimately, in the realm of food recommendation systems, all habit changes begin with the acceptance of recommendations and are translated into actual actions. However, meeting this requirement presents a significant challenge for traditional recommendation system algorithms (Min et al., 2019a).

Referring to the RS structure proposed by (Oh et al., 2010), the conceptual workflow of a potentially explainable emotion and busyness-aware persuasive healthy food recommender system is illustrated in Figure 4.2. Following a reported change in the user's emotional state, the contextual information provider and context manager collaborate to provide regulating information for the recommendation list. The primary recommendation list could be generated based on the user's past eating preferences and multimodal food-based factors. After being modified by the model adapter, the system should be able to generate contextual recommendation lists that are more suitable for the user's current context. Then, after incorporating recipe nutritional information, the user will be given the freedom to filter the results as well. Finally, explanations and persuasive results will be provided to the user to increase the transparency of the algorithm and make recommendations more convincing, trustworthy, and reliable.

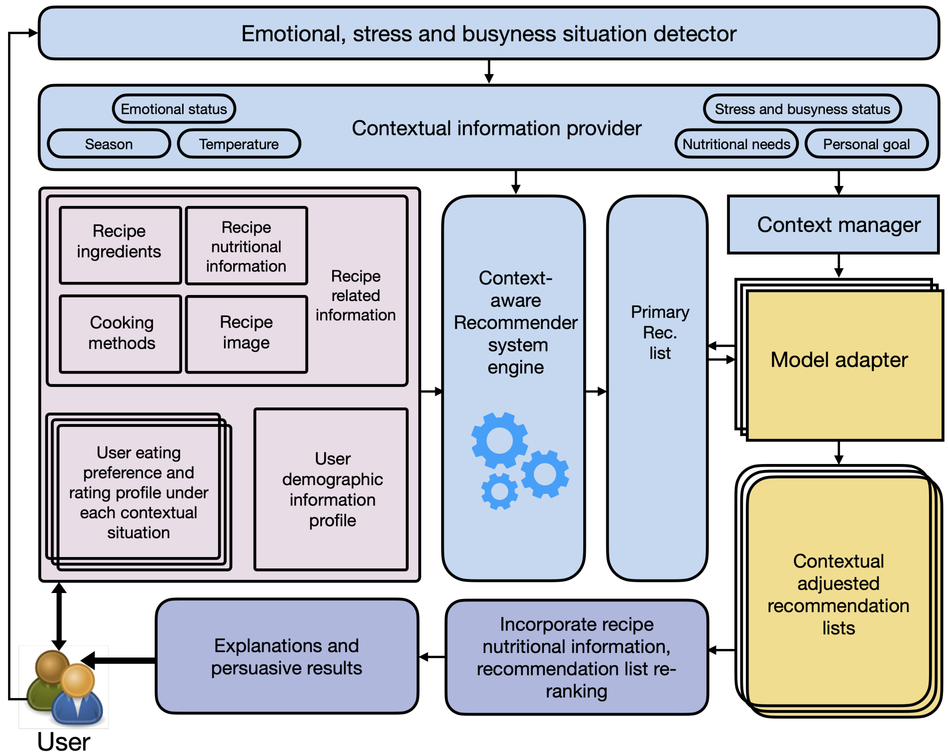


Figure 4.2: *Emotion and busyness-aware persuasive healthy food recommender system workflow*

4.7 Discussion

Based on the previous analysis, the key static factors identified in response to RQ1 were: cultural background and personal goals. The former supports the findings of Zhang et al. (2020a), which demonstrate that knowing cultural biases can significantly help in food classification tasks. That said, the use of current location may not specifically afford more precise food recommendations (Ramirez-Garcia & García-Valdez, 2014). Instead, home-country and ethnic origin may a more useful feature to incorporate into the food recommender system and adjust the recommendation result, giving individuals a taste of home, no matter where they are in the world.

In terms of users with clear personal dietary goals, the desire for diversity in eating choices was relatively low; for such users the recommendation should be better target monitoring of nutritional information, and offering healthier alternative ingredients.

The most influential dynamic factors for RQ2 were emotions, busyness, seasons, physical activities, and sustainability. These are novel influential contextual factors for food recommender systems, as no prior research has integrated these features into FRS to enable adaptable contextual recommendations. Users' food choices and intake behaviours present sizable changes under these factors; emotions and busyness in particular shape individual's food intake behaviour. These identified factors support established findings for the circularity of emotions and eating (Macht, 2008) and that emotional and stress-induced eating have been linked as behavioural mechanisms in depression and development of obesity (Torres & Nowson, 2007; Konttinen, 2020). Given that FRS could influence people's daily lives, it becomes paramount to consider such emotional factors and food's role in regulating these.

Where research has indicated seasonality in people's recipe development, cooking and eating behaviour (Kusmierczyk et al., 2015a), this study further demonstrates the relatively distinct emotional, busyness-related, seasonal patterns. Therefore, the integration of these features into a recommender system may provide the potential to enhance traditional collaborative filtering and content-based recommendation results; this integration might offer algorithms more prior information, enabling them to uncover hidden (e.g., temporal) patterns and provide personalised recommendations more effectively (Cheng et al., 2017; Maia & Ferreira, 2018).

Further tailoring of recommendations could account for interaction effects across factors. For example, for people increasing protein intake after physical activities, recipe recommendations with high protein and low cooking effort may be especially appealing - they can replenish energy quickly despite being tired. Alternatively, food sustainability has garnered increasing attention (Lajoie-O'Malley et al., 2020; El Bilali et al., 2019; Garcia et al., 2020); as individuals become more aware of the 'food waste problem', environmentally sustainable considerations and moderation of consumption should be taken into account in the design of the next generation of recommender systems.

As for RQ3, the findings further support work indicating that people engaging in high levels of physical activity tend to take greater care on their food and nutrient intakes (Raine, 2005). Otherwise, when people facing problems with their health tend to care more about nutritional information; this outcome is consistent with research indicating people may adjust diets for health reasons over short periods (Devine, 2005; Lee et al., 2019). Similarly, this study identified that unhealthy eating habits are hard to break and habits from childhood establish pre-existing preferences. This aligns with research on the development of eating preferences of Wolstenholme et al. (2020). In terms of establishing good eating habits, it may be difficult for typical families to incorporate scientific knowledge and guidance in prompting the families' healthy eating behaviour.

In practice, busyness can shape people's food choices and increase calorie intake via opting for food delivery or pre-prepared meal (Dixon et al., 2014; Kalenkoski & Hamrick, 2013). This may be compounded by feelings of stress or other negative emotions associated with busyness, that further prompt unhealthy food choices (Adriaanse et al., 2011). Interactions across multiple contexts that can shape healthy eating behavior underscores the importance of considering these in health-oriented recommendations.

In regard to RQ4, the findings highlight that online recipe searches are highly susceptible to influence from the surrounding environmental context. The visual presentation of a recipe has the potential to significantly impact individuals' real-life cooking experiences and intentions. The presence of disliked ingredients emerges as the primary factor leading to the rejection of a recipe. Thus, constructing a multimodal feature matrix may substantially enhance the likelihood of a recommendation being accepted (Liu et al., 2023; Zhang et al., 2023).

As for RQ5, this research sheds light on several key aspects when designing the next generation of food recommender systems. Systems built with contextually relevant information embedded throughout development-such as the identified influencing factors-may better account for, explain, and even predict user recipe ratings made under specific circumstances (Tran et al., 2021). For example, a person may rate a meal as having low appeal, not because of the meal itself, but because it needed more preparation time than their busy context afforded. The development of an emotion and busyness-aware persuasive healthy food recommender system becomes increasingly relevant in providing useful and appealing recommendations (Trattner et al., 2017a). By further understanding and incorporating how participants' health intake behaviour can change, responsive recommendations may even lead to less rejection of the recommendation and achieve the aim of cultivating healthier eating (Cohen & Babey, 2012).

Furthermore, the findings revealed that offering users more control over the system - filter and re-rank recommendation results based on factors like sugar level, calorie content, or cooking complexity - could greatly enhance the system's potential usability. Additionally, a system that provides diverse, context aware recommendations and offers explanations for these would significantly improve the trustworthiness of the system (Trattner et al., 2017a).

Chapter 5

The Effect of Simulated Contextual Factors on Recipe Rating and Implied Nutritional Intake Behaviour

5.1 Introduction

This chapter presents findings from the second stage of the research: a large-scale experimental study on understanding how contextual factors influence individuals' online recipe rating and implied nutritional intake behaviours. The focus is on assessing whether significant differences emerge across different simulated contextual situations. Furthermore, the analyses preliminary explored the potential benefits of integrating contextual factors into the prediction model to enhance its performance. The chapter also investigates whether applying a pre-filtering technique, based on dynamic contextual groups, improves the performance of the rating prediction model compared to a generic (context-free) approach.

The chapter is structured around addressing four research questions:

- RQ 2.1: Do people's recipe rating behaviour vary among different simulated contextual situations?
- RQ 2.2: To what extent do contextual factors affect people's implied nutritional intake behaviour?
- RQ:2.3 Can integration of these contextual factors improve recommendation performance?
- RQ:2.4 Which contextual factors are the most influential factors when recommending foods?

First, the descriptive statistics of the dataset collected from the experimental study will be presented in Section 5.2. Then, in Section 5.3, the results of hypothesis testing using one-way

ANOVAs and Tukey's HSD tests will be discussed. Next, the comparison between a baseline model and a model integrating contextual features will be discussed in Section 5.4, along with an analysis of the pre-filtering contextual scenarios group model against the generic group model. Finally, the discussion will be presented.

5.2 Descriptive Statistics and Data Visualisation

Following the data cleaning and preprocessing steps discussed in the Methodology section 3.4.7, Table 5.1 presents the resulting cleaned dataset. The dataset includes 397 individual participants and 75 unique recipes, with a total of 11,910 ratings collected. These ratings were provided under seven simulated contextual scenarios as well as one context-free condition. The overall mean rating is 2.927, with a standard deviation of 1.417.

Table 5.1: *Dataset statistics*

#User	#Item	#Dynamic context	No. of Ratings	No. of Ratings/user	Mean rating	Std. Dev. rating
397	75	7+1	11910	30	2.927	1.417

Before developing a rating prediction model, it is crucial to thoroughly explore and understand the dataset, such as how data distribute within each group, and whether notably patterns exists between features (Vigni et al., 2013). Given that recipe ratings were provided under seven distinct contextual situations and one context-free setting, it is insightful to examine whether recipe preferences and recipe rating behaviour vary across these contexts. This analysis will provide valuable insights into whether incorporating contextual factors could potentially improve the model's performance.

After grouping the data by contextual scenario, ratings lower than 4 were filtered out, and the remaining recipes were considered preferred and likely to be consumed within the given context. For each contextual scenario, the frequency with which each recipe received a rating above 3 was recorded. This facilitates tracking the degree to which each recipe was favoured, while also enabling the identification of the most popular recipes across different contexts. As shown in Figure 5.1, the top 10 most popular recipes are presented. It is evident that certain recipes are consistently favoured; however, the most popular recipes vary across different contexts.

During hot summer days, lighter options such as fresh spring rolls, veggie stir-fries, smoothies, and fruit salads were popular. In contrast, on cold winter days, heartier dishes like potato and taco soups were favoured, possibly because of the warmth these recipes provide. Additionally, heavier foods such as dumplings, roast beef, and more complex recipes like lasagna were also popular during the winter. Unsurprisingly, cakes such as cheesecake, layer cake, and chocolate cake were preferred when people felt happy. Seafood chowder and soup were popular when people felt sad, possibly because soup serves as a great comfort food. During busy and stressful weekdays, bread pudding, zucchini bread, and Shepherd's pie were popular, likely due to more carbohydrates were needed during these times, and these recipes require less effort to prepare. Notably, two snacks - panna cotta and pecan treats - were popular during relaxing weekends, which may indicate that people had more free time to enjoy and prepare such snacks. After physical activities, dishes like ribs, Zuppa Toscana, and eggs were preferred, possibly due to the need for high protein, as well as carbohydrates like zucchini bread and sandwiches. In the generic (i.e., context-free) group, no clear pattern of preference emerged.

To gain deeper insights into how each recipe was preferred or disliked across the simulated contextual situations, a line chart was provided to visualise whether the likelihood of preference for each recipe varied across different context. Due to the variability in participant numbers under each contextual situation, and the fact that each participant rated only 30 out of the 75 available recipes, the number of ratings each recipe received was not uniform. To ensure a fair comparison, the mean rating of each recipe within each contextual scenario was calculated. Figure 5.2 illustrates that preferences for these recipes are not consistent across all contextual circumstances, as they are relatively varied. In certain contexts, some recipes were preferred over others. For example, recipe number 61, representing “Slow Cooker Chicken Taco Soup”, was generally preferred in the generic group (mean rating = 3.5) but was highly favoured during a “cold winter’s day” (mean rating = 4.1) and disliked on a “hot summer’s day” (mean rating = 2.8) and “busy weekday” (mean rating = 2). Another example is recipe number 13, ’Buffalo Style Chicken Pizza.’ While this recipe was generally not favoured, receiving a mean rating of only 2.75 in the generic group, it suddenly became popular during a ’busy and stressful weekday,’ achieving a significantly higher mean rating of 4.24.

Figure 5.3 shows the rating distributions across each simulated contextual scenarios. Particularly noteworthy is the rating distribution in the generic baseline group (G), which displays low variance, with most ratings falling within the range [2.75, 3.25]. In contrast, the ratings for the ‘cold winter day’ (CWD) and ‘hot summer day’ scenarios exhibit considerably more dispersed distributions. The rating distribution for the busy weekday contextual scenario (B) is notably lower than those of the other scenarios. Surprisingly, the score distributions of the groups experiencing happy and sad emotions appear similar; however, the sad emotion group demonstrated a marginally higher mean rating range when compared to the happy emotion group.

5.3 Hypothesis Testing

Since the ANOVA test assumes that the data follows a normal distribution (Field, 2024), it is important to assess the normality of the data to determine if this assumption holds. To assess the data distribution, the mean rating distribution for each recipe under each contextual scenario is presented in Figure 5.4. It is evident that the mean ratings differ across the groups. Notably, the distributions for the cold winter day, sad emotion, busy weekday, and generic groups exhibit a bell-shaped curve, suggesting an approximate normal distribution.

Further assessment of the normality of the residuals was conducted using the Shapiro-Wilk test (see Table 5.2). The p-values for seven of the groups were found to be non-significant, indicating that the data was drawn from a normal distribution. However, note that the data for the ‘after physical activities’ group exhibited a significant departure from normality. We, however, proceed under the assumption that all the data are normally distributed, as this is

Figure 5.1: *Most popular recipes across each contextual scenario group*

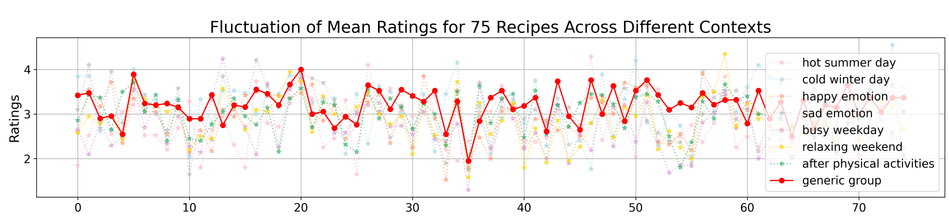


Figure 5.2: *Fluctuation of Mean Ratings for 75 Recipes Across Different Contexts*

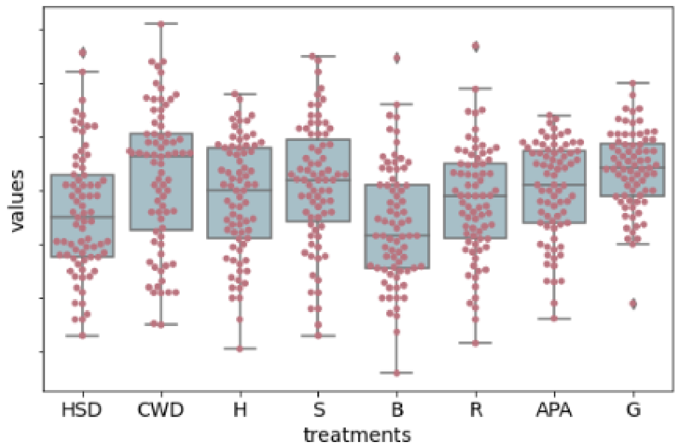


Figure 5.3: *The data distribution between each contextual scenario group (The above abbreviation stands for information below. HSD: hot summer day (contextual group), CWD: cold winter day (contextual group), H: happy (contextual group), S: sad (contextual group), B: busy (contextual group), R: relax (contextual group), APA: after physical activities (contextual group), G: generic group)*

common practice when dealing with data derived from Likert scales.

5.3.1 Examining the Influence of Contextual Factors on Recipe Rating Behaviours

To address RQ1, hypothesis testing was initiated. The null hypothesis (H0) posited that there are no significant differences (no variation in means) among the contextual scenario groups.

A one-way ANOVA test was conducted to explore whether variations existed in individuals' implied eating behaviours and recipe rating responses across the different contextual scenarios. The result revealed that there were significant differences between the groups (F(7, 592)=7.564, $p \ll 0.001$), allowing us to reject the null hypothesis. As the ANOVA test doesn't test the relationships between each group, multiple pairwise comparisons were subsequently conducted using Tukey's Honest Significant Difference (HSD) test. The outcomes of Tukey's test are shown in Table 5.3.

Participants exhibited notable distinctions in recipe preferences when compared with the generic group in certain contextual situations. These included hot summer day (HSD; $F(7, 592) = 6.451$, $p = 0.001$, $\mu_{G1} = 2.795$, $\mu_{G2} = 3.196$), busy weekdays (B; $F(7, 592) = 8.311$, $p = 0.001$, $\mu_{G1} = 2.679$, $\mu_{G2} = 3.196$), and relaxing weekends (R; $F(7, 592) = 4.889$, $p = 0.014$, $\mu_{G1} = 2.892$, $\mu_{G2} = 3.196$). In contrast, the remaining contextual scenarios did not exhibit statistically significant variations in mean ratings.

As expected, participants in the 'hot summer day' group and 'cold winter day' group shows significant difference in recipe preference ($F(7, 592)=5.304$, $p=0.005$, $\mu_{G1} = 2.795$, $\mu_{G2} = 3.124$). More interestingly, the results show that, during busy weekdays, implied eating behaviour and recipe preference significantly diverged from that observed during 'cold winter

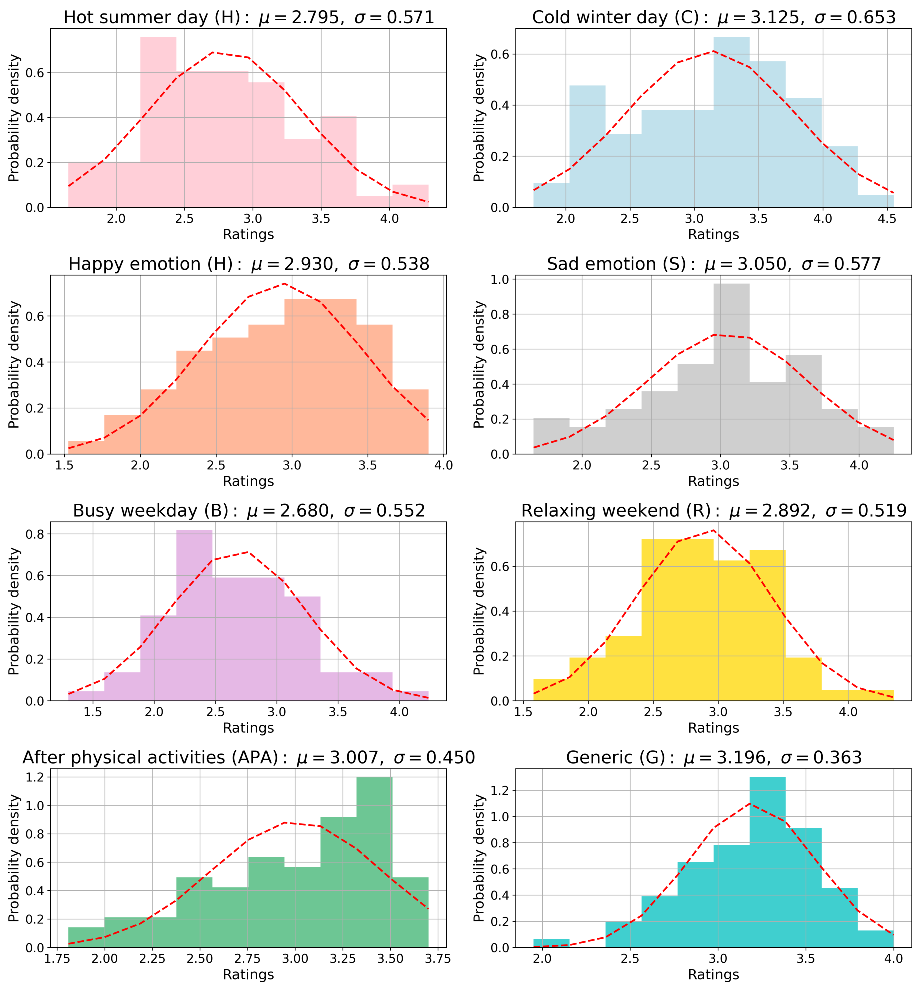


Figure 5.4: *Distribution of mean ratings for 75 recipes across each contextual scenario group*

day' (F(7,592)=7.165, p=0.001, $\mu_{G1} = 3.124$, $\mu_{G2} = 2.679$), as well as during states of 'sad emotion' (F(7, 592)=5.958, p=0.001, $\mu_{G1} = 3.049$, $\mu_{G2} = 2.679$) and 'after physical activities' (F(7, 592)=5.277, p=0.005, $\mu_{G1} = 2.679$, $\mu_{G2} = 3.007$).

5.3.2 Examining the Influence of Contextual Factors on Implied Nutritional Intake Behaviour

The determination of recipe health levels primarily relies on the FSA standard (Fsa, 2016) in this section, as this is one of the most appropriate standard for evaluating a single recipe. The other two standards focus on measuring appropriate daily nutritional intake. In alignment with this standard, a higher FSA score signifies a less healthy recipe, with scores ranging from 4 (extremely healthy recipe) to 12 (extremely unhealthy recipe). After data aggregation, it was observed that recipes with the highest FSA score of 12 were remarkably popular. In fact, such recipes garnered the highest mean ratings among six contextual scenario groups:

Table 5.2: *Shapiro-Wilk test normal distribution result*

SW value	HSD	CWD	H	S	B	R	APA	G
statistic	0.983	0.972	0.973	0.986	0.989	0.994	0.945	0.981
p-value	0.409	0.095	0.115	0.591	0.778	0.984	0.003	0.356

Note: The abbreviations are the same as in Figure 5.3.

Table 5.3: *Tukey's Honest Significant Difference (HSD) test results*

G1	**G2**	**Diff.**	μ_{G1}	μ_{G2}	**Conf. Int.**	**q**	**p**
HSD	CWD	0.329	2.795	3.124	[0.062, 0.597]	5.304	0.005
HSD	G	0.401	2.795	3.196	[0.134, 0.668]	6.451	0.001
CWD	B	0.445	3.124	2.679	[0.178, 0.712]	7.165	0.001
S	B	0.370	3.049	2.679	[0.103, 0.637]	5.958	0.001
B	APA	0.328	2.679	3.007	[0.061, 0.595]	5.277	0.005
B	G	0.517	2.679	3.196	[0.249, 0.784]	8.311	0.001
R	G	0.304	2.892	3.196	[0.036, 0.571]	4.889	0.014

Note: G1 = Group1, G2 = Group2, q = q-value, p = p-value. The group abbreviations are the same as in Figure 5.3.

'hot summer day' (mean rating of 3.21), 'happy emotion' (3.35), 'sad emotion' (3.53), 'busy weekday' (3.57), 'relaxing weekend' (3.75), and 'after physical activities' (3.7), as well as the 'generic' group (mean rating of 3.89), see Table 5.4. It is notable that, across most groups, recipes with better health ratings tend to have been given lower scores than their less healthy counterparts.

In the generic group the mean rating remained relatively consistent, with only minor fluctuations (ranging from 3.144 to 3.889 score), suggesting that, in the absence of contextual factors, individuals' recipe preferences don't result in significant changes in health outcomes. However, in the 'after physical activities' group, healthy recipes were generally preferred over less healthy ones. Conversely, 'during relaxing weekend' unhealthy recipes are more popular. Similarly, both 'happy' and 'sad emotion' groups show a preference towards unhealthy recipes, particularly in the case of the 'sad emotion' group.

A One-way ANOVA test was conducted once again to assess whether preferences for healthy recipes show significant differences between the contextual scenario groups. The results indicate significant differences among the groups (F(7, 64) = 4.942, p$\ll$0.001). The Tukey's HSD tests were subsequently conducted for pairwise comparisons (See Table 5.5), which revealed several significant differences. Preferences for healthy recipes during 'hot summer day' showed significant differences compared to 'cold winter day' (F(7, 64) = 5.094, p = 0.013, $\mu_{G1} = 2.722$, $\mu_{G2} = 3.315$) and during 'feeling sad' (F(7, 64) = 4.552, p = 0.040, $\mu_{G1} = 2.722$, $\mu_{G2} = 3.252$), as well as the 'generic' group (F(7, 64) = 5.159, p = 0.012, $\mu_{G1} = 2.722$, $\mu_{G2} = 3.323$). Preferences for healthy recipes during 'busy weekday' exhibited significant differences compared to 'cold winter day' (F(7, 64) = 5.525, p = 0.005, $\mu_{G1} = 3.315$, $\mu_{G2} = 2.672$) and while 'feeling sad' (F(7, 64) = 4.983, p = 0.017, $\mu_{G1} = 3.252$, $\mu_{G2} = 2.672$), as well as the 'generic' group (F(7, 64) = 5.589, p = 0.005, $\mu_{G1} = 2.672$, $\mu_{G2} = 3.323$). Furthermore, 'after physical activities' displayed significant differences in preferences for healthy recipes when compared to 'busy weekday' (F(7, 64) = 4.707, p = 0.030,

$\mu_{G1} = 2.672$, $\mu_{G2} = 3.220$).

Additionally, the characteristics of the favoured recipes were analysed. The most popular recipes, defined as those receiving ratings of 4 or 5 in each contextual situation, were extracted for further investigation. These recipes were then aggregated based on recipe categories (e.g., Main dish, Soup, Salad and Dessert/Snack) and three grouped levels of FSA scores: Low (FSA levels 4,5 and 6), Medium (FSA levels 7, 8 and 9), and High (FSA levels 10, 11 and 12), after Starke et al. (2021) (See Figure 5.5). It has been found that during 'cold winter days' and 'after physical activities', people tend to prefer main dishes, most of which fall into the (Medium) health category. Desserts and snacks are favoured when people are 'feeling sad' and 'feeling happy', with many of these items belonging to the unhealthy (High) food category. Predictably, salads are preferred during 'hot summer days' as they are relatively healthy compared to other categories. Soups are more popular during 'cold winter days' and when people are feeling sorrow, suggesting that soups might be an effective comfort food. Most soups fall into the healthy (Low) and general (Medium) health categories. In general, emotional changes may lead to increased consumption of unhealthy recipes.

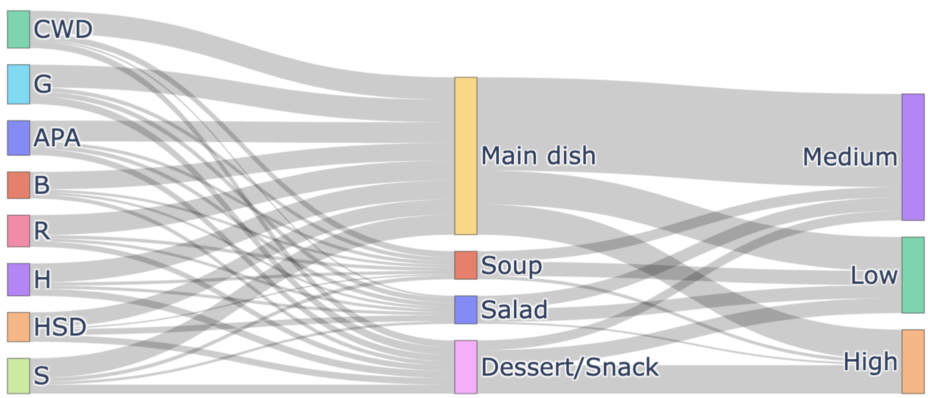


Figure 5.5: *Sankey Diagram of contextual impact on likelihood of recipe categories and grouped FSA health levels. Low (FSA levels 4,5 and 6), Medium (FSA levels 7, 8 and 9), and High (FSA levels 10, 11 and 12).)*

Table 5.4: *Recipe mean rating among different FSA health levels for each contextual scenario group*

FSA	$HSD\mu r$	$CWD\mu r$	$H\mu r$	$S\mu r$	$B\mu r$	$R\mu r$	$APA\mu r$	$G\mu r$
4	2.250	3.450	2.684	3.100	2.579	2.682	3.316	3.318
5	3.048	2.333	3.273	2.950	2.714	2.800	3.450	3.158
6	2.450	3.100	2.905	2.750	2.700	2.789	3.048	3.200
7	2.605	3.333	3.051	3.075	2.711	3.056	3.301	3.244
8	2.714	3.350	2.850	3.105	2.550	2.842	2.900	3.200
9	2.657	4.025	3.303	4.178	2.284	3.159	3.489	3.419
10	2.949	3.167	3.316	3.364	2.597	3.071	2.881	3.144
11	2.620	3.084	3.167	3.225	2.345	2.629	2.900	3.338
12	3.211	4.000	3.350	3.526	3.571	3.750	3.700	3.889

Note: The abbreviations are the same as in Figure 5.3.

Table 5.5: *Tukey's Honest Significant Difference (HSD) test result demonstration on preference of healthy recipes (only statistically significant results are reported)*

G1	G2	Diff.	μ_{G1}	μ_{G2}	Conf. Int.	q	p
HSD	CWD	0.593	2.722	3.315	[0.077, 1.109]	5.094	0.013
HSD	S	0.530	2.722	3.252	[0.014, 1.046]	4.552	0.040
HSD	G	0.601	2.722	3.323	[0.085, 1.117]	5.159	0.012
CWD	B	0.643	3.315	2.672	[0.127, 1.159]	5.525	0.005
S	B	0.580	3.252	2.672	[0.064, 1.096]	4.983	0.017
B	APA	0.548	2.672	3.220	[0.032, 1.064]	4.707	0.030
B	G	0.651	2.672	3.323	[0.135, 1.167]	5.589	0.005

Note: The abbreviations are the same as in Table 5.3 and Figure 5.3.

5.4 Preliminary Development of Rating Prediction Model

The following section reports the results of the preliminary contextual modelling and pre-filtering rating prediction models. Section 5.4.1 addresses whether integrating contextual features leads to improvements in model performance, while Section 5.4.2 explores which contextual elements are the most influential when developing a rating prediction model using the pre-filtering approach.

5.4.1 Evaluating the Influence of Contextual Factor Integration on Recommendation Performance

In this study, it is worth highlighting that the task of rating prediction was approached as a regression problem, specifically utilising the XGBoost regression model. Details of the model selection and its justification can be found in the Methodology section 3.4.9. Model evaluation involved the calculation of several key metrics, including Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and R-squared (R2). The primary comparison entailed assessing models trained with the inclusion of all available features (contextual scenario feature included) against a baseline model that excluded contextual features. This comparative analysis aimed to identify the impact of the contextual factors on model performance. Additionally, the top 5 most important features were identified to discern the most influential variables for rating prediction.

The XGBoost model employed a comprehensive set of 29 features encompassing demographic attributes (e.g., age, gender, ethnic origin, physical activity level, survey ID), recipe-related details (e.g., recipe name, category, nutritional information), information regarding cooking (e.g., cooking frequency, skill level), recipe health level (WHO, FDA, FSA and WHO_adj, FDA_adj), and the contextual scenario information. Since each health standard emphasises a slightly different nutritional aspect, which may potentially impact model performance, all of them were included in the model-training process. As demonstrated in Table 5.6, the model utilising the complete feature set exhibited superior performance compared to the model without contextual scenario features. The all-feature model yielded an MSE of 1.621, RMSE of 1.273, and MAE of 1.048, all of which outperformed the baseline model (MSE 1.681, RMSE 1.296, and MAE 1.067). The all-feature model achieved a higher R2 score (0.208) compared to the baseline (0.179), signifying a superior goodness of fit.

More importantly, the feature importance analysis revealed that the contextual scenario features were the most significant among the 29 features considered, as evidenced by Figure 5.6. This underscores the impact of *Contextual information* on both human implied behaviour and algorithmic comprehension thereof. The incorporation of contextual factors enhances the capacity of the model to intelligently discern and uncover the hidden relationships within individuals' preferences in the dataset. This, in turn, leads to more precise rating predictions. *Cooking skill level* may influence recipe choice in a straightforward manner, making it a prominent feature in determining preferences. Individuals with advanced cooking skills may explore a wide variety of recipes, while those with lower cooking skills might prefer to avoid challenging or complex dishes, opting instead for simpler options. *Home country code* emerges as the third most important feature in influencing recipe choice. This finding aligns with the results from the first stage of the study, which highlighted the lasting impact of cultural and familial food traditions. The cuisine and meal preferences ingrained during childhood, particularly those tied to family meals and local dishes from one's hometown, continue to play a significant role in shaping individuals' food choices well into adulthood. Surprisingly, *gender* emerges as the fourth most important feature, suggesting potentially substantial differences in rating and dietary preferences between men and women. The importance of the *Survey ID* feature lies in its role in identifying and distinguishing an individual's implied eating and rating behaviour. This feature is commonly utilised in traditional matrix factorisation recommender systems, and so it comes as no surprise that it ranks as the third most important feature.

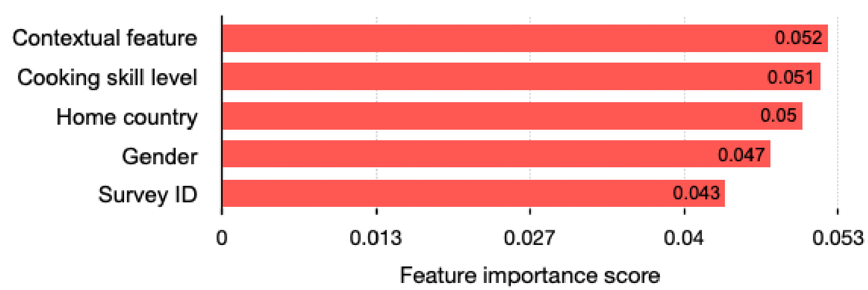


Figure 5.6: *XGBoost feature importance of all feature model*

Table 5.6: *XGBoost test results summarisation and comparison of all feature model*

	MSE	RMSE	MAE	R2
All feature model	**1.621**	**1.273**	**1.048**	**0.208**
Baseline model	1.681	1.296	1.067	0.179

5.4.2 Determining Influential Contextual Factors Through a Pre-Filtering Approach

The XGBoost model was further employed in a pre-filtering approach to predict ratings segmented by contextual scenario, investigating which group led to better model performance. The results show that using data from the 'happy emotion' group achieved the highest model performance (MSE=1.615, RMSE=1.271, MAE=1.032, and R2=0.185), as shown in Table 5.7. Data from 'hot summer day', 'cold winter day' 'happy emotion', 'sad emotion' generally performed better than the other groups. When compared to the generic baseline group and considered from a model fitting perspective, the datasets from all simulated contextual scenario groups achieved higher R2 values. There was a relatively lower performance observed in the models for the 'busy weekday', 'relaxing weekend', 'after physical activities' groups compared to other contexts. This may be due to the fact that, even though an attempt was made to specify a single contextual scenario to the participants, other potential independent factors might still influence their preferences in that scenario. For example, during a relaxing weekend, one may also feel happy, and after physical activities, one may feel both tired and energetic simultaneously.

The demographic factors, such as the *time spent living in one's current country of residence, gender, home country, currently living country, and cooking skill level*, are frequently shown to be the most important factors in each contextual scenario group, as indicated in Table 5.8. Remarkably, the *FDA* and *FDA_adj* health level show as the most important factor in the 'cold winter day' 'sad emotion' and 'busy weekday' groups. WHO health level shows as the third most important feature in the 'happy emotion' group. This may indicate that, under these contextual situations, the healthiness of the recipe may potentially drive or shape people's food choices. This could imply that people may unconsciously be even more prone to choosing relatively unhealthy recipes over healthy ones. Both the *WHO health level* and *FSA health level* emerge as the most important features in the 'after physical activities' group. This may suggest that people might consciously prefer healthier recipes after physical activity, making it easier to predict favoured recipes when combined with nutritional health indicators. *Salt, saturates and fat* are identified as important features in the 'relaxing weekend,' 'cold winter day,' 'sad emotion,' and 'happy emotion' groups. This may be attributed to a preference for larger meals and snacks during relaxing weekends, which often contain higher levels of calories, sugar, and salt. Similarly, on cold winter days, hearty main dishes are favoured, likely leading to an increase in calorie and saturated fat consumption. Both happy and sad emotions may drive a tendency to consume higher-calorie, less healthy foods.

These analyses further supports the earlier suggestion that the contextual scenario may be capable of acting as a moderator of other predictive features in recommender system models, providing additional prior information and potentially leading to improved model performance. Although each individual model for the contextual groups did not perform particularly well, this may be due to the limitations of the pre-filtering approach (Ricci et al., 2015). After splitting the data, only a limited number of ratings remained in each group (ranging from 1,470 to 1,500 ratings). With a larger dataset, model performance is expected to be improved.

Table 5.7: *XGBoost model test results summarisation and comparison among each contextual scenario group*

	HSD	CWD	**H**	S	B	R	APA	G
MSE	1.821	1.897	**1.615**	1.747	2.043	1.929	2.090	1.755
RMSE	1.350	1.377	**1.271**	1.322	1.429	1.389	1.446	1.325
MAE	1.071	1.119	**1.032**	1.049	1.159	1.132	1.173	1.043
R2	0.133	0.071	**0.185**	0.067	-0.089	-0.022	-0.098	-0.101

Note: The abbreviations are the same as in Figure 5.3. Under certain contexts, the fitted model performed worse than the null model and achieved a negative R2. In such situations, the average rating would be used for rating prediction.

Table 5.8: *Summary of top 5 most important features for each contextual scenario model*

	Features
HSD	current living country, cooking skill level, gender, online recipe searching frequency, current country living duration
CWD	FDA_adj health level, saturates, current country living duration, category, cooking frequency
H	home country, gender, WHO health level, cook book use frequency, saturates
S	FDA health level, fat, ethnic origin, cook book use frequency, home country
B	FDA_adj health level, current living country, home country, cooking complexity, WHO_adj health level
R	saturates, online recipe searching frequency, total weight per portion, sugar, salt
APA	WHO health level, gender, home country, cooking skill level, FSA health level
G	current living country, ethnic origin, cooking complexity, sugar, cooking frequency

Note: The abbreviations are the same as in Figure 5.3.

5.5 Discussion

This study investigated whether people's recipe rating and implied eating and nutritional intake behaviour changed under different simulated contextual situations. Additionally, the study examined whether integrating contextual features could improve model-based recommendation performance and identified which features were the most important in each contextual scenario.

The results indicate that people's eating preference and the likelihood of consuming healthy recipes during busy weekdays differ significantly from other contextual situations. This difference is supported by the results of a one-way ANOVA and Tukey's HSD test, which showed a significant variation in mean ratings during busy weekdays compared to situations such as cold winter days, sad emotions, after physical activities, and the generic group. The demanding and busy work schedule during weekdays often leaves individuals with limited time for cooking. This constraint may restrict their freedom to think and choose preferred food or recipes (Pinho et al., 2018). In this situation, the primary goal of cooking and eating becomes satisfying hunger, and meals should preferably be prepared and completed quickly. Consequently, recipes that are easy and quick to make, often involving refined or processed products and other potentially less healthy options, are preferred during busy weekdays. These findings on distinct implied eating behaviour during busy weekdays align with the research conducted by Pinho et al. (2018), which suggests that a hectic lifestyle may lead to reduced consumption of vegetables and home-cooked meals.

In addition, it is worth considering that the varying stress levels associated with a busy lifestyle, as supported by Hyldelund et al. (2022), may result in individuals exhibiting a shift in their dietary choices, characterised by a reduction in main meals and an increase in snack consumption. This phenomenon could also explain the observed significant differences, such as those evident when comparing busy weekdays with cold winter days. In the latter case, individuals may gravitate towards carbohydrate-rich options, such as main dishes, potentially contributing to this distinction in eating preferences.

As anticipated, there is a significant difference in recipe preference between 'hot summer's days' and 'cold winter days', providing evidence for the seasonality of food preferences over time (Spence, 2021). Furthermore, the analysis of recipe health levels revealed a notably higher consumption of main dishes during the 'cold winter day' contextual group. This aligns with the findings of Capita & Alonso-Calleja (2005), who concluded that both men and women tend to consume more energy during the winter months.

Surprisingly, the Tukey's test results did not show significant differences in recipe rating behaviours between participants in the happy and sad emotion groups. Individuals appeared to exhibit similar food preferences, even under these extremely different emotional situations. Under both of these scenarios, participants showed an increased demand for unhealthy food (Bartkiene et al., 2019), relative to the general condition. However, there may be a limitation in the ability of the external emotion stimuli method to effectively evoke such strong and polarised emotions.

Perhaps unsurprisingly, individuals prioritised nutritional information and opted for relatively healthier recipes after engaging in physical activities (Aguiar-Bloemer & Diez-Garcia, 2018). This may indicate that social norms and health-conscious behaviours play a significant role in influencing food choices after physical exertion.

Beyond its theoretical implications, this study offers a novel perspective on the development of context-aware food recommender systems. Previous research has predominantly focused on location-aware or gender-aware food recommender systems (Al-Ghobari et al., 2021; Rostami et al., 2022). In this algorithmic experiments, contextual features emerged as the most influential features, leading to improved accuracy in rating predictions. Importantly, the datasets within each contextual scenario group exhibited higher R2 values compared to the baseline group. This is likely due to the baseline group encompassing a wide range of random possibilities, making it challenging for the model to discern meaningful patterns. In contrast, the inclusion of contextual features provides the model with a clearer direction for uncovering hidden patterns. Notably, home country code and gender also emerged as key factors influencing recipe choice. This finding aligns with the results from the first stage of the study, as well as the work of Wolstenholme et al. (2020), which emphasized the lasting impact of cultural and familial food traditions. Additionally, this finding is consistent with Cavazza et al. (2015), who found that men and women may express different food preferences, with women typically favouring smaller, elegantly presented meals, while men often prefer larger, more robust dishes.

Based on the experimental results of the pre-filtering approach, the findings, therefore, suggest that emotion-aware systems could represent the next generation of food recommender systems, as highlighted by (Raza & Ding, 2019). This could also be combined with season- and stress level-awareness. While the results are promising, they require validation on a larger, more naturalistic dataset for practical implementation. As the analysis of model performance and feature importance continues, nutritional information frequently emerged as an influential factor. This may indicate a clear nutritional dominance (with individuals possibly preferring less healthy recipes) when selecting recipes under certain contextual situations, particularly during busy weekdays and sad emotions, as incorporating this information contributed to improved model performance. This highlights the need to provide healthy recommendations that balance the recipe preference and nutritional needs (Harvey et al., 2013).

Previous food recommendation systems have primarily focused on either context-aware or healthy recommendations (Rostami et al., 2022; Starke & Trattner, 2021). This research has demonstrated examples where the contextual scenario acts as a moderator, allowing other features to perform better than they would without the context as a precedent. For example, during busy weekdays, there was a noticeable increase in the consumption of unhealthy food. Addressing how to incorporate nutritional information into recommender systems to encourage healthy eating habits during hectic lifestyles could present a novel approach to balancing the trade-offs involved in healthy food recommendations. Currently, such systems have not been proposed in the field of food recommendation.

Chapter 6

Incorporating Knowledge of Contextual Situations for Rating Prediction and Healthy Food Recommendations

6.1 Introduction

Building upon the previous stage of the research, experimentation was continued with the aim of proposing more robust and better-performing food and healthy food recommender systems. This chapter presents the results of developing advanced contextual modelling algorithms by integrating multi-modal features, including recipe ingredients, cooking directions, recipe images, nutritional information, and user demographic data, alongside contextual factors. These features were incorporated into three commonly used machine learning algorithms (XGBoost, Ridge Regression, and SVR) and a deep learning algorithm (MLP). The optimal combination of these features, which led to the best model performance, was identified. A preliminary one-stage contextual modelling approach is demonstrated in Section 6.3.1.

To further enhance performance, a proposed two-stage model is discussed in Section 6.3.2, which integrates the decomposed features of traditional RS algorithms, such as SVD and NMF, as key inputs into the model learning process. This approach incorporates preference-priority information aims to improve the accuracy of user rating predictions. Moreover, to address the healthiness-tastiness trade-off challenge, a novel approach was introduced in Section 6.3.3 to deliver contextually-appropriate healthy recommendations, aligned with healthiness-tastiness evaluation methods. This chapter will conclude with a discussion and implications of the main findings in Section 6.4.

This study mainly addressed the following research questions:

- RQ 3.1: Will integrating dynamic contextual features enhance performance across all models?

- RQ 3.2: Which combination of features, when integrated with contextual factors, yields superior performance?
- RQ 3.3: Which model architecture and feature combination achieve optimal rating prediction accuracy?
- RQ 3.4: What strategies can be employed in building healthy recommendation models to effectively balance preferences and health considerations?

6.2 Experimental Settings

All models were implemented using the scikit-learn (sklearn) package in Python. Both the one-stage and two-stage approaches were ensured to utilise the exact same training and test datasets. To reduce the likelihood of overfitting caused by any particular data split and to provide a more reliable estimate of the model's performance, 5-fold cross-validation was applied across all four models during training (Wong & Yeh, 2019), using a random seed of 42 to ensure fair and reproducible results. Additionally, hyperparameter tuning was performed for each model to enhance performance and ensure robustness (see Table 6.1 for details). Given the varying foundations of each employed model, different methods were adopted to calculate feature importance. Information gain was used to report the feature importance for the XGBoost model (Elsweiler et al., 2017), while ridge coefficients (Rodríguez Sánchez et al., 2022) and dual coefficient (Ning et al., 2016) were used for ridge regression and SVM, respectively. The feature importance of the MLP model is more complex. Following the study by Egmont-Petersen et al. (1998), the importance of each input feature was calculated by examining the weights of the connections between the input and hidden layers. These weights were summed for each feature and the values were normalised to provide a relative measure of feature importance. The detailed processes for data preprocessing, embedding, and feature engineering can be referred back to in Methodology Section 3.5.

Table 6.1: *Details of hyperparameter optimisation for each model*

Models	Hyperparameters
XGBoost	"n_estimators": {50, 100, 200}, "learning_rate": {0.01, 0.1, 0.2}, "max_depth": {3, 4, 5, 6, 7, 8, 9, 10}
Ridge regression	"alpha": {0.1, 1.0, 10.0, 11.0, 12.0}
MLP	"hidden_layer_sizes": {(32,), (64,), (128,), (32, 32), (64, 64), (128, 128), (128, 64, 32)}, "activation": {'relu', 'tanh', 'logistic'}, "alpha": {0.001, 0.01, 0.1}, "solver": {'adam', 'sgd', 'lbfgs'}, batch_size: {512, 1024}, "learning_rate": {'constant', 'invscaling', 'adaptive'}
SVR	"kernel": {'linear', 'poly', 'rbf'}, "C": {0.1, 1, 10}, "epsilon": {0.1, 0.2, 0.5}, "gamma": {'scale', 'auto'}

6.3 Results

6.3.1 Contextual Modelling One-stage Approach

To address **RQ1** (Does integrating dynamic contextual features enhance performance across all models?) and **RQ2** (Which combination of features, when integrated with contextual factors, yields the best performance?), a comprehensive ablation study was conducted by systematically incorporating different sets of features into the models. The key rating prediction model test results are presented in Table 6.2. Full evaluation results can be found in Appendix Table G.1

It is evident that incorporating dynamic contextual features (cs) into the baseline User*Item (UI) model led to performance improvements across all models. The XGBoost model showed the most significant gains: RMSE decreased from 1.299 to 1.276, reflecting an improvement of approximately 1.8%; MAE decreased from 1.109 to 1.078, representing a 2.8% improvement; and R^2 increased by 20.3%, from 0.148 to 0.178, compared to the baseline UI model. For all four models, further integrating user demographic information (*UI+cs+udi*) generally resulted in better performance than incorporating basic recipe content and nutritional information (*UI+cs+nibri*).

Table 6.2: *Evaluation results of prediction experiments*

	XGBoost			**Ridge Regression**			**MLP**			**SVR**		
Feature set	**RMSE**	**MAE**	**R2**	**RMSE**	**MAE**	**R2**	**RMSE**	**MAE**	**R2**	**RMSE**	**MAE**	**R2**
Stage 1: baseline group												
UI (baseline)	1.299	1.109	0.148	1.408	1.203	-0.001	1.4	1.207	0.01	1.407	1.206	0.001
UI+cs	1.276	1.078	0.178	1.402	1.204	0.008	1.394	1.201	0.018	1.403	1.201	0.005
UI+cs+udi	1.256	1.056	0.203	1.374	1.183	0.047	1.315	1.111	0.127	1.352	1.154	0.077
UI+cs+nibri	1.262	1.054	0.195	1.375	1.18	0.045	1.362	1.163	0.063	1.366	1.171	0.057
Stage 2: UIC+recipe content feature set												
UI+cs+nibri+tfidfrn	1.287	1.076	0.163	1.363	1.168	0.061	1.35	1.142	0.08	1.368	1.166	0.055
UI+cs+nibri+tfidfingr	1.281	1.07	0.172	1.364	1.169	0.061	1.349	1.149	0.08	1.37	1.169	0.052
UI+cs+nibri+BERTembcd	1.269	1.06	0.187	1.366	1.172	0.058	1.357	1.15	0.07	1.367	1.17	0.056
UI+cs+nibri+if	1.288	1.085	0.162	1.363	1.169	0.061	1.36	1.158	0.066	1.383	1.179	0.034
UI+cs+nibri+Gloveembcd	1.273	1.062	0.182	1.366	1.172	0.058	1.352	1.146	0.076	1.366	1.166	0.057
UI+cs+nibri+tfidfcmm	1.262	1.054	0.196	1.37	1.175	0.052	1.353	1.149	0.076	1.368	1.17	0.055
Stage 3: UIC+user demographic information + recipe content feature set												
UI+cs+udi+nibri	1.228	1.02	0.239	1.347	1.155	0.084	1.293	1.074	0.155	1.291	1.089	0.159
UI+cs+udi+nibri+tfidfrn	1.227	1.02	0.24	1.336	1.142	0.099	1.286	1.086	0.165	1.31	1.101	0.134
UI+cs+udi+nibri+tfidfingr	1.247	1.046	0.215	1.336	1.142	0.099	1.283	1.072	0.169	1.311	1.103	0.133
UI+cs+udi+nibri+BERTembcd	1.229	1.023	0.237	1.338	1.146	0.096	1.286	1.083	0.165	1.301	1.096	0.145
UI+cs+udi+nibri+if	1.223	1.02	0.245	1.349	1.157	0.081	1.295	1.088	0.153	1.294	1.092	0.154
Best combination feature set												
UI+cs+udi+nibri+BERTembcd+if	**1.223**	**1.02**	**0.244**	**1.339**	**1.146**	**0.094**	**1.288**	**1.075**	**0.162**	**1.305**	**1.099**	**0.14**
Best combination feature set without contextual factors and contextual mediators												
UI+udi+nibri+BERTembcd+if	1.259	1.057	0.2	1.343	1.149	0.089	1.318	1.121	0.122	1.343	1.136	0.089
All combination feature set												
UI+cs+udi+nibri+tfidfingr+BERTembcd+if	1.243	1.039	0.219	1.336	1.142	0.099	1.284	1.075	0.168	1.323	1.115	0.116

Note: Abbreviations of feature groups shows as follows, UI: user_id and recipe_id, cs: contextual scenarios, udi: user demographic information, nibri: nutritional information and basic recipe information, tfidfrn:tfidf recipe name, tfidfingr: tfidf ingredients, BERTembcd: BERT embedding cooking direction, Gloveembcd: Glove embedding cooking direction, tfidfcmm: tfidf cooking methods matching, if: image features. UIC: UI+cs.

When different recipe content features were added, no consistent performance pattern emerged across all models. However, the XGBoost model consistently outperformed the other models within the *UIC+recipe content feature set* group. For the XGBoost model, the contribution of recipe content information was less significant compared to user demographic information. Among the various recipe content feature groups, UIC combined with cooking methods embedding outperformed ingredients, image, and recipe name feature sets. Notably, extracting the cooking methods and encoding this list as TF-IDF features demonstrated similar performance to BERT Sentence Transformer embeddings of the cooking directions. The Ridge regression model appears to be insensitive to recipe content features, as its performance does not vary much when different sets of recipe content feature groups are added. The SVM model, on the other hand, performs relatively worse on the rating prediction task. Interestingly, for the MLP model, recipe ingredients emerge as the most influential feature set. Since integrating recipe content information did not lead to an improvement in model performance, the next step involved experimenting with the addition of user demographic context information, combined with recipe content information, to explore the best combination of feature sets.

Based on the experimental results presented in Table 6.2, it is evident that combining user demographic information with recipe content features can lead to considerable improvements in the performance of all models. Specifically, combining user demographic features with image features and basic recipe and nutritional features under the XGBoost model yielded the best performance (RMSE=1.223, MAE=1.02, R^2=0.245). As different parts of the recipe content features were incrementally incorporated using a forward-step approach, combining UIC with user demographic features, basic recipe and nutritional information, BERT-embedded cooking directions, and recipe image features (*UI+cs+udi+nibri+BERTembcd+if*) was identified as the best feature combination, leading to the highest model performance. These feature sets may have the potential to provide more comprehensive prior information, enabling the model to better understand the preferences of different user groups and more effectively match them with the types of recipes they are likely to prefer. Compared to the *UI+cs+udi+nibri+if* feature combination, this combination demonstrates superior model generalization ability, as the test results show stronger performance in both Ridge Regression and MLP models compared to other feature combinations. Surprisingly, yet logically, image features emerged as significant contributors to model performance across multiple models. Notably, during the data collection process in this study, the recipe images were not shown to participants. This may indicate that when participants, upon encountering the recipe names or ingredients, may have been able to mentally visualise the corresponding recipe image. Their choices may have been influenced by these mental visualizations (Missbach et al., 2015). This cognitive process may potentially explain why image features influenced the model's performance and may have contributed to its improvement, as Muñoz-Vilches et al. (2020) suggest that imagining food may alter food desire at an implicit level. Additionally, these image features may capture underlying patterns or characteristics that correlate with participant ratings, these features could serve as "hidden context", enriching the model's ability to predict ratings more accurately.

It is worth highlighting that, contrary to expectations, combining all feature sets did not yield the best model performance in this study. One possible reason for this could be the limited size of the dataset. As the number of features increases while the number of instances remains constant, particularly in the case of the SVR model, the growing number of features may significantly increase computational complexity and lead to diminished performance. After identifying the best set of features, an additional experiment was conducted, where the contextual factors and mediator features were removed to assess their importance (This feature set is abbreviated as *UI+udi+nibri+BERTembcd+if*). As shown in Table 6.2 and Figure 6.1, the performance of all four models decreased significantly. These results show that the removal of contextual factors increased prediction RMSE errors by 2.1%, 0.3%, 2.3%, and 2.9%, respectively. This again highlights the utility of integrating contextual features and contextual mediator into the models to achieve more accurate user preference predictions. The details regarding the creation of the contextual mediator can be traced back to the Methodology section 3.5.3.

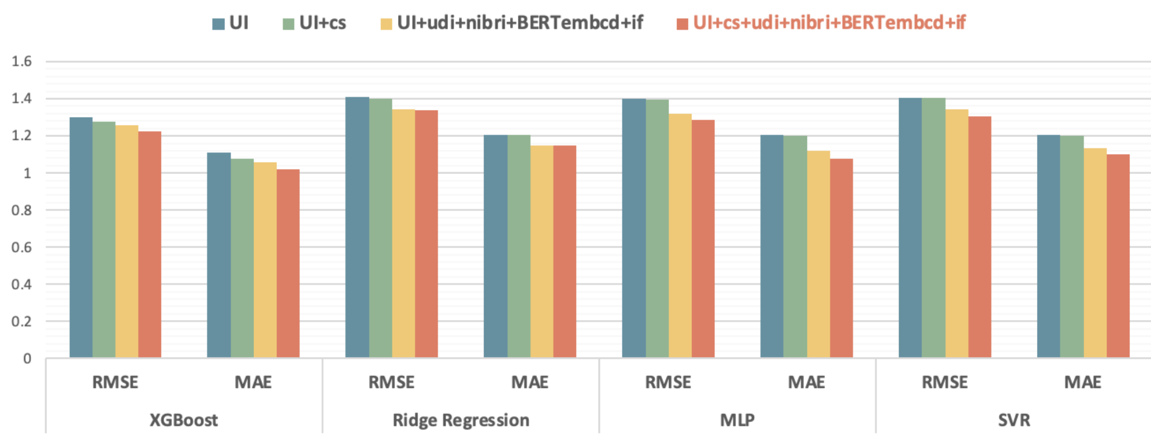


Figure 6.1: *Model comparison results*

To further analyse the best-performing model (*UI+cs+udi+nibri+BERTembcd+if*), Table 6.3 presents the top 10 most important features. This allows for a deeper exploration of the results at the individual feature level, providing more granular insights into which features drive the model's performance. Various approaches have been employed to determine feature importance; further details can be found in Section 6.2. The recipe category features, cooking methods, and dynamic contextual scenario features appear to be particularly influential in the XGBoost model. Nutritional information and image features, on the other hand, are clearly significant contributors in the ridge regression model. In the MLP model, dynamic contextual features combined with user demographic features, such as hot summer's day(HSD), cold winter's day (CWD), gender, and home country feature, frequently appear as the top important features. This implies that, for predicting recipe ratings, combining contextual features with user demographic information may be sufficient for the hidden layers of the MLP to identify patterns and deliver reasonable performance. User demographic features also contributed to the SVR model's performance, with the top five features primarily related to home country and current country attributes. Additionally, features indicating busy and stressful weekdays appear in the top 10 important features, suggesting that this combination of features may be effective in helping the SVR model find the optimal hyperplane.

These findings indicate that, although different models benefit from distinct feature sets, contextual and demographic features play a crucial role across all models. This further underscores the importance of incorporating both dynamic and static contextual factors, such as ethnic origin and home country, are typically included in the user demographic feature set. It is also notable that even integrate with more types of features into models, the home country feature remains more influential than the current living country feature. This again suggests that individuals retain a sense of their home country in their preferences, regardless of where they currently reside. This phenomenon underscores the deep-rooted cultural and emotional connections people have with their native cuisines, which can even detected by algorithm. In this study, the one-stage model using alternative data sources, such as recipe content features, user demographics, and contextual information, showed promise in mitigating the cold-start problem without relying on previous user ratings.

Table 6.3: *Top 10 influential features in each model under the best combination feature set based on feature importance*

XGBoost	Ridge regression	MLP	SVR
Main dish	Carbohydrate%	HSD	HomeCountry_Philippines
gender_other	Protein%	CWD	HomeCountry_Zimbabwe
pca_BERTEMBCD_10	salt/100g	Non-binary or gender diverse	HomeCountry_Hungary
B	Saturates%	Mixed or multiple ethnic groups	CurrentLivingCountry_Philippine
gender_pnts	Colourfulness	HomeCountry_Poland	HomeCountry_Taiwan
pca_BERTEMBCD_1	Fat%	user_id	Energy(kcal)
HSD	total_weight/perportion	HomeCountry_Bulgaria	gender_pnts
Cholesterol%	Calcium(g)	APA	pca_BERTEMBCD_19
Salad	pca_BERTEMBCD_8	cook_book_using_frequency	Total_NE(g)
pca_BERTEMBCD_8	pca_BERTEMBCD_18	HomeCountry_Mexico	B

notes: The above abbreviation stands for information below. HSD: hot summer day (contextual group), CWD: cold winter day (contextual group), H: happy (contextual group), S: sad (contextual group), B: busy (contextual group), R: relax (contextual group), APA: after physical activities (contextual group), G: generic group)

6.3.2 Contextual Modelling Two-stage Approach

To further enhance model performance and address **RQ3** (Which model architecture and feature combination achieve optimal rating prediction accuracy?), a two-stage model structure was proposed. This approach integrates traditional recommender system algorithms with machine learning and deep learning techniques, enabling the use of user historical ratings to predict new preferred recipes. SVD and NMF were applied to decompose the user-item matrix, and the resulting decomposition matrices were used as features for the machine learning models to learn from. With a sufficiently large training dataset, the decomposition matrices can be highly representative, enabling machine learning and deep learning models to effectively capture hidden patterns. The complete model-building workflow is illustrated in Figure 6.2. The same training-test split was adopted, ensuring no new users or items in the test set. The SVD and NMF models were applied to the training set to extract the user matrix W and item matrix H. The number of components for both approaches was tested from 3 to 10. This two-stage model was then applied to the previously best-performing feature set, *UI+cs+udi+nibri+BERTembcd+if* for further experimentation. The performance of this two-stage model is shown in Table 6.4 and Table 6.5.

For the XGBoost model, the best performance was achieved with SVD when the number of components was set to 6 (RMSE=1.239, MAE=1.023, R^2=0.225). The performance of the ridge regression model was not significantly impacted by the number of components, with the best results obtained at SVD n_components = 5 (RMSE=1.274, MAE=1.07, R^2=0.181). The MLP model's performance varied with changes in the number of components, achieving the best results at SVD n_components = 8 (RMSE=1.23, MAE=1.004, R^2=0.236). The SVR model's performance improved as more SVD features were integrated, peaking at n_components = 10 (RMSE=1.234, MAE=1.025, R^2=0.231).

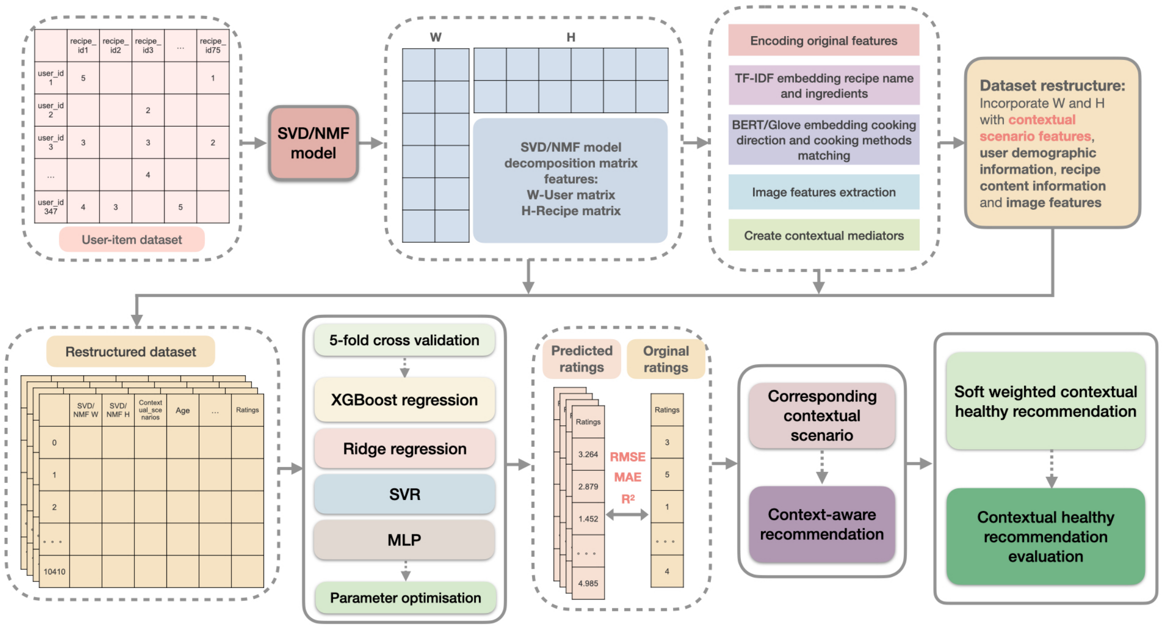


Figure 6.2: *Two stage model building workflow*

Table 6.4: *Two stage model evaluation results - SVD-derived features integrated with UI+cs+udi+nibri+BERTembcd+if features*

	XGBoost Regression Model			Ridge Regression Model			MLP			SVR		
SVD n_components	**RMSE**	**MAE**	**R2**	**RMSE**	**MAE**	**R2**	**RMSE**	**MAE**	**R2**	**RMSE**	**MAE**	**R2**
n=3	1.252	1.037	0.208	1.276	1.072	0.177	1.236	1.025	0.229	1.25	1.039	0.212
n=4	1.259	1.041	0.2	1.277	1.073	0.177	1.245	1.035	0.217	1.247	1.035	0.215
n=5	1.256	1.038	0.204	**1.274**	**1.071**	**0.18**	1.25	1.036	0.211	1.242	1.033	0.221
n=6	**1.239**	**1.023**	**0.225**	1.274	1.07	0.181	1.239	1.025	0.225	1.24	1.031	0.224
n=7	1.25	1.032	0.211	1.275	1.071	0.18	1.269	1.045	0.187	1.237	1.027	0.227
n=8	1.251	1.029	0.21	1.275	1.071	0.179	**1.23**	**1.004**	**0.236**	1.236	1.027	0.229
n=9	1.256	1.037	0.203	1.275	1.071	0.179	1.248	1.017	0.214	1.236	1.027	0.228
n=10	1.249	1.025	0.212	1.276	1.072	0.178	1.239	1.026	0.225	**1.234**	**1.025**	**0.231**

The results indicate that integrating SVD/NMF features leads to significant improvements in ridge regression, MLP, and SVD models, as shown in Table 6.6. For the SVR model, the two-stage SVD approach demonstrates the most significant improvements, with a 0.071 reduction in RMSE and a 0.074 reduction in MAE. This corresponds to a 5.44% improvement in RMSE and a 6.73% improvement in MAE compared to the one-stage model. Additionally, the R^2 value increased by 0.091, indicating a dramatic improvement of 65%. The NMF two-stage model also shows improved performance, with reductions of around 0.068 in RMSE and MAE and an 0.087 increase in R^2. Ridge regression also achieved significant improvements with the two-stage approach, with both SVD and NMF decompositions providing similar performance, showing a decrease of 0.065 in RMSE and 0.076 in MAE, along with an increase of 0.087 in R^2. These results correspond to a 4.85% improvement in RMSE, a 6.63% improvement in MAE, and a substantial 92.55% improvement in R^2 compared to the one-stage approach. The MLP model's performance improved with the two-stage SVD model, showing a reduction of

Figure 6.3: *Two stage model performance comparison*

0.058 in RMSE (a 4.5% relative improvement), 0.071 in MAE (a 6.60% relative improvement), and a significant increase of 0.074 inR^2(45.68% relative improvement). Similarly, with the two-stage NMF model, the MLP achieved improvements of 0.051 in RMSE (a 3.96% relative improvement), 0.046 (a 4.28% relative improvement) in MAE, and 0.065 in R^2 (40.12% relative improvement).

Interestingly, the two-stage approach does not yield a significant improvement for the XGBoost model. Both approaches exhibit comparable performance, with the two-stage model even slightly underperforming relative to the one-stage model. A potential explanation for the one-stage XGBoost model outperforming the two-stage XGBoost model could be the information loss incurred during the SVD/NMF decomposition. In this experiment, the XGBoost model has been fine-tuned and applies both L1 and L2 regularisation to improve generalisation. As a result, having a fully representative user-item matrix is beneficial for the model's performance. On the other hand, this also implied that individual recipe preferences are

Table 6.5: *Two stage model evaluation results - NMF-derived features integrated with UI+cs+udi+nibri+BERTembcd+if features*

	XGBoost Regression Model			Ridge Regression Model			MLP			SVR		
NMF n_components	**RMSE**	**MAE**	**R2**	**RMSE**	**MAE**	**R2**	**RMSE**	**MAE**	**R2**	**RMSE**	**MAE**	**R2**
n=3	1.259	1.048	0.2	1.278	1.073	0.176	1.271	1.055	0.184	1.252	1.047	0.208
n=4	1.266	1.048	0.19	1.277	1.073	0.176	**1.237**	**1.029**	**0.227**	1.247	1.043	0.215
n=5	**1.259**	**1.048**	**0.2**	1.276	1.072	0.178	1.259	1.045	0.2	1.244	1.039	0.219
n=6	1.275	1.06	0.179	**1.274**	**1.07**	**0.18**	1.273	1.054	0.182	1.241	1.037	0.222
n=7	1.274	1.051	0.181	1.276	1.072	0.177	1.276	1.049	0.178	1.239	1.033	0.225
n=8	1.28	1.057	0.173	1.278	1.072	0.176	1.237	1.027	0.227	1.239	1.034	0.225
n=9	1.294	1.064	0.154	1.278	1.073	0.176	1.24	1.027	0.223	**1.237**	**1.031**	**0.227**
n=10	1.282	1.066	0.17	1.277	1.072	0.176	1.237	1.021	0.227	1.239	1.032	0.225

complex, with no clear (liner) relationships emerging even when historical ratings are available. Simply knowing a person's preferred recipes does not directly translate to accurately predicting their preferences for new, unknown recipes. The decomposed matrices, in this case, may not provide additional value and could even hinder performance. It is expected that with a larger training set, the decomposed matrices may better represent individual preferences, potentially leading to improved XGBoost model performance. However, without further experimentation, this outcome remains uncertain.

Overall, SVD decomposition features perform better than NMF, as clearly evidenced in Figure 6.3 and Table 6.6. For the XGBoost and MLP models, the SVD features significantly outperform the NMF model. The SVD-XGBoost model shows a 0.022 improvement in RMSE and MAE, and a 0.028 improvement in R^2 compared to the NMF decomposition features. The MLP model demonstrates a 0.009 improvement in RMSE, a 0.011 improvement in MAE, and a 0.012 improvement in R^2 by incorporating SVD decomposition features compared to NMF features. Both decomposition methods perform similarly for ridge regression and SVR, but the SVD slightly outperforms NMF in every evaluation metric. This may be due to the non-negative processing of NMF, which might not be well-suited for re-entering into the machine learning model during training, as it could limit the representation of complex relationships in the data. In contrast, SVD features may better capture the underlying patterns of user preferences, more effectively reflecting how users make choices or express preferences.

Table 6.6: *Two stage model evaluation average results comparison*

	XGBoost Regression Model			Ridge Regression Model			MLP			SVR		
	RMSE	**MAE**	**R2**	**RMSE**	**MAE**	**R2**	**RMSE**	**MAE**	**R2**	**RMSE**	**MAE**	**R2**
One stage: UI+cs+udi+nibri+BERTembcd+if	1.223	1.02	0.244	1.339	1.146	0.094	1.288	1.075	0.162	1.305	1.099	0.14
Best SVD two stage performance	1.239	1.023	0.225	1.274	1.07	0.181	1.23	1.004	0.236	1.234	1.025	0.231
Best NMF two stage performance	1.259	1.048	0.2	1.274	1.07	0.18	1.237	1.029	0.227	1.237	1.031	0.227
SVD two stage average	1.252	1.033	0.209	1.275	1.071	0.179	1.245	1.027	0.218	1.240	1.031	0.223
NMF two stage average	1.274	1.055	0.181	1.277	1.072	0.177	1.254	1.038	0.206	1.242	1.037	0.221
Pairwise model differences												
Best SVD two stage - One stage	0.016	0.003	-0.019	-0.065	-0.076	0.087	-0.058	-0.071	0.074	-0.071	-0.074	0.091
(%)	(1.31%)	(0.29%)	(-7.79%)	(-4.85%)	(-6.63%)	(92.55%)	(-4.50%)	(-6.60%)	(45.68%)	(-5.44%)	(-6.73%)	(65.00%)
Best NMF two stage - One stage	0.036	0.028	-0.044	-0.065	-0.076	0.086	-0.051	-0.046	0.065	-0.068	-0.068	0.087
(%)	(2.94%)	(2.75%)	(-18.03%)	(-4.85%)	(-6.63%)	(91.49%)	(-3.96%)	-4.28%	(40.12%)	(-5.21%)	(-6.19%)	(62.14%)
NMF average - SVD average	0.022	0.022	-0.028	0.002	0.001	-0.002	0.009	0.011	-0.012	0.002	0.006	-0.002
(%)	(1.76%)	(2.13%)	(-13.40%)	(0.16%)	(0.09%)	(-1.12%)	(0.72%)	(1.07%)	(-5.50%)	(0.16%)	(0.58%)	(-0.90%)

Recipe SVD feature 1 and *user SVD feature 1* have been identified as the most influential features for the XGBoost regression, MLP, and SVR models, see Table 6.7. Contextual features, such as *"cold winter's day" and "hot summer's day"*, are significant for both the XGBoost and MLP models. For Ridge regression, the image feature *"colourfulness"* is the most influential, followed by several *nutritional features*. Static contextual features, particularly those related to the *home country*, are more influential in the SVR model. Overall, SVD decomposition features and contextual features (both static and dynamic) consistently emerge as the most important feature types across all four models. Following the core research design, after identifying the optimal rating prediction model, addressing the trade-offs between user preferences and nutritional needs becomes essential. The next section outlines a strategy to gradually balance taste and health, offering healthier alternatives that users may still enjoy.

6.3.3 Contextual Healthy Recommendation and Evaluation

Traditional healthy recommendations often involve creating a subset of healthy recipes and matching user-preferred recipes with those in the healthy subset. This approach can lead to several issues. First, the similarity-matching algorithm may not work accurately, which can further degrade the quality of the recommendations. Second, if the created healthy subset is extremely healthy, it may be difficult for users with relatively unhealthy eating habits to accept these recommendations. It can be argued that the key to effective healthy recommendations is to subtly integrate healthy aspects while aligning them with the user's past preferences. Instead of making strict and rigid healthy recommendations, the approach should offer a range of options. By encouraging users to accept slightly healthier choices, the system can gradually introduce even healthier recipes over time, fostering a more sustainable shift in eating habits.

More importantly, users may not always choose unhealthy or high-calorie food all the time. Contextual factors significantly impact eating behaviour, making it difficult to eliminate these influences. Under different contextual situations, users' eating and nutritional intake behaviours can vary. It is therefore important to link healthy recommendations with the user's past preferences under different contexts.

To address **RQ4** (What strategies can be employed in building healthy recommendation models to effectively balance preferences and health considerations?), this research builds on the work of Zheng et al. (2013) and Harvey et al. (2013). The aim is to explore the trade-off between offering users the best prediction of what they want and recommending healthier options than what they typically choose in specific contexts.

Table 6.7: *Two stage model best performance (SVD decomposition) top-10 feature importance*

XBGoost regression, SVD n=6	Ridge regression, SVD n=5	MLP, SVD n=8	SVR, SVD n=10
recipe_SVDn6_feature1	Colorfulness	user_SVDn8_feature1	user_SVDn10_feature1
user_SVDn6_feature1	Iron(g)	Bulgaria	recipe_SVDn10_feature1
Total_NE(g)	Vitamin C%	Non-binary or gender diverse	Taiwan
CWD	pca_BERTEMBCD_7	CWD	pca_BERTEMBCD_14
HSD	Protein%	Black, Black British, Caribbean or African	gender_other
recipe_SVDn6_feature3	Calcium(g)	Ghana	Philippines
pca_BERTEMBCD_12	Potassium(g)	user_SVDn8_feature5	Hungary
Salad	Fat%	Salad	Zimbabwe
general	contextual_FDA	Vegan	Japan
pca_BERTEMBCD_3	contextual_scenario_label	HSD	pca_BERTEMBCD_19
Cholesterol(g): 0.011	Sugars%	Poland	gender_pnts

This could operationalise as one metric consisting of a weighted linear combination of three constituent scores shown in the equation below. Here, i represents a given recipe, $\hat{pr}_i$ is the predicted recipe rating of the best preforming model based on the user's past preferences, identified in the previous section (The one-stage XGBoost model was taken as an example and utilised in this context), n_i is the normalised nutritional value of the recommended recipe, and c_i is the normalised label-encoding values of each corresponding contextual scenario the user is encountering. λ_c is a free parameter corresponding to the impact of the user's current context on their eating behaviour, while λ_n corresponds to the user's expectation of healthy eating. Both λ_c and λ_n can be adjusted to suit user's priorities. The score can be calculated as shown below, where chr stands for contextual healthy ratings. Note that the user would have the freedom to decide the nutritional expectation λ_n and λ_c, which indicates the degree to which they prioritize healthier and more nutritious eating, as well as the extent to which they believe context influences their food choices and nutritional expectations.

$$\begin{aligned}\mathrm{chr}_i = {} & \lambda_c \cdot c_i \\ & + \lambda_n \cdot n_i \\ & + (1 - \lambda_c - \lambda_n) \cdot \hat{pr}_i \end{aligned} \tag{6.1}$$

In this experiment, after identifying the best rating prediction approach based on various feature sets and combinations of user preferences, the proposed contextual health rating approach was applied to generate a new healthy recommendation score. Three international standards FSA, WHO, and FDA have been utilised to generate health ratings, in alignment with the previously proposed WHO_adj and FDA_adj measures, which reasonably assess the healthiness of individual recipes. It is important to note that, since a higher FSA score indicates an unhealthier recipe, when using the FSA standard, the score needs to be modified to 1−scaled_FSA_score to ensure healthier recipes carry more weight. The potential impact of the contextual scenario on user's eating and nutritional intake behaviour (λ_c) and expectation for healthy eating (λ_n) are both set to 1/3, reflecting an equal balance between these three components. This proposed approach could also demonstrate inherent flexibility, allowing it to seamlessly transition between pure contextual weighted recommendation (when λ_n=0), and pure healthy weighted recommendation (when λ_c=0). Pure healthy recommendations, where the focus is solely on balance predicted rating (user preference) with recipe health level to provide re-ranking results, where the list operates from the most preferred and healthiest options at the top to the least preferred and unhealthiest at the bottom.

Contextual healthy recommendations may have the potential to offer subtler and more diverse suggestions, as supported by the findings from the first and second stage of the study. An individual's nutritional intake may be influenced by various contextual factors. More importantly, nutritional intake expectations can vary significantly across different contexts. For example, when feeling sad, individuals may be more inclined to consume comfort foods, such as candy or high-calorie dishes, even though they are aware of the need to eat healthier. In such situations, recommending extremely healthy options may not necessarily increase the acceptance of the recommendation. This proposed approach highlights the interaction between contextual situations and the willingness to eat healthily, while also providing users with greater flexibility to adjust the influence of contextual factors on their eating preferences and their expectations for healthy eating.

Generating new ratings (and therefore ranked lists) by combining the weights of the contextual scenario and nutritional value would obviously lead to a drop in traditional recommender systems evaluation measures (e.g., RMSE). Following the approach of Elsweiler & Harvey (2015), who proposed a novel method for offline evaluation of healthy recommendations, the weight of nutritional value should be incorporated into the calculation of the root error to balance the trade-off between user preferences and the need for healthy intake. The adjusted healthy recommendation RMSE score is presented below:

$$\mathrm{RMSE_h} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\lambda \cdot (r_i - \hat{chr}_i) + (1-\lambda)\cdot(n_i))^2} \tag{6.2}$$

where:

r_i is the actual rating,

n_i is the normalised nutritional value,

$\hat{chr}_i$ is the adjusted contextual health rating as previously proposed,

λ is the indicator to balance user preference and nutritional expection, pre-set up at 0.5.

The final healthy recommendation evaluation results comparison can be seen in Table 6.8. Four one-stage XGBoost models were employed to provide examples of contextual healthy recommendations, as XGBoost generally outperforms the other three models. These models include the baseline user and item feature (UI) model, the user*item model with contextual features (UI+cs), the best feature combination model (UI+cs+udi+nibri+BERTembcd+if), and the best feature combination model without contextual factors and mediators (UI+udi+nibri+BERTembcd+if). Three international health standards were examined, along with two adjusted standards. The best feature combination group achieved the highest performance across all health standards (FSA-RMSEh=1.2, WHO-RMSEh=1.207, FDA-RMSEh=1.222, WHO_adj-RMSEh=1.222, FDA_adj-RMSEh=1.221). When providing health recommendations, the FSA standard appears to be the most appropriate, as it not only offers a more precise health evaluation for individual recipes but also provides more accurate healthy recommendations based on novel evaluation metrics. Notably, the models with contextual features outperformed those without. The UI+cs model achieved a lower error score (RMSEh=1.204) under the FSA standard, showing an improvement of 0.006 compared to the UI model. The best feature combination set achieved the best performance (RMSEh=1.2) under the FSA standard, outperforming the best feature combination set without contextual features and mediators by 0.007. The proposed WHO_adj and FDA_adj approaches did not perform well in this experiment, suggesting that simply narrowing the scope of these international standards may not be a reasonable approach and could not provide robust recommendation results.

To elaborate on this contextual healthy recommendation results, one participant_id:110 were randomly selected to compare the predicted rating, contextual healthy rating (The FSA

Table 6.8: *Healthy recommendation model evaluation comparison*

	RMSEh				
	FSA	WHO	FDA	WHO_adj	FDA_adj
UI	**1.21**	1.217	1.232	1.232	1.231
UI+cs	**1.204**	1.212	1.227	1.226	1.225
UI+udi+nibri+BERTembcd+if	**1.203**	1.211	1.226	1.225	1.225
UI+cs+udi+nibri+BERTembcd+if	**1.2**	**1.207**	**1.222**	**1.222**	**1.221**

notes: WHO_adj and FDA_adj refer to adjusted values derived from the WHO and FDA standards, respectively. These adjustments divide the standard recommended daily nutritional intake by three, aiming to represent the health level of individual recipes approximately. The details of the experiment can be found in Section 3.4.3.

standard was utilised as it better performed), and original user rating. The results are shown in Table 6.9. It can be observed that the user gave a score of 5 to both recipe ID 50 and 34, with both recipes having an FSA health level of 6, which is generally considered medium health. The contextual health recommendation score also provides the very high rating for both recipes. According to the original prediction results, the user is not expected to like recipe ID 71, having rated it with a score of 1. However, due to the prediction error in the XGBoost model, a much higher rating was incorrectly assigned to recipe ID 71, which the participant is unlikely to accept. With the implementation of the contextual healthy recommendation, this situation is improved. By incorporating the weight of contextual factors and the FSA health level, recipe ID 34 is recommended instead, an option that the user is more likely to enjoy, and that is also relatively healthy.

Table 6.9: *Healthy recommendation results of example user: 110*

user_id	recipe_id	contextual_scenario_label	FSA_health_level	user_ratings	prediction_results	ch_adjusted_ratings
110	19	4	7	1	3.441	1.522
110	12	4	6	1	2.147	1.132
110	50	4	6	5	2.952	1.401
110	48	4	8	3	2.900	1.300
110	18	4	8	3	3.361	1.454
110	52	4	7	1	2.503	1.209
110	34	4	6	5	**3.237**	**1.496**
110	6	4	7	1	**3.344**	**1.490**
110	71	4	9	1	3.356	1.410
110	0	4	8	4	2.435	1.145
110	46	4	5	1	3.196	1.524
110	20	4	10	4	3.490	1.413

note: ch_adjusted_ratings stands for FSA standard contextual_health_adjusted_ratings

After adjusting for contextual healthy recommendations, the recipe with the highest predicted rating became Recipe ID 46. Although this recipe may not typically align with the user's preferences, it is one of the healthiest options on the list based on FSA standard. According to well-established psychological and behavioural theories, such as priming (Harris et al., 2009) and environmental cues (Wansink et al., 2006), exposure to images or environments featuring healthy foods may potentially influence individuals' desires and choices, including food preferences. Regular exposure to such stimuli (whether through media, advertisements, or daily experiences) may prime individuals to make healthier food choices. Given that recommender systems are integrated into daily life, presenting healthier alternatives at the top of recommendations may not immediately capture user interest. However, including a few items that match their preferences within the list could be a more effective strategy for promoting healthier eating habits. This approach is likely more advantageous than a top list consisting entirely of unhealthy but preferred items.

It is also worth noting that, based on current experimental results, delivering more effective contextual healthy recommendations depends on accurate rating predictions, as inaccurate predictions of user preferences can lead to a deviation from recipes the user actually likes, and recommending recipes that users do not enjoy, which further negatively impacts the quality of healthy recommendations. After all, if the healthy recommendations are based on recipes the user dislikes, it can only worsen the overall experience. Further scientific and personalised adjustments to the weights of nutritional expectation and contextual scenario influence might enhance the effectiveness of the recommendations. To better evaluate the performance and application of this proposed idea, conducting user studies to gather actual feedback on this approach would be meaningful.

6.4 Discussion

The main findings of this research related to each RQ can be summarised as follows:

- RQ1: The experimental results underscored the critical role of incorporating dynamic contextual features into the model-building process, as this integration led to enhanced performance across all models. The XGBoost model (UI+cs) demonstrated the most significant enhancement, with a 20.3% increase in R^2, and improvements of 1.8% and 2.8% in RMSE and MAE, respectively, compared to the baseline UI model. As additional feature sets were continuously incorporated, the XGBoost model consistently outperformed the other models. The MLP model demonstrated similar, but slightly better performance than the SVR model, while Ridge regression generally performed worse. This may suggest that linear regression models may not be well-suited for this prediction task, further confirming that predicting food preferences is a multifaceted challenge involving a variety of intricate and interrelated relationships (Sobal et al., 2014; Elsweiler et al., 2017). Overall, recipe content information contributed less significantly compared to user demographic information, as incorporating user-oriented features tended to result in better performance.

- RQ2: After conducting a comprehensive feature ablation study on various combinations of extracted feature sets, the results indicate that the combination of UI features with dynamic contextual features, user demographic features, basic recipe and nutritional information, BERT-embedded cooking directions, and recipe image features (*UI+cs+udi+nibri+ BERTembcd+if*) achieved the best performance across nearly all four models. This combination can be identified as the most effective and suitable for the rating prediction task. The XGBoost model emerged as the most suitable algorithm for this rating prediction task, consistently outperforming other models across almost all feature combinations, the best performance achieving under *UI+cs+udi+nibri+BERTembcd+if* feature combination, with an RMSE of 1.223, MAE of 1.02, and R^2 of 0.244. Although the most important features varied among models, dynamic contextual features and user demographic features consistently emerged as influential across all models. This finding emphasises the importance of both dynamic and static contextual features in predicting users' recipe preferences.

- RQ3: The proposed two-stage model structure demonstrates enhanced generalisation ability, as evidenced by significant performance improvements across three models—Ridge Regression, MLP, and SVR. The improvement of RMSE ranged from approximately 3.96% to 5.44% across all models, while the improvement of MAE was around 6%, and R^2 increased by approximately 40.12% to 92.55%. SVD-derived features generally outperformed their NMF decomposition, although improvements were also observed with NMF. This may be due to the non-negative processing in NMF, which could limit its ability to fully capture the hidden relationships in the data. The two-stage XGBoost model did not outperform the one-stage model, possibly due to information loss during the SVD/NMF decomposition. It is worth highlighting that the proposed one-stage model, which uses only alternative features (e.g., recipe content, image features, user demographics, and contextual information) without relying on previous user ratings, shows promise in mitigating the cold-start problem. Although further experiments specifically targeting cold-start scenarios—such as new users or new items—would be needed to gain deeper insights into its effectiveness.

- RQ4: This study proposed a novel approach to generating and evaluating healthy recommendations. The approach prioritises recipes closely aligned with the user's historical preferences, and subsequently suggests contextual healthier alternatives based on a balance between the nutritional expectation (health level of the recipes) and the potential contextual influence on people's eating and nutritional intake preferences. The idea has been primarily tested on four representative models (UI, UI+cs, UI+udi+nibri+BERTembcd+if and UI+cs+udi+nibri+BERTembcd+if, explanation for each abbreviation can be found in Figure 6.2). The best feature combination group (UI+cs+udi+nibri+BERTembcd+if)), utilising the FSA standard, achieved the highest performance (RMSEh=1.2). When comparing different international health standards, the FSA standard appears to be the most appropriate, as it provides more accurate healthy recommendations based on novel evaluation metrics, suggesting that it may offer a more precise health evaluation for individual recipes. While the results are promising, conducting user studies to gather real-world feedback would provide valuable insights for more accurately assessing the performance and practical application of this proposed approach.

Taken together, these findings substantiate the pivotal role of contextual information in modelling and predicting individuals' dietary preferences, as well as in providing healthy recommendations. To the best of our knowledge, this study is the first to investigate and model the impact of contextual factors on food recommender systems. Based on the model evaluation results, the importance of integrating contextual features, particularly dynamic ones, into FRS has been confirmed. These findings align with previous studies (Harvey et al., 2013; Zhang et al., 2023; Elsweiler et al., 2017), which indicate that not only food-based factors, such as ingredients or the aesthetic appeal of food images, impact individuals' eating behaviour. Emotional states, stress levels, and environmental factors such as temperature or weather can also significantly influence food choices, often without individuals being consciously aware of it (Macht, 2008; Konttinen, 2020; Kusmierczyk et al., 2015b).

To better integrate contextual features and achieve more precise rating prediction performance, various embedding methods were employed, including TF-IDF, BERT sentence embeddings, GloVe, and image feature extraction, in conjunction with machine learning and deep learning algorithms. This positions the study as the first to develop a multi-modal food recommender system. By identifying the optimal feature combination, the research offers valuable insights into the future development of context-aware healthy food recommender systems.

The model-building results demonstrate that the XGBoost model outperforms ridge regression, MLP, and SVR models. Given the dataset's characteristics, where individual features exhibit weak correlations with user ratings, linear regression may not effectively capture the underlying relationships. XGBoost, an advanced tree-based model, excels in handling complex feature interactions and relationships by aggregating predictions from multiple decision trees, thereby enhancing its learning capability. XGBoost has shown impressive performance in both movie recommendation (Qomariyah et al., 2020) and product recommendation domains (Shahbazi & Byun, 2019; Shahbazi et al., 2020). By adopting a machine learning approach for rating prediction, the proposed method partially overcomes the limitations of traditional recommender systems, which often decompose features in a black box manner. This approach provides explainable results based on feature importance (e.g., "People from your hometown tend to enjoy these recipes when they're feeling happy, making them a great match for your mood!"). This could enhance the transparency of model outputs, potentially increasing user trust. Such an approach may offer valuable insights to drive the development of context-aware, explainable food recommender systems (Zhang et al., 2020b).

By analysing model performance and identifying influential feature sets, these findings differ from those of previous research by Teng et al. (2012), Chhipa et al. (2022) and Maheshwari & Chourey (2019). Unlike these studies, recipe ingredients did not significantly improve model performance in this case. This discrepancy could be due to the limited number of recipes (75) included, where the recipe names alone may suffice to capture user preferences. Instead, this study emphasises the importance of integrating cooking methods and recipe image features, as suggested by Elsweiler et al. (2017). These features not only contributed to the best model performance, but also have a foreseeable impact on a recipe's health level. For instance, deep-fried chicken contains substantially more calories than stir-fried chicken. Additionally, the visual appearance of a recipe can be an indicator of its healthiness. For example, images predominantly featuring yellow and orange hues often correlate with less healthy dishes compared to those with vibrant green and red colours, which tend to represent healthier ingredients like vegetables and fruits.

The most pressing challenge in the food recommendation domain is balancing individuals' eating preferences with their nutritional needs. A key contribution of this research addresses this challenge by extending the work of Elsweiler & Harvey (2015) and Harvey et al. (2013). A novel approach was developed to provide subtle contextual healthy recommendations that align with users' eating preferences while also considering the balance between nutritional needs and contextual influences. Experimental results demonstrate that this approach may offer healthy substitutions that users are likely to enjoy. This serves as a promising starting point, as users' acceptance of recommendations could further promote positive behaviour change. The proposed approach also demonstrates potential for fostering interdisciplinary research. Setting the weights for nutritional expectations and contextual scenario influences could be refined through further scientific and behavioural studies to determine the most effective balance. From a broader perspective, this idea could potentially be applied to a wider range of applications, such as balancing and regulating emotions in music recommendations.

Chapter 7

Conclusions

This thesis has explored the influence of contextual factors on people's food choices, nutritional intake, and daily food decision-making processes, identifying key potential influential contextual elements (features) by capturing individuals' insights into their real-life cooking experiences and expectations. The study emphasises the importance of developing context-aware healthy food recommender systems to better align with users' preference and nutritional needs. To generalise these initial findings, this research demonstrated how people's recipe rating behaviour and implied nutritional intake vary under specific contextual situations, utilising a large-scale, between-subjects experimental design to justify the significance of contextual factors from a data-driven perspective. These findings contribute to the advancement of context-aware healthy food recommender systems and have led to the development of one-stage and two-stage contextual modelling approaches, as well as a novel strategy for contextual healthy recommendations.

In this chapter, the main finding of this research will be summarised and presented in Section 7.1. Then, the contribution of the research will be highlighted in Section 7.2. Finally, Section 7.3 will discuss the limitations of the current study and explore potential directions for future work, providing insights into how the research can be further applied, developed and improved.

7.1 Summary of Research Findings

All three stages of the study lead to one primary conclusion: "contextual factors matter". They influence people's daily food choices and decision-making, recipe ratings and implied nutritional intake behaviour, as well as the development of more effective food and healthy recommender systems. To better integrate contextual knowledge into food recommender systems, enhance the performance of rating prediction models, and propose healthy recommendations that align with users' past preferences and nutritional expectations while addressing trade-off challenges, it is crucial to explore and identify potential influential contextual factors. Given the vast array of contextual elements present in our daily lives, it can be extremely difficult to measure them all comprehensively.

The study began with semi-structured interviews to gaining an understanding of people's perspectives and experiences regarding their daily eating habits, and inductively gather their perceptions and views on potential contextual factors that reshaping food choices and nutritional intake behaviours. Following this, a carefully designed large-scale experimental study was conducted, placing participants in simulated contexts and environments to collecting their recipe ratings, and statistically to examined whether these conditions led to variations in rating behaviour and, consequently, different implied nutritional intake. A pre-filtering rating prediction model was then developed to determine whether incorporating contextual factors would improve model performance. After identifying potential influential factors through both qualitative and quantitative methods, the need for developing context-aware recommender systems (RSs) became evident. Novel one-stage and two-stage contextual modelling approaches incorporating contextual features and multimodel embedded feature sets, have been proposed for the development of food recommender systems, which laying the groundwork for more effective healthy recommendations. Additionally, a novel weighted contextual healthy recommendation approach and an healthy evaluation metric were introduced to deliver recommendations that not only align with users' daily preferences but also prioritise healthier recipes at the top of the list. Specifically, this research addressed the following key research questions:

- **RQ1**: What contextual factors affect people's (online) food choices?
- **RQ2**: What impact do these same factors have on people's nutritional intake?
- **RQ3**: Can integrating these contextual factors enhance the performance of recommendation systems?
- **RQ4**: How to combine this knowledge of contextual factors to recommend people healthy recipes that they will enjoy?

This conclusion highlights how each key research question (RQ) and corresponding sub-research question (sub-RQ) was addressed across different stages of the research. In response to RQ1, this study identified key influential factors that affect eating preferences through semi-structured interview. The static factors include cultural background and personal goals (RQ 1.1), while the dynamic factors encompass emotions, busyness, seasons, physical activities, and sustainability (RQ 1.2). These factors were spontaneously mentioned by participants during discussions about their daily lives eating habits and preferences. These are novel influential contextual factors for food recommender systems, as no prior research has integrated these features into FRS to enable adaptable contextual recommendations. Users' food choices and intake behaviours present sizable changes under these factors; emotions and busyness in particular shape individual's food intake behaviour. These identified factors support established findings for the circularity of emotions and eating. A detailed discussion of these findings is presented in Chapter 4. The qualitative interview results were tested and confirmed through a large-scale quantitative experimental study. The results confirmed significant differences in individuals' implied eating behaviours and recipe rating responses across various contextual scenarios. Specifically, people's eating and recipe rating behaviours during busy weekdays differed significantly from those observed during a cold winter day, a sad emotional state, and after physical activities. As anticipated, a substantial variation in recipe preferences is observed between hot summer days and cold winter days, corroborating the concept of seasonality in food preferences over time (RQ 2.1). A detailed discussion is provided in Chapter 5.

To address RQ2, the findings indicate that nutritional information may be undervalued in food choice decisions. Although individuals recognise the importance of nutritional content, they often neglect it when making decisions or reject foods based on their nutritional profile. An exception is observed in individuals engaging in high levels of physical activity, who tend to be more mindful of their food and nutrient intake. However, attitudes toward nutrition often shift dramatically when individuals face health problems, underscoring the importance of preventive measures. Unhealthy eating habits are difficult to break, and preferences established during childhood often persist, reflecting existing research on the development of eating behaviours. Additionally, increased stress levels and emotional fluctuations can lead to higher consumption of calories and sugar, or even contribute to eating disorders (RQ 1.3). A detailed discussion of this topic can be found in Chapter 4. The results, presented in Chapter 5 further highlight significant differences in preferences for healthy recipes across various contextual scenarios. Notably, preferences during hot summer days differed significantly from those during cold winter days, feelings of sadness, and the generic group. Similarly, participants' preferences during busy weekdays showed significant differences compared to those during cold winter days, feelings of sadness, after physical activities, and the generic group (RQ 2.2).

To address the gap in integrating contextual knowledge into food recommender systems, and in response to RQ3, this study confirmed the beneficial role of contextual information in enhancing the accuracy of rating prediction models using both pre-filtering and contextual modeling approaches. When pre-filtering the dataset based on contextual groups, all contextual models outperformed the context-free group, with the 'happy emotion' group achieving the highest performance (MSE=1.615, RMSE=1.271, MAE=1.032, and R^2=0.185) (RQ 2.4). When building models with all eight groups (7 contextual groups + 1 context-free group), contextual factors emerged as the most important contributors based on the XGBoost gain, significantly enhancing the accuracy of rating predictions (RQ 2.3). To delve deeper and incorporate more complex user and recipe information, the proposed one-stage and two-stage contextual modelling approaches also demonstrated that integrating contextual factors at the model-building level significantly outperformed the baseline user-item (UI) model. Among the tested models, the XGBoost model integreated with dynamic contextual features (UI+cs) showed the most significant enhancement, with a 20.3% increase in R^2, and improvements of 1.8% and 2.8% in RMSE and MAE, respectively, compared to the baseline UI model (RQ 3.1). These findings are discussed in greater detail in Chapter 5 and Chapter 6.

The key idea behind providing healthy recommendations in this research is to balance user preferences with nutritional intake, thereby facilitating smoother transitions to healthier eating habits. In this context, developing a more accurate rating prediction model that aligns with users' preferences is crucial. To achieve the best rating prediction model, contextual features and multimodel features were incorporated to enhance model performance. In addressing RQ4, this research identified the optimal feature combination through a comprehensive ablation study, which included UI features, dynamic contextual features, user demographic information, basic recipe and nutritional data, BERT-embedded cooking directions, and recipe image features (*UI+cs+udi+nibri+BERTembcd+if*). This combination achieved the best performance across nearly all four tested models (XGBoost, Ridge regression, MLP, and SVR). The XGBoost model continued to emerge as the best-performing model, with an RMSE of 1.223, MAE of 1.02, and R^2 of 0.245, indicating that it may be the most suitable algorithm for this rating prediction task. Recipe content information show relatively less influential compared to user demographic information. Dynamic contextual features and user demographic features consistently emerged as influential across all models, which emphasises the importance of both dynamic and static contextual features in predicting users' recipe preferences (RQ 3.2). The proposed two-stage model structure demonstrates enhanced generalization ability; however, the one-stage model, which uses alternative data sources without relying on previous user ratings, shows promise in mitigating the cold-start problem (RQ 3.3).

By adopting the proposed weighted contextual healthy recommendation approach and evaluation metrics, which prioritises recipes closely aligned with the user's historical preferences, and subsequently suggests contextual healthier alternatives based on a balance between the nutritional expectation (health level of the recipes) and the potential contextual influence on people's eating and nutritional intake preferences. The results suggest that the best feature combination group (UI+cs+udi+nibri+BERTembcd+if)), utilising the FSA standard, achieved the highest performance. Under this proposed approach, the user is given the flexibility and freedom to set their nutritional expectations, ideally enabling them to re-rank and filter the recommendation results as desired (RQ 1.5). The proposed approach also shows promise for encouraging interdisciplinary research. Fine-tuning the weights for nutritional expectations and contextual influences could benefit from additional scientific and behavioural studies to find the most effective balance (RQ 3.4). A detailed discussion of these findings is presented in Chapter 6.

7.2 Contributions

This work contributes to the expansion of existing concepts regarding which contextual factors can influence people's eating and nutritional intake behaviour. By considering a wide range of variables, this research has identified the most potentially influential static and dynamic contextual factors that are particularly beneficial for the development of the next generation of context-aware food recommender systems. Unlike other research in the fields of psychology and recommender systems, which often focuses on the relationship between individual factors and eating behaviour, or examined contextual factors being selected based on literatures rather than human-centered experiments. This study takes a comprehensive approach, by starting with an understanding of peoples' perceptions and expectations regarding food and online food choices, their attitudes toward nutritional information, and the extent to which they believe contextual factors influence their eating preferences, valuable insights were obtained. Significant practical implications have been derived from gaining insights into users' perspectives, and acquired deeper understanding of how users perceive nutritional information, identifying where they often face challenges, and recognizing the contexts in which their nutritional intake behaviours are likely to change. These insights can be leveraged to develop healthy food recommender systems that more effectively balance user preferences with nutritional intake, and these systems can be tailored to better meet individual needs. Additionally, this research could inspire the design of specialised algorithms for specific types of food recommendations, such as fitness-oriented foods, tailored to individuals with strong personal goals or specific health requirements.

This research also made methodologically contribute to human-centered study on contextual food recommender systems. Accurately collecting, measuring, and analysing contextual factors presents significant challenges, particularly in ways that can reliably influence or predict behaviour. In this research, the semi-structured interview and experimental study approach offer a feasible method for understanding and collecting contextual data. Given the vast number of contextual factors and their potential interactions, it would be unrealistic and prohibitively expensive to study all of them. Therefore, selecting and focusing on the most influential factors becomes critically important. This research confirms the value of contextual factors in influencing eating and nutritional intake behaviour through real-life recall, statistical analysis, and algorithmic modelling. The analysis has identified that the most influential contextual factors likely include emotions, busyness and cultural background. These factors should be considered for integration into the next generation of recommender systems to enable more personalised recommendations. Incorporating such context allows the system to explain how recommendations are tailored to the user's current situation, which can make the suggestions more persuasive and potentially increase user trust. More importantly, this research addresses a significant gap in the availability of datasets for studying context-aware recommender systems. Through a carefully designed large-scale experimental study, it was able to collect user recipe ratings across various contextual scenarios, including a context-free control group. This study introduces a novel method to researching context-aware recommender systems and designs an experiment to collect contextual data. By doing so, it provides valuable insights into user behaviour under different scenarios, offering a unique dataset that enhances the understanding of context in recommendation models. This approach combines insights from psychology, social sciences, and computer science, providing an interdisciplinary contribution that advances the development of context-aware explainable healthy food recommender systems. By bridging these fields, it offers a more holistic understanding of user behaviour and context, paving the way for more effective and personalised recommendation models.

This research makes a notable technological contribution to the advancement of healthy food recommender systems by proposing a re-ranking methods that better align recommendations with individual preferences while also offering healthier options that users are more likely to accept. This approach seeks to balance the trade-off between maintaining user satisfaction and encouraging healthier eating habits, laying a strong foundation for designing more intelligent, health-focused recommender systems, thus facilitating gradual and sustainable behaviour change. Rather than relying solely on traditional recommender system algorithms that draw recommendations based on user-item similarity, this study integrates contextual knowledge with multimodal features extracted from both recipe information and user data. This combination seeks to achieve optimal model performance by considering a richer set of factors that influence user preferences. By identifying the optimal feature combinations through testing with various machine learning and deep learning models, this research may potentially inspire further exploration into integrating advanced algorithms into food recommendation systems. It also highlights the potential for incorporating additional user bias and richer recipe images features to improve recommendation accuracy. After gaining a deeper understanding of user preferences and identifying the best performing model, the adoption of the proposed re-ranking methods would further prioritise healthier recipes while still aligning with user preferences. This study may further offer valuable insights into designing personalised, health-focused recommendation systems by highlighting the importance of specific ingredients and nutritional factors, such as in the context of weight loss recommendation systems, personalised meal suggestions that focus on calorie control, macronutrient balance, and portion sizes while adapting to individual preferences and dietary restrictions to ensure long-term adherence. By striking a balance between nutritional optimisation and user preferences, these systems may not only improve dietary compliance but also enhance user engagement, making it more likely that individuals will consistently follow the recommended guidelines. The key idea behind this approach may provide implications to other domains beyond healthy food recommendation, such as music recommendation, where personalisation plays a crucial role in user experience. By incorporating similar re-ranking methods, a music recommendation system could prioritise songs that align with a user's mood, emotional state, or therapeutic needs, while still considering their listening preferences. For instance, the system could suggest calming music to reduce anxiety, uplifting tracks to enhance motivation, or balanced playlists to help regulate emotions throughout the day. Such an approach could be particularly beneficial in mental health applications, stress management, and wellness-focused digital platforms. In the meantime, these findings might shed light on advancements in the industry, such as the development of smart refrigerators capable of capturing user preferences and recommending recipes that align with the ingredients currently available. These systems may be able to consider factors like users' typical meal preferences at different times of the day or seasonal variations in taste. By aligning recommendations with users' needs and available ingredients, these smart ridges may not only make meal planning more efficient and user-friendly, but also encourage healthier eating habits through tailored suggestions that fit the user's lifestyle and preferences.

However, integrating contextual and user demographic features, as well as nutritional information, may introduce bias into the algorithm and potentially lead to security and privacy-related ethical issues. Several limitations may exist, which will be discussed in the next section.

7.3 Limitations and Future Work

7.3.1 Insights into the Ethical Implications of Healthy Recommender Systems

This research integrates contextual features to enhance model performance and further assigns weights to achieve more refined and health-conscious recommendations. However, it is important to reflect on the challenges of collecting contextual factors in real-life settings, as this may involve integrating wearable devices to collect physiological signals or relying on user self-reported emotional and mental states. Collecting and utilising such personal data could potentially raise privacy concerns. Future research should rigorously adhere to the ethical and responsible AI guidelines outlined by GDPR (General Data Protection Regulation) and HIPAA (Health Insurance Portability and Accountability Act) to ensure that the collection, processing, and use of personal data are conducted lawfully, transparently, and ethically.

Additionally, those involved in the development and implementation of AI and recommender system algorithms must incorporate robust data security measures, such as encryption and anonymization techniques, to further safeguard personal data and uphold ethical AI practices in real-world applications.

While the proposed approach demonstrates effectiveness in context-aware healthy recommendations, it does not explicitly incorporate privacy-aware mechanisms. Given that privacy concerns have been widely recognised as an important aspect of recommender systems research, it is crucial to explore ways to provide privacy-protected recommendations (Knijnenburg & Kobsa, 2013). This could potentially be achieved by offering recommendations based on item identification, encrypting user data, or grouping user identification by location or regional clusters. However, such approaches may reduce the level of personalization to some extent, making it essential to strike an optimal balance between personalised recommendations, ethical considerations, and privacy protection (Ali et al., 2021). Future research could further explore novel algorithms that dynamically adjust privacy levels based on user consent and contextual factors, ensuring robust privacy-preserving techniques while maintaining recommendation quality. Addressing these challenges remains a critical direction for future research to enhance both user trust and the overall effectiveness of privacy-aware recommender systems.

From a narrower perspective, this research has certain limitations, particularly in its ability to directly measure how the proposed healthy recommendations are preferred by users and whether they are more effective than other health re-ranking algorithms in influencing and improving people's dietary patterns, as well as their physical and mental health. Future follow-up studies could further explore users' real needs and preferences regarding healthy recommendations and examine whether the proposed subtle healthy recommendation approach is genuinely preferred in real-life environments (Jinnette et al., 2021; Celis-Morales et al., 2015). This could be assessed through A/B testing in an online setting (Kohavi & Longbotham, 2015). Additionally, it would be meaningful to conduct a longitudinal study to understand whether users' health conditions improve as a result of specific healthy recommendations and to assess whether the proposed approach effectively promotes long-term changes in healthier eating behaviours, as tracking user behaviour changes over time would provide valuable insights into the sustainability of the approach (Schwarzer, 2008).

Moreover, it is crucial to critically evaluate and refine the proposed recommender algorithms while continuously monitoring user behaviour over time. This ensures that the system functions as intended and does not inadvertently lead users to become overly dependent on its recommendations (O'Donovan & Smyth, 2005; Klingbeil et al., 2024). Future research should explore how users interact with the system, assess whether they develop excessive reliance on automated suggestions, and identify strategies to promote informed decision-making. This could be achieved by incorporating user control, explainability, and transparency features, ensuring that the system empowers users rather than passively shaping their choices (Zhang et al., 2020b).

There are also additional avenues for enhancing the healthy recommendations approach, such as expanding beyond the international standards for nutritional assessment. Tailoring recommendations to individuals based on specific nutritional needs could further improve personalisation (Mustaqeem et al., 2020; Sookrah et al., 2019). In addition, the current recommendation model does not consider sequential relationships. Expanding this research to develop a sequential model could be highly beneficial. This model could recommend recipes tailored for specific meals, such as breakfast, lunch, or dinner, and extend to longer periods, like creating weekly meal plans that incorporate users' current contexts. This enhancement might provide greater flexibility and more effective healthy recommendations (Zioutos et al., 2023).

7.3.2 Considerations in Experimental Study and Algorithm Development

As is common practice in the food recommendation literature, this research takes ratings as a proxy for intent to consume. As such, the behaviour measured in this research is implied behaviour - none of the recipes were actually consumed (Rozin, 2007). More in-depth research is needed to investigate whether people's actual behaviour changes under different contextual situations. Despite efforts to control for the impact of individual contextual factors in our experimental design, the results may still be affected by uncontrolled real-world variables. The user studies conducted were necessarily somewhat contrived and simulating emotions is clearly not the same as experiencing them naturally. Manipulating simulated contexts may lead to the representation of an artificial nature, introducing a potential conflict between simulated scenarios and the real world. Therefore, a study of real-life contexts is needed to thoroughly confirm the findings presented in this study (Reis, 2018).

Additionally, the examination of user ratings was based on a limited sample of recipes (n = 75) and only included four main categories (main dishes, soup, desserts/snacks, and salads). A larger and more diverse set of recipes could lead to more generalisable study results. Notably, the improvement of each model through different feature combination is not exceptional. This may be due to the relatively small dataset used in this research, particularly for the deep learning (MLP) model. A larger experimental study, encompassing a greater number of participants and recipes, would be needed to achieve better model performance.

Although, a novel perspective on integrating contextual features into food recommender system models and balancing healthy recommendations with users' past preferences have been provided. Due to the difficulty of collecting and controlling contextual factors, participants were exposed to only a single contextual scenario during the data collection phase. This limitation restricts the diversity of data available for analysis. Given that, the contextual scenarios investigated are not necessarily mutually exclusive. It would be both important and interesting to study the impact of the interaction of dynamic contextual factors on individuals' eating behaviour. Furthermore, a within-subject experimental design that measures how individual participants behave under various contextual situations would further enrich the findings of this research and allow for more personalised contextual healthy recommendations to be made (Charness et al., 2012). An experimental design involving the collection of quantitative biological measurements (such as, electroencephalogram and electrocardiogram signal) could potentially contribute to developing more intelligent systems. By predicting users' emotions in advance, the system may be able to suggest healthier recipes that align with their preferences and, potentially, regulate their emotions—an interesting direction for future research. With a broader vision in mind, developing an online app or a food website that enables users to self-report their current context, collect real-time user ratings, and test the effectiveness of the recommendation algorithm in an online environment would be a significant step forward (Maruyama et al., 2012).

It is worth noting that during the third stage of this research, the recipe image features, which were extracted and utilised in the model training process, were not shown to participants. Since the objective of this stage was to achieve advanced model performance, it aligns with common practices in the fields of machine learning and recommender systems to enrich the dataset (Dang et al., 2021; Di Noia et al., 2012). When participants viewed the recipe names and ingredients, they may have mentally visualised the corresponding recipe images. This could potentially explain why integrating recipe image features helped improve the model's ability to predict user preferences more accurately (Missbach et al., 2015; Muñoz-Vilches et al., 2020). However, the selection of the recipe image database may introduce potential biases. For example, the level of visual appeal for each recipe image cannot be guaranteed, as the recipes and images were sourced from different chefs. Although researchers attempted to mitigate such biases by selecting images of similar levels of visual appeal, some biases may still persist. Additionally, the selected images may differ from how participants mentally visualised the recipes. For future work, a comparative experimental design could help address these concerns. For instance, one group of participants could be provided with recipe images, while another group is not. By comparing the preferences for each recipe between the two groups, which could determine whether significant differences exist and evaluate how recipe images may influence recipe preference and introduce bias.

Five embedding methods and four prediction algorithms were examined in this study. However, there are other state-of-the-art algorithms and combinations that warrant future experiments. For example, besides BERT sentence embedding, fine-tuning BERT specifically for food and recipe tasks or incorporating BERT's attention mechanism could provide valuable insights (Zhang et al., 2024). It is also worth implementing feature selection approaches to identify the most influential features at a granular level. Moreover, creating a large recipe image database and training deep learning networks (such as, VGG16, ResNet) for multi-label ingredient classification tasks would be valuable (Zhang et al., 2020a; Liang, 2020). This approach would allow for the extraction of additional ingredient features based on images and reveal hidden features within the recipe images themselves, which could be incorporated into rating prediction models, may potentially leading to significant improvements in model performance. Furthermore, additional user bias features, such as hobbies and perceived impact of contextual factors on behaviour, could be incorporated into the model to assess their potential contribution to performance improvement (Harvey et al., 2013). Conducting a feature selection step might further enhance model performance and provide insights into the best performing model at individual feature level.

Bibliography

Abdollahi, B. & Nasraoui, O. (2018). Transparency in fair machine learning: the case of explainable recommender systems. *Human and machine learning: Visible, explainable, trustworthy and transparent*, 21–35.

Aberg, J. (2006). Dealing with malnutrition: A meal planning system for elderly. *AAAI spring symposium: argumentation for consumers of healthcare*, 1–7.

Abusair, M., Sharaf, M., Bozeya, M., & Beiruti, A. (2021). Context-aware recommender system based on content filtering. *2021 IEEE 15th International Conference on Application of Information and Communication Technologies (AICT)*, 1–6.

Adomavicius, G. & Kwon, Y. (2011). Maximizing aggregate recommendation diversity: A graph-theoretic approach. *Proc. of the 1st International Workshop on Novelty and Diversity in Recommender Systems (DiveRS 2011)*, 3–10.

Adomavicius, G. & Tuzhilin, A. (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE transactions on knowledge and data engineering*, 17(6), 734–749.

Adomavicius, G. & Tuzhilin, A. (2011). Context-aware recommender systems. *Recommender systems handbook*, 217–253. Springer.

Adriaanse, M. A., de Ridder, D. T., & Evers, C. (2011). Emotional eating: eating when emotional or emotional about eating? *Psychology and Health*, 26(1), 23–39.

Afchar, D., Melchiorre, A., Schedl, M., Hennequin, R., Epure, E., & Moussallam, M. (2022). Explainability in music recommender systems. *AI Magazine*, 43(2), 190–208.

Agarwal, D. & Chen, B.-C. (2010). flda: matrix factorization through latent dirichlet allocation. *Proceedings of the third ACM international conference on Web search and data mining*, 91–100.

Agarwal, D., Chen, B.-C., & Long, B. (2011). Localized factor models for multi-context recommendation. *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, 609–617.

Aggarwal, C. C. et al. (2016). *Recommender systems*, volume 1. Springer.

Aghdam, M. H. (2019). Context-aware recommender systems using hierarchical hidden markov model. *Physica A: Statistical Mechanics and Its Applications*, 518, 89–98.

Aguiar-Bloemer, A. C. & Diez-Garcia, R. W. (2018). Influence of emotions evoked by life events on food choice. *Eating and Weight Disorders-Studies on Anorexia, Bulimia and Obesity*, 23, 45–53.

Ahn, H. J. (2008). A new similarity measure for collaborative filtering to alleviate the new user cold-starting problem. *Information sciences*, 178(1), 37–51.

AL-Bakri, N. F. & Hashim, S. H. (2019). Collaborative filtering recommendation model based on k-means clustering. *Al-Nahrain Journal of Science*, 22(1), 74–79.

Al-Ghobari, M., Muneer, A., & Fati, S. M. (2021). Location-aware personalized traveler recommender system (lapta) using collaborative filtering knn. *Computers, Materials & Continua*, 69(2).

Alharthi, H., Inkpen, D., & Szpakowicz, S. (2018). A survey of book recommender systems. *Journal of Intelligent Information Systems*, 51, 139–160.

Ali, W., Kumar, R., Deng, Z., Wang, Y., & Shao, J. (2021). A federated learning approach for privacy protection in context-aware recommender systems. *The Computer Journal*, 64(7), 1016–1027.

Alpaydin, E. (2021). *Machine learning*. MIT press.

Amatriain, X. & Basilico, J. (2015). Recommender systems in industry: A netflix case study. *Recommender systems handbook*, 385–419. Springer.

Arulprakash, M., Joshikha, M., & Khemka, C. (2024). Food recommender system using content based filtering. *AIP Conference Proceedings*, volume 3075.

Aschner, P., Karuranga, S., James, S., Simmons, D., Basit, A., Shaw, J. E., Wild, S. H., Ogurtsova, K., Saeedi, P., Group, I. D. F. D. E. G. W., et al. (2021). *The international diabetes federation's guide for diabetes epidemiological studies*.

Avery, R. J., Mathios, A., Shanahan, J., & Bisogni, C. (1997). Food and nutrition messages communicated through prime-time television. *Journal of Public Policy & Marketing*, 16(2), 217–227.

Balabanović, M. & Shoham, Y. (1997). Fab: content-based, collaborative recommendation. *Communications of the ACM*, 40(3), 66–72.

Balog, K., Radlinski, F., & Arakelyan, S. (2019). Transparent, scrutable and explainable user models for personalized recommendation. *Proceedings of the 42nd international acm sigir conference on research and development in information retrieval*, 265–274.

Baltrunas, L. & Amatriain, X. (2009). Towards time-dependant recommendation based on implicit feedback. *Workshop on context-aware recommender systems (CARS'09)*, 25–30.

Baltrunas, L. & Ricci, F. (2009). Context-dependent items generation in collaborative filtering. *Proceedings of the 2009 Workshop on Context-Aware Recommender Systems*, 22–25.

Bartkiene, E., Steibliene, V., Adomaitiene, V., Juodeikiene, G., Cernauskas, D., Lele, V., Klupsaite, D., Zadeike, D., Jarutiene, L., Guiné, R. P., et al. (2019). Factors affecting consumer food preferences: food taste and depression-based evoked emotional expressions with the use of face reading technology. *BioMed Research International*, 2019.

Bauman, A. (2015). Qualitative online interviews: Strategies, design, and skills. *Qualitative research in organizations and management: An international journal*, 10(2), 201–202.

Bavaresco, R., Barbosa, J., Vianna, H., Büttenbender, P., & Dias, L. (2020). Design and evaluation of a context-aware model based on psychophysiology. *Computer Methods and Programs in Biomedicine*, 189, 105299.

Beecroft, A. (2010). *Authorship and cultural identity in early Greece and China: Patterns of literary circulation*. Cambridge University Press.

Bellogín, A., Cantador, I., Díez, F., Castells, P., & Chavarriaga, E. (2013). An empirical comparison of social, collaborative filtering, and hybrid recommenders. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 4(1), 1–29.

Bennett, J., Lanning, S., et al. (2007). The netflix prize. *Proceedings of KDD cup and workshop*, volume 2007, 35.

Berkovsky, S. & Freyne, J. (2010). Group-based recipe recommendations: analysis of data aggregation strategies. *Proceedings of the fourth ACM conference on Recommender systems*, 111–118.

Biernacki, P. & Waldorf, D. (1981). Snowball sampling: Problems and techniques of chain referral sampling. *Sociological methods & research*, 10(2), 141–163.

Billsus, D. & Pazzani, M. J. (2000). User modeling for adaptive news access. *User modeling and user-adapted interaction*, 10, 147–180.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan), 993–1022.

Boas, T. C., Christenson, D. P., & Glick, D. M. (2020). Recruiting large online samples in the united states and india: Facebook, mechanical turk, and qualtrics. *Political Science Research and Methods*, 8(2), 232–250.

Bolek, S. (2020). Consumer knowledge, attitudes, and judgments about food safety: A consumer analysis. *Trends in Food Science & Technology*, 102, 242–248.

Bourdieu, P. (2018). Distinction a social critique of the judgement of taste. *Inequality*, 287–318. Routledge.

Bove, C. F., Sobal, J., & Rauschenbach, B. S. (2003). Food choices among newly married couples: convergence, conflict, individualism, and projects. *Appetite*, 40(1), 25–41.

Braun, V. & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative research in psychology*, 3(2), 77–101.

Braun, V. & Clarke, V. (2012). *Thematic analysis.* American Psychological Association.

Breese, J. S., Heckerman, D., & Kadie, C. (2013). Empirical analysis of predictive algorithms for collaborative filtering. *arXiv preprint arXiv:1301.7363.*

Bronfenbrenner, U. (1979). *The ecology of human development: Experiments by nature and design.* Harvard university press.

Brunson, K. (2008). Electronic textbooks: An empirical study of adoption potential. *The Liberty Business Review*, 1(1), 44–55.

Bryman, A. (2016). *Social research methods.* Oxford university press.

Burke, R. (2002). Hybrid recommender systems: Survey and experiments. *User modeling and user-adapted interaction*, 12, 331–370.

Burke, R. (2007). Hybrid web recommender systems. *The adaptive web: methods and strategies of web personalization*, 377–408.

Button, K. S., Ioannidis, J. P., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S., & Munafò, M. R. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nature reviews neuroscience*, 14(5), 365–376.

Cai, R. (2024). Dietary and social stratification: An anthropological perspective. *COMPASS*, volume 4, 56–65.

Cameron, D. S., Bertenshaw, E. J., & Sheeran, P. (2015). The impact of positive affect on health cognitions and behaviours: A meta-analysis of the experimental evidence. *Health Psychology Review*, 9(3), 345–365.

Cao, J., Hu, H., Luo, T., Wang, J., Huang, M., Wang, K., Wu, Z., & Zhang, X. (2015). Distributed design and implementation of svd++ algorithm for e-commerce personalized recommender system. *Embedded System Technology: 13th National Conference, ESTC 2015, Beijing, China, October 10–11, 2015, Revised Selected Papers 13*, 30–44.

Capita, R. & Alonso-Calleja, C. (2005). Differences in reported winter and summer dietary intakes in young adults in spain. *International journal of food sciences and nutrition*, 56(6), 431–443.

Carrer-Neto, W., Hernández-Alcaraz, M. L., Valencia-García, R., & García-Sánchez, F. (2012). Social knowledge-based recommender system. application to the movies domain. *Expert Systems with applications*, 39(12), 10990–11000.

Castillo-Montoya, M. (2016). Preparing for interview research: The interview protocol refinement framework. *Qualitative report*, 21(5).

Cavazza, N., Guidetti, M., & Butera, F. (2015). Ingredients of gender-based stereotypes about food. indirect influence of food type, portion size and presentation on gendered intentions to eat. *Appetite*, 91, 266–272.

Celis-Morales, C., Lara, J., & Mathers, J. C. (2015). Personalising nutritional guidance for more effective behaviour change. *Proceedings of the Nutrition Society*, 74(2), 130–138.

Chandak, M., Girase, S., & Mukhopadhyay, D. (2015). Introducing hybrid technique for optimization of book recommender system. *Procedia Computer Science*, 45, 23–31.

Chandon, P. & Wansink, B. (2012). Does food marketing need to make us fat? a review and solutions. *Nutrition reviews*, 70(10), 571–593.

Charness, G., Gneezy, U., & Kuhn, M. A. (2012). Experimental methods: Between-subject and within-subject design. *Journal of economic behavior & organization*, 81(1), 1–8.

Chavan, P., Thoms, B., & Isaacs, J. (2021). A recommender system for healthy food choices: building a hybrid model for recipe recommendations using big data sets.

Chen, S., Chen, M., Wu, X., Lin, S., Tao, C., Cao, H., Shao, Z., & Xiao, G. (2022). Global, regional and national burden of low back pain 1990–2019: a systematic analysis of the global burden of disease study 2019. *Journal of orthopaedic translation*, 32, 49–58.

Chen, T. & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 785–794.

Chen, Y., Chen, C.-H., & Zaki, M. J. (2020). Combining user preferences and health needs in personalized food recommendation. *AMIA*.

Cheng, H., Rokicki, M., & Herder, E. (2017). The influence of city size on dietary choices and food recommendation. *Proceedings of the 25th conference on user modeling, adaptation and personalization*, 359–360.

Chhipa, S., Berwal, V., Hirapure, T., & Banerjee, S. (2022). Recipe recommendation system using tf-idf. *ITM web of conferences*, volume 44, 02006.

Choong, A. C. H. & Lee, N. K. (2017). Evaluation of convolutionary neural networks modeling of dna sequences using ordinal versus one-hot encoding method. *2017 International Conference on Computer and Drone Applications (IConDA)*, 60–65.

Christakopoulou, E. & Karypis, G. (2016). Local item-item models for top-n recommendation. *Proceedings of the 10th ACM conference on recommender systems*, 67–74.

Christakou, C., Vrettos, S., & Stafylopatis, A. (2007). A hybrid movie recommender system based on neural networks. *International Journal on Artificial Intelligence Tools*, 16(05), 771–792.

Claypool, M., Gokhale, A., Miranda, T., Murnikov, P., Netes, D., & Sartin, M. (1999). Combining content-based and collaborative filters in an online newspaper. *Proceedings of ACM SIGIR workshop on recommender systems*, volume 60, 1853–1870.

Codina, V., Ricci, F., & Ceccaroni, L. (2013). Exploiting the semantic similarity of contextual situations for pre-filtering recommendation. *User Modeling, Adaptation, and Personalization: 21th International Conference, UMAP 2013, Rome, Italy, June 10-14, 2013 Proceedings 21*, 165–177.

Cohen, D. A. & Babey, S. H. (2012). Contextual influences on eating behaviours: heuristic processing and dietary choices. *Obesity Reviews*, 13(9), 766–779.

Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. routledge.

Collis, J. & Hussey, R. (2021). *Business research: A practical guide for students*. Bloomsbury Publishing.

Colombo-Mendoza, L. O., Valencia-García, R., Rodríguez-González, A., Alor-Hernández, G., & Samper-Zapater, J. J. (2015). Recommetz: A context-aware knowledge-based mobile recommender system for movie showtimes. *Expert Systems with Applications*, 42(3), 1202–1222.

Condliff, M. K., Lewis, D. D., Madigan, D., & Posse, C. (1999). Bayesian mixed-effects models for recommender systems. *ACM SIGIR*, volume 99, 23–30.

Connors, M., Bisogni, C. A., Sobal, J., & Devine, C. M. (2001). Managing values in personal food systems. *Appetite*, 36(3), 189–200.

Contento, I. R. & Koch, P. A. (2020). *Nutrition education: Linking research, theory, and practice: Linking research, theory, and practice*. Jones & Bartlett Learning.

Counihan, C., Van Esterik, P., et al. (2013). *Food and culture*. Routledge New York, NY.

Creswell, J. W. & Clark, V. L. P. (2017). *Designing and conducting mixed methods research*. Sage publications.

Creswell, J. W. et al. (2009). Research design: Qualitative and mixed methods approaches. *London and Thousand Oaks: Sage Publications*.

Creswell, J. W. & Poth, C. N. (2016). *Qualitative inquiry and research design: Choosing among five approaches*. Sage publications.

Dang, C. N., Moreno-García, M. N., & Prieta, F. D. l. (2021). An approach to integrating sentiment analysis into recommender systems. *Sensors*, 21(16), 5666.

De Choudhury, M., Sharma, S., & Kiciman, E. (2016). Characterizing dietary choices, nutrition, and language in food deserts via social media. *Proceedings of the 19th acm conference on computer-supported cooperative work & social computing*, 1157–1170.

De Pessemier, T., Dooms, S., & Martens, L. (2013). A food recommender for patients in a care facility. *Proceedings of the 7th ACM conference on Recommender systems*, 209–212.

Dean, A. & Voss, D. (1999). *Design and analysis of experiments*. Springer.

Degemmis, M., Lops, P., & Semeraro, G. (2007). A content-collaborative recommender that exploits wordnet-based user profiles for neighborhood formation. *User Modeling and User-Adapted Interaction*, 17, 217–255.

Deldjoo, Y., Elahi, M., Quadrana, M., & Cremonesi, P. (2018). Using visual features based on mpeg-7 and deep learning for movie recommendation. *International journal of multimedia information retrieval*, 7, 207–219.

Denzin, N. K. & Lincoln, Y. S. (2011). *The Sage handbook of qualitative research*. sage.

Deshpande, M. & Karypis, G. (2004). Item-based top-n recommendation algorithms. *ACM Transactions on Information Systems (TOIS)*, 22(1), 143–177.

Desrosiers, C. & Karypis, G. (2010). A comprehensive survey of neighborhood-based recommendation methods. *Recommender systems handbook*, 107–144.

Devine, C. M. (2005). A life course perspective: understanding food choices in time, social location, and history. *Journal of nutrition education and behavior*, 37(3), 121–128.

Devine, C. M., Wolfe, W. S., Frongillo Jr, E. A., & Bisogni, C. A. (1999). Life-course events and experiences: association with fruit and vegetable consumption in 3 ethnic groups. *Journal of the American Dietetic Association*, 99(3), 309–314.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Di Noia, T., Mirizzi, R., Ostuni, V. C., Romito, D., & Zanker, M. (2012). Linked open data to support content-based recommender systems. *Proceedings of the 8th international conference on semantic systems*, 1–8.

Diaz, A. (2008). Through the google goggles: Sociopolitical bias in search engine design. *Web search: Multidisciplinary perspectives*, 11–34. Springer.

Díaz, A., García, A., & Gervás, P. (2008). User-centred versus system-centred evaluation of a personalization system. *Information Processing & Management*, 44(3), 1293–1307.

Dingemans, A. E., Martijn, C., van Furth, E. F., & Jansen, A. T. (2009). Expectations, mood, and eating behavior in binge eating disorder. beware of the bright side. *Appetite*, 53(2), 166–173.

Dixon, J., Woodman, D., Strazdins, L., Banwell, C., Broom, D., & Burgess, J. (2014). Flexible employment, flexible eating and health risks. *Critical Public Health*, 24(4), 461–475.

Dotdashmeredith, O. (2023). *Terms of service*. `https://www.dotdashmeredith.com/brands-termsofservice`. Accessed June 19

Drucker, H., Burges, C. J., Kaufman, L., Smola, A., & Vapnik, V. (1996). Support vector regression machines. *Advances in neural information processing systems*, 9.

Egmont-Petersen, M., Talmon, J. L., Hasman, A., & Ambergen, A. W. (1998). Assessing the importance of features for multi-layer perceptrons. *Neural networks*, 11(4), 623–635.

Ehatisham-Ul-Haq, M., Javed, A., Azam, M. A., Malik, H. M., Irtaza, A., Lee, I. H., & Mahmood, M. T. (2019). Robust human activity recognition using multimodal feature-level fusion. *IEEE Access*, 7, 60736–60751.

Ekstrand, M. D. & Willemsen, M. C. (2016). Behaviorism is not enough: better recommendations through listening to users. *Proceedings of the 10th ACM conference on recommender systems*, 221–224.

El Bilali, H., Callenius, C., Strassner, C., & Probst, L. (2019). Food and nutrition security and sustainability transitions in food systems. *Food and energy security*, 8(2), e00154.

El-Dosuky, M., Rashad, M. Z., Hamza, T., & El-Bassiouny, A. (2012). Food recommendation using ontology and heuristics. *Advanced Machine Learning Technologies and Applications: First International Conference, AMLTA 2012, Cairo, Egypt, December 8-10, 2012. Proceedings 1*, 423–429.

Elahi, M., Ricci, F., & Rubens, N. (2016). A survey of active learning in collaborative filtering recommender systems. *Computer Science Review*, 20, 29–50.

Elsweiler, D. & Harvey, M. (2015). Towards automatic meal plan recommendations for balanced nutrition. *Proceedings of the 9th ACM Conference on Recommender Systems*, 313–316.

Elsweiler, D., Harvey, M., Ludwig, B., & Said, A. (2015). Bringing the" healthy" into food recommenders. *Dmrs*, 1533, 33–36.

Elsweiler, D., Trattner, C., & Harvey, M. (2017). Exploiting food choice biases for healthier recipe recommendation. *Proceedings of the 40th international acm sigir conference on research and development in information retrieval*, 575–584.

Erlebacher, A. (1977). Design and analysis of experiments contrasting the within-and between-subjects manipulation of the independent variable. *Psychological Bulletin*, 84(2), 212.

ESHAresearch (2023). *Nutrition and supplement solutions analysis — labeling — compliance, your choice for nutrition software, databases, and consulting services.* `https://esha.com/`. Accessed June 21

Fakhri, A. A., Baizal, Z., & Setiawan, E. B. (2019). Restaurant recommender system using user-based collaborative filtering approach: a case study at bandung raya region. *Journal of Physics: Conference Series*, volume 1192, 012023.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G* power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods*, 39(2), 175–191.

Fda (2022). *How to understand and use the nutrition facts label.* `https://www.fda.gov/food/new-nutrition-facts-label/how-understand-and-use-nutrition-facts-label`. Accessed June 20

Felfernig, A., Jeran, M., Ninaus, G., Reinfrank, F., & Reiterer, S. (2013). Toward the next generation of recommender systems: applications and research challenges. *Multimedia Services in Intelligent Environments: Advances in Recommender Systems*, 81–98.

Ferdousi, Z. V., Negre, E., & Colazzo, D. (2017). Context factors in context-aware recommender systems. *AISR 2017: Atelier interdisciplinaire sur les systèmes de recommandation*.

Fereday, J. & Muir-Cochrane, E. (2006). Demonstrating rigor using thematic analysis: A hybrid approach of inductive and deductive coding and theme development. *International journal of qualitative methods*, 5(1), 80–92.

Fernández, M., Bellogín, A., & Cantador, I. (2021). Analysing the effect of recommendation algorithms on the amplification of misinformation. *arXiv preprint arXiv:2103.14748*.

Field, A. (2024). *Discovering statistics using IBM SPSS statistics*. Sage publications limited.

Forbes, P. & Zhu, M. (2011). Content-boosted matrix factorization for recommender systems: experiments with recipe recommendation. *Proceedings of the fifth ACM conference on Recommender systems*, 261–264.

Francis, J. J., Johnston, M., Robertson, C., Glidewell, L., Entwistle, V., Eccles, M. P., & Grimshaw, J. M. (2010). What is an adequate sample size? operationalising data saturation for theory-based interview studies. *Psychology and health*, 25(10), 1229–1245.

Freyne, J. & Berkovsky, S. (2010). Intelligent food planning: personalized recipe recommendation. *Proceedings of the 15th international conference on Intelligent user interfaces*, 321–324.

Freyne, J., Berkovsky, S., Baghaei, N., Kimani, S., & Smith, G. (2011). Personalized techniques for lifestyle change. *Conference on Artificial Intelligence in Medicine in Europe*, 139–148.

Fsa (2016). *Guide to creating a front of pack (fop) nutrition label for pre-packed products sold through retail outlets*. `https://www.food.gov.uk/sites/default/files/media/document/fop-guidance_0.pdf`. Accessed June 20

Fu, Z., Xian, Y., Gao, R., Zhao, J., Huang, Q., Ge, Y., Xu, S., Geng, S., Shah, C., Zhang, Y., et al. (2020). Fairness-aware explainable recommendation over knowledge graphs. *Proceedings of the 43rd international ACM SIGIR conference on research and development in information retrieval*, 69–78.

Funk, S. (2006). *Netflix update: Try this at home*. `https://sifter.org/simon/journal/20061211.html`. Accessed: 2023-09-09.

Furst, T., Connors, M., Bisogni, C. A., Sobal, J., & Falk, L. W. (1996). Food choice: a conceptual model of the process. *Appetite*, 26(3), 247–266.

Gabrielsson, A. & Lindström, E. (2010). The role of structure in the musical expression of emotions. *Handbook of music and emotion: Theory, research, applications*, 367400, 367–44.

Gao, X., Feng, F., He, X., Huang, H., Guan, X., Feng, C., Ming, Z., & Chua, T.-S. (2019). Hierarchical attention network for visually-aware food recommendation. *IEEE Transactions on Multimedia*, 22(6), 1647–1659.

Garcia, S. N., Osburn, B. I., & Jay-Russell, M. T. (2020). One health for food safety, food security, and sustainable food production. *Frontiers in Sustainable Food Systems*, 4, 1.

Ge, M., Elahi, M., Fernaández-Tobías, I., Ricci, F., & Massimo, D. (2015). Using tags and latent factors in a food recommender system. *Proceedings of the 5th international conference on digital health 2015*, 105–112.

Ghauth, K. I. B. & Abdullah, N. A. (2009). Building an e-learning recommender system using vector space model and good learners average rating. *2009 Ninth IEEE International Conference on Advanced Learning Technologies*, 194–196.

Gibbs, G. R. (2018). Analyzing qualitative data.

Goldberg, D., Nichols, D., Oki, B. M., & Terry, D. (1992). Using collaborative filtering to weave an information tapestry. *Communications of the ACM*, 35(12), 61–70.

Gong, B., Kaya, M., & Tintarev, N. (2020). Contextual personalized re-ranking of music recommendations through audio features. *arXiv preprint arXiv:2009.02782*.

Goodman, J. K., Cryder, C. E., & Cheema, A. (2013). Data collection in a flat world: The strengths and weaknesses of mechanical turk samples. *Journal of Behavioral Decision Making*, 26(3), 213–224.

Goody, J. (1982). *Cooking, Cuisine and Class: A Study in Comparative Sociology*. Cambridge University Press.

Gravetter, F. J., Forzano, L.-A. B., & Rakow, T. (2009). *Research methods for the behavioral sciences*. Wadsworth Cengage Learning Belmont, CA.

Grigorev, V., Li, J., Ma, W., He, Z., Zhang, M., Liu, Y., Yan, M., & Zhang, J. (2024). Situnes: A situational music recommendation dataset with physiological and psychological signals. *Proceedings of the 2024 Conference on Human Information Interaction and Retrieval*, 417–421.

Guan, X., Li, C.-T., & Guan, Y. (2017). Matrix factorization with rating completion: An enhanced svd model for collaborative filtering recommender systems. *IEEE access*, 5, 27668–27678.

Guest, G., Bunce, A., & Johnson, L. (2006). How many interviews are enough? an experiment with data saturation and variability. *Field methods*, 18(1), 59–82.

Guest, G., MacQueen, K. M., & Namey, E. E. (2012). *Applied thematic analysis*. sage.

Gunawardana, A. & Meek, C. (2009). A unified approach to building hybrid recommender systems. *Proceedings of the third ACM conference on Recommender systems*, 117–124.

Gunawardana, A., Shani, G., & Yogev, S. (2012). Evaluating recommender systems. *Recommender systems handbook*, 547–601. Springer.

Gupta, R., Nair, K., Mishra, M., Ibrahim, B., & Bhardwaj, S. (2024). Adoption and impacts of generative artificial intelligence: Theoretical underpinnings and research agenda. *International Journal of Information Management Data Insights*, 4(1), 100232.

Halcomb, E. J. & Davidson, P. M. (2006). Is verbatim transcription of interview data always necessary? *Applied nursing research*, 19(1), 38–42.

Hariri, N., Mobasher, B., & Burke, R. (2013). Query-driven context aware recommendation. *Proceedings of the 7th ACM conference on Recommender systems*, 9–16.

Harper, F. M. & Konstan, J. A. (2015). The movielens datasets: History and context. *Acm transactions on interactive intelligent systems (tiis)*, 5(4), 1–19.

Harris, J. L., Bargh, J. A., & Brownell, K. D. (2009). Priming effects of television food advertising on eating behavior. *Health psychology*, 28(4), 404.

Harvey, M. & Elsweiler, D. (2015). Automated recommendation of healthy, personalised meal plans. *Proceedings of the 9th ACM Conference on Recommender Systems*, 327–328.

Harvey, M., Ludwig, B., & Elsweiler, D. (2012). Learning user tastes: a first step to generating healthy meal plans. *First international workshop on recommendation technologies for lifestyle change (lifestyle 2012)*, volume 18.

Harvey, M., Ludwig, B., & Elsweiler, D. (2013). You are what you eat: Learning user tastes for rating prediction. *String Processing and Information Retrieval: 20th International Symposium, SPIRE 2013, Jerusalem, Israel, October 7-9, 2013, Proceedings 20*, 153–164.

Herlocker, J. L., Konstan, J. A., & Riedl, J. (2000). Explaining collaborative filtering recommendations. *Proceedings of the 2000 ACM conference on Computer supported cooperative work*, 241–250.

Hevner, K. (1935). The affective character of the major and minor modes in music. *The American Journal of Psychology*, 47(1), 103–118.

Hidasi, B., Karatzoglou, A., Baltrunas, L., & Tikk, D. (2015). Session-based recommendations with recurrent neural networks. *arXiv preprint arXiv:1511.06939*.

Hoerl, A. E. & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1), 55–67.

Hofmann, T. (1999). Probabilistic latent semantic indexing. *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*, 50–57.

Hofmann, T. (2001). Unsupervised learning by probabilistic latent semantic analysis. *Machine learning*, 42, 177–196.

Hofmann, T. (2004). Latent semantic models for collaborative filtering. *ACM Transactions on Information Systems (TOIS)*, 22(1), 89–115.

Howard, S., Adams, J., & White, M. (2012). Nutritional content of supermarket ready meals and recipes by television chefs in the united kingdom: cross sectional study. *Bmj*, 345.

Hu, W., Huang, G., Li, L., Zhang, L., Zhang, Z., & Liang, Z. (2020). Video-triggered eeg-emotion public databases and current methods: a survey. *Brain Science Advances*, 6(3), 255–287.

Hu, W., Zhang, Z., Zhao, H., Zhang, L., Li, L., Huang, G., & Liang, Z. (2023). Eeg microstate correlates of emotion dynamics and stimulation content during video watching. *Cerebral Cortex*, 33(3), 523–542.

Hyldelund, N. B., Frederiksen, C., Byrne, D. V., & Andersen, B. V. (2022). Is stress taking the pleasure out of food?—a characterization of the food pleasure profiles, appetite, and eating behaviors of people with chronic stress. *Foods*, 11(13), 1980.

Imani, M. & Montazer, G. A. (2019). A survey of emotion recognition methods with emphasis on e-learning environments. *Journal of Network and Computer Applications*, 147, 102423.

Isinkaye, F. O., Folajimi, Y. O., & Ojokoh, B. A. (2015). Recommendation systems: Principles, methods and evaluation. *Egyptian informatics journal*, 16(3), 261–273.

Islam, M. R., Trenholm, J., Rahman, A., Pervin, J., Ekström, E.-C., & Rahman, S. M. (2019). Sociocultural influences on dietary practices and physical activity behaviors of rural adolescents—a qualitative exploration. *Nutrients*, 11(12), 2916.

Jähne, B. (2005). *Digital image processing*. Springer Science & Business Media.

Jain, S., Grover, A., Thakur, P. S., & Choudhary, S. K. (2015). Trends, problems and solutions of recommender system. *International conference on computing, communication & automation*, 955–958.

James, G., Witten, D., Hastie, T., Tibshirani, R., et al. (2013). *An introduction to statistical learning*, volume 112. Springer.

Jannach, D., Zanker, M., Felfernig, A., & Friedrich, G. (2010). *Recommender systems: an introduction*. Cambridge University Press.

Jayedi, A., Soltani, S., Abdolshahi, A., & Shab-Bidar, S. (2020). Healthy and unhealthy dietary patterns and the risk of chronic disease: an umbrella review of meta-analyses of prospective cohort studies. *British Journal of Nutrition*, 124(11), 1133–1144.

Jeong, S.-Y. & Kim, Y.-K. (2021). Deep learning-based context-aware recommender system considering contextual features. *Applied Sciences*, 12(1), 45.

Jia, Y., Zhang, C., Lu, Q., & Wang, P. (2014). Users' brands preference based on svd++ in recommender systems. *2014 ieee workshop on advanced research and technology in industry applications (wartia)*, 1175–1178.

Jiawei, H. & Micheline, K. (2006). *Data mining: concepts and techniques*. Morgan kaufmann.

Jinnette, R., Narita, A., Manning, B., McNaughton, S. A., Mathers, J. C., & Livingstone, K. M. (2021). Does personalized nutrition advice improve dietary intake in healthy adults? a systematic review of randomized controlled trials. *Advances in Nutrition*, 12(3), 657–669.

Johnson, R. B. & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational researcher*, 33(7), 14–26.

Johnston, J. & Baumann, S. (2014). *Foodies: Democracy and distinction in the gourmet foodscape*. Routledge.

Joyce, J. (2006). Pandora and the music genome project, song structure analysis tools facilitate new music discovery. *Scientific Computing*, 23(14), 40–41.

Kabbur, S., Ning, X., & Karypis, G. (2013). Fism: factored item similarity models for top-n recommender systems. *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, 659–667.

Kalenkoski, C. M. & Hamrick, K. S. (2013). How does time poverty affect behavior? a look at eating and physical activity. *Applied Economic Perspectives and Policy*, 35(1), 89–105.

Karatzoglou, A., Amatriain, X., Baltrunas, L., & Oliver, N. (2010). Multiverse recommendation: n-dimensional tensor factorization for context-aware collaborative filtering. *Proceedings of the fourth ACM conference on Recommender systems*, 79–86.

Katarya, R. & Verma, O. P. (2017). An effective collaborative movie recommender system with cuckoo search. *Egyptian Informatics Journal*, 18(2), 105–112.

Kearney, J. (2010). Food consumption trends and drivers. *Philosophical transactions of the royal society B: biological sciences*, 365(1554), 2793–2807.

Keren, G. (2014). Between-or within-subjects design: A methodological dilemma. *A handbook for data analysis in the behaviorial sciences*, 1, 257–272.

Khanal, S. S., Prasad, P., Alsadoon, A., & Maag, A. (2020). A systematic review: machine learning based recommendation systems for e-learning. *Education and Information Technologies*, 25, 2635–2664.

Khusro, S., Ali, Z., & Ullah, I. (2016). Recommender systems: issues, challenges, and research opportunities. *Information science and applications (ICISA) 2016*, 1179–1189.

Kim, H.-Y. (2015). Statistical notes for clinical researchers: A one-way repeated measures anova for data with repeated observations. *Restorative dentistry & endodontics*, 40(1), 91.

Kim, J. & Gambino, A. (2016). Do we trust the crowd or information system? effects of personalization and bandwagon cues on users' attitudes and behavioral intentions toward a restaurant recommendation website. *Computers in Human Behavior*, 65, 369–379.

Kim, Y. & Shim, K. (2014). Twilite: A recommendation system for twitter using a probabilistic model based on latent dirichlet allocation. *Information Systems*, 42, 59–77.

King, R. S. (2015). *Cluster analysis and data mining: An introduction*. Mercury Learning and Information.

Kiran, R., Kumar, P., & Bhasker, B. (2020). Dnnrec: A novel deep learning based hybrid recommender system. *Expert Systems with Applications*, 144, 113054.

Klassen, A. C., Creswell, J., Plano Clark, V. L., Smith, K. C., & Meissner, H. I. (2012). Best practices in mixed methods for quality of life research. *Quality of life Research*, 21, 377–380.

Klingbeil, A., Grützner, C., & Schreck, P. (2024). Trust and reliance on ai—an experimental study on the extent and costs of overreliance on ai. *Computers in Human Behavior*, 160, 108352.

Knijnenburg, B. P. & Kobsa, A. (2013). Making decisions about privacy: information disclosure in context-aware recommender systems. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 3(3), 1–23.

Kohavi, R., Deng, A., Frasca, B., Walker, T., Xu, Y., & Pohlmann, N. (2013). Online controlled experiments at large scale. *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, 1168–1176.

Kohavi, R. & Longbotham, R. (2015). Online controlled experiments and a/b tests. *Encyclopedia of machine learning and data mining*, 1–11.

Konstan, J. A., Miller, B. N., Maltz, D., Herlocker, J. L., Gordon, L. R., & Riedl, J. (1997). Grouplens: Applying collaborative filtering to usenet news. *Communications of the ACM*, 40(3), 77–87.

Konttinen, H. (2020). Emotional eating and obesity in adults: the role of depression, sleep and genes. *Proceedings of the Nutrition Society*, 79(3), 283–289.

Koohi, H. & Kiani, K. (2016). User based collaborative filtering using fuzzy c-means. *Measurement*, 91, 134–139.

Koren, Y. (2008). Factorization meets the neighborhood: a multifaceted collaborative filtering model. *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, 426–434.

Koren, Y., Bell, R., & Volinsky, C. (2009). Matrix factorization techniques for recommender systems. *Computer*, 42(8), 30–37.

Koren, Y., Rendle, S., & Bell, R. (2021). Advances in collaborative filtering. *Recommender systems handbook*, 91–142.

Kouki, P., Fakhraei, S., Foulds, J., Eirinaki, M., & Getoor, L. (2015). Hyper: A flexible and extensible probabilistic framework for hybrid recommender systems. *Proceedings of the 9th ACM Conference on Recommender Systems*, 99–106.

Kula, M. (2015). Metadata embeddings for user and item cold-start recommendations. *arXiv preprint arXiv:1507.08439*.

Kunaver, M. & Požrl, T. (2017). Diversity in recommender systems–a survey. *Knowledge-based systems*, 123, 154–162.

Kusmierczyk, T., Trattner, C., & Nørvåg, K. (2015a). Temporal patterns in online food innovation. *Proceedings of the 24th international conference on world wide web*, 1345–1350.

Kusmierczyk, T., Trattner, C., & Nørvåg, K. (2015b). Temporality in online food recipe consumption and production. *Proceedings of the 24th international conference on world wide web*, 55–56.

Kvale, S. & Brinkmann, S. (2009). *Interviews: Learning the craft of qualitative research interviewing*. sage.

Lai, C.-H., Liu, D.-R., & Lien, K.-S. (2021). A hybrid of xgboost and aspect-based review mining with attention neural network for user preference prediction. *International Journal of Machine Learning and Cybernetics*, 12, 1203–1217.

Lajoie-O'Malley, A., Bronson, K., van der Burg, S., & Klerkx, L. (2020). The future (s) of digital agriculture and sustainable food systems: An analysis of high-level policy documents. *Ecosystem Services*, 45, 101183.

Lapadat, J. C. & Lindsay, A. C. (1999). Transcription in research and practice: From standardization of technique to interpretive positionings. *Qualitative inquiry*, 5(1), 64–86.

Lee, A., Cardel, M., & Donahoo, W. T. (2019). Social and environmental factors influencing obesity. *Endotext [Internet]*.

Lee, D. & Seung, H. S. (2000). Algorithms for non-negative matrix factorization. *Advances in neural information processing systems*, 13.

Lee, T. Q., Park, Y., & Park, Y.-T. (2007). A similarity measure for collaborative filtering with implicit feedback. *Advanced Intelligent Computing Theories and Applications. With Aspects of Artificial Intelligence: Third International Conference on Intelligent Computing, ICIC 2007, Qingdao, China, August 21-24, 2007. Proceedings 3*, 385–397.

Lee, W.-P., Chen, C.-T., Huang, J.-Y., & Liang, J.-Y. (2017). A smartphone-based activity-aware system for music streaming recommendation. *Knowledge-Based Systems*, 131, 70–82.

Lhuissier, A., Tichit, C., Caillavet, F., Cardon, P., Masullo, A., Martin-Fernandez, J., Parizot, I., & Chauvin, P. (2013). Who still eats three meals a day? findings from a quantitative survey in the paris area. *Appetite*, 63, 59–69.

Li, H., Cui, J., Shen, B., & Ma, J. (2016). An intelligent movie recommendation system through group-level sentiment analysis in microblogs. *Neurocomputing*, 210, 164–173.

Li, P. & Yamada, S. (2004). A movie recommender system based on inductive learning. *IEEE Conference on Cybernetics and Intelligent Systems, 2004.*, volume 1, 318–323.

Liang, J. (2020). Image classification based on resnet. *Journal of Physics: Conference Series*, volume 1634, 012110.

Lin, C.-J., Kuo, T.-T., & Lin, S.-D. (2014). A content-based matrix factorization model for recipe recommendation. *Advances in Knowledge Discovery and Data Mining: 18th Pacific-Asia Conference, PAKDD 2014, Tainan, Taiwan, May 13-16, 2014. Proceedings, Part II 18*, 560–571.

Linden, G. D., Jacobi, J. A., & Benson, E. A. (2001). *Collaborative recommendations using item-to-item similarity mappings*. U.S. Patent No. 6,266,649. Washington, DC: U.S. Patent and Trademark Office.

Liu, Q., Hu, J., Xiao, Y., Gao, J., & Zhao, X. (2023). Multimodal recommender systems: A survey. *arXiv preprint arXiv:2302.03883*.

Liu, Y.-l., Yan, W., Hu, B., Li, Z., & Lai, Y. L. (2022). Effects of personalization and source expertise on users' health beliefs and usage intention toward health chatbots: Evidence from an online experiment. *Digital Health*, 8, 20552076221129718.

Livne, A., Unger, M., Shapira, B., & Rokach, L. (2019). Deep context-aware recommender system utilizing sequential latent context. *arXiv preprint arXiv:1909.03999*.

Lombardi, S., Anand, S. S., & Gorgoglione, M. (2009). Context and customer behaviour in recommendation.

Lopez-Arevalo, I., Aldana-Bobadilla, E., Molina-Villegas, A., Galeana-Zapién, H., Muñiz-Sanchez, V., & Gausin-Valle, S. (2020). A memory-efficient encoding method for processing mixed-type data on machine learning. *Entropy*, 22(12), 1391.

Lops, P., De Gemmis, M., & Semeraro, G. (2011). Content-based recommender systems: State of the art and trends. *Recommender systems handbook*, 73–105.

Low, M. X., Yap, T. T. V., Soo, W. K., Ng, H., Goh, V. T., Chin, J. J., & Kuek, T. Y. (2022). Comparison of label encoding and evidence counting for malware classification. *Journal of System and Management Sciences*, 12(6), 17–30.

Lu, J., Wu, D., Mao, M., Wang, W., & Zhang, G. (2015). Recommender system application developments: a survey. *Decision support systems*, 74, 12–32.

Lü, L., Medo, M., Yeung, C. H., Zhang, Y.-C., Zhang, Z.-K., & Zhou, T. (2012). Recommender systems. *Physics reports*, 519(1), 1–49.

Lu, Y., Dong, R., & Smyth, B. (2018). Why i like it: multi-task learning for recommendation and explanation. *Proceedings of the 12th ACM Conference on Recommender Systems*, 4–12.

Macht, M. (2008). How emotions affect eating: A five-way model. *Appetite*, 50(1), 1–11.

Mack, N. (2005). Qualitative research methods: A data collector's field guide.

Magnini, B. & Strapparava, C. (2000). Experiments in word domain disambiguation for parallel texts. *ACL-2000 Workshop on Word Senses and Multi-linguality*, 27–33.

Magnini, B. & Strapparava, C. (2001). Improving user modelling with content-based techniques. *User Modeling 2001: 8th International Conference, UM 2001 Sonthofen, Germany, July 13–17, 2001 Proceedings 8*, 74–83.

Maheshwari, S. & Chourey, M. (2019). Recipe recommendation system using machine learning models. *International Research Journal of Engineering and Technology (IRJET)*, 6(9), 366–369.

Maia, R. & Ferreira, J. C. (2018). Context-aware food recommendation system. *Context-aware food recommendation system*, 349–356.

Maier, A. & Cash, P. (2023). Designing for human behaviour in a systemic world. *Handbook of Engineering Systems Design*, 1–34. Springer.

Maruyama, T., Kawano, Y., & Yanai, K. (2012). Real-time mobile recipe recommendation system using food ingredient recognition. *Proceedings of the 2nd ACM international workshop on Interactive multimedia on mobile and portable devices*, 27–34.

Mason, M. et al. (2010). Sample size and saturation in phd studies using qualitative interviews. *Forum qualitative Sozialforschung/Forum: qualitative social research*, volume 11.

Matthews, R. & Ross, E. (2010). *Research methods: A practical guide for the social sciences*. Pearson Education Ltd.

Maxwell, S. E., Delaney, H. D., & Kelley, K. (2017). *Designing experiments and analyzing data: A model comparison perspective*. Routledge.

Melville, P., Mooney, R. J., Nagarajan, R., et al. (2002). Content-boosted collaborative filtering for improved recommendations. *Aaai/iaai*, 23, 187–192.

Milosavljevic, M., Navalpakkam, V., Koch, C., & Rangel, A. (2012). Relative visual saliency differences induce sizable bias in consumer choice. *Journal of consumer psychology*, 22(1), 67–74.

Min, S.-H. & Han, I. (2005). Recommender systems using support vector machines. *International Conference on Web Engineering*, 387–393.

Min, W., Jiang, S., & Jain, R. (2019a). Food recommendation: Framework, existing solutions, and challenges. *IEEE Transactions on Multimedia*, 22(10), 2659–2671.

Min, W., Jiang, S., Liu, L., Rui, Y., & Jain, R. (2019b). A survey on food computing. *ACM Computing Surveys (CSUR)*, 52(5), 1–36.

Mintz, S. W. (1986). *Sweetness and power: The place of sugar in modern history*. Penguin.

Mintz, S. W. & Du Bois, C. M. (2002). The anthropology of food and eating. *Annual review of anthropology*, 31(1), 99–119.

Missbach, B., Florack, A., & König, J. (2015). Mental imagery and food consumption. *Frontiers in psychiatry*, 6, 48.

Miyahara, K. & Pazzani, M. J. (2000). Collaborative filtering with the simple bayesian classifier. *Pacific Rim International conference on artificial intelligence*, 679–689.

Mohamed, M. H., Khafagy, M. H., & Ibrahim, M. H. (2019). Recommender systems challenges and solutions survey. *2019 international conference on innovative trends in computer engineering (ITCE)*, 149–155.

Molnar, A. (2019). Smartriqs: A simple method allowing real-time respondent interaction in qualtrics surveys. *Journal of Behavioral and Experimental Finance*, 22, 161–169.

Monterrosa, E. C., Frongillo, E. A., Drewnowski, A., de Pee, S., & Vandevijvere, S. (2020). Sociocultural influences on food choices and implications for sustainable healthy diets. *Food and Nutrition Bulletin*, 41(2_suppl), 59S–73S.

Montoya, A. K. (2023). Selecting a within-or between-subject design for mediation: Validity, causality, and statistical power. *Multivariate Behavioral Research*, 58(3), 616–636.

Morgan, D. L. (2014). Pragmatism as a paradigm for mixed methods research. *Integrating qualitative and quantitative methods*, 55, 25–44.

Mozaffarian, D. (2016). Dietary and policy priorities for cardiovascular disease, diabetes, and obesity: a comprehensive review. *Circulation*, 133(2), 187–225.

Mu, Y., Xiao, N., Tang, R., Luo, L., & Yin, X. (2019). An efficient similarity measure for collaborative filtering. *Procedia computer science*, 147, 416–421.

Mummolo, J. & Peterson, E. (2019). Demand effects in survey experiments: An empirical assessment. *American Political Science Review*, 113(2), 517–529.

Muñoz-Vilches, N. C., Van Trijp, H. C., & Piqueras-Fiszman, B. (2020). Pleasure or health? the role of mental simulation in desire and food choices. *Foods*, 9(8), 1099.

Murphy, A. B., Jordan-Bychkov, T. G., & Jordan, B. B. (2020). *The European culture area: A systematic geography*. Rowman & Littlefield Publishers.

Mustaqeem, A., Anwar, S. M., & Majid, M. (2020). A modular cluster based collaborative recommender system for cardiac patients. *Artificial intelligence in medicine*, 102, 101761.

Musto, C. (2010). Enhanced vector space models for content-based recommender systems. *Proceedings of the fourth ACM conference on Recommender systems*, 361–364.

Musto, C., Franza, T., Semeraro, G., De Gemmis, M., & Lops, P. (2018). Deep content-based recommender systems exploiting recurrent neural networks and linked open data. *Adjunct Publication of the 26th conference on user modeling, adaptation and personalization*, 239–244.

Musto, C., Semeraro, G., De Gemmis, M., & Lops, P. (2016). Learning word embeddings from wikipedia for content-based recommender systems. *Advances in Information Retrieval: 38th European Conference on IR Research, ECIR 2016, Padua, Italy, March 20–23, 2016. Proceedings 38*, 729–734.

Myers, A. & Hansen, C. (2006). *Experimental psychology*. Thomson Wadsworth.

Nanda, A., Mohapatra, B. B., Mahapatra, A. P. K., Mahapatra, A. P. K., & Mahapatra, A. P. K. (2021). Multiple comparison test by tukey's honestly significant difference (hsd): Do the confident level control type i error. *International Journal of Statistics and Applied Mathematics*, 6(1), 59–65.

Naumov, M., Mudigere, D., Shi, H.-J. M., Huang, J., Sundaraman, N., Park, J., Wang, X., Gupta, U., Wu, C.-J., Azzolini, A. G., et al. (2019). Deep learning recommendation model for personalization and recommendation systems. *arXiv preprint arXiv:1906.00091*.

Neysiani, B. S., Soltani, N., Mofidi, R., & Nadimi-Shahraki, M. H. (2019). Improve performance of association rule-based collaborative filtering recommendation systems using genetic algorithm. *International Journal of Information Technology and Computer Science*, 11(2), 48–55.

Nidhi, R. & Annappa, B. (2017). Twitter-user recommender system using tweets: A content-based approach. *2017 International Conference on Computational Intelligence in Data Science (ICCIDS)*, 1–6.

Ning, J., Yang, J., Jiang, S., Zhang, L., & Yang, M.-H. (2016). Object tracking via dual linear structured svm and explicit feature map. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 4266–4274.

Ning, X. & Karypis, G. (2011). Slim: Sparse linear methods for top-n recommender systems. *2011 IEEE 11th international conference on data mining*, 497–506.

Noorian, A. (2024). A personalized context and sequence aware point of interest recommendation. *Multimedia Tools and Applications*, 1–30.

Nowell, L. S., Norris, J. M., White, D. E., & Moules, N. J. (2017). Thematic analysis: Striving to meet the trustworthiness criteria. *International journal of qualitative methods*, 16(1), 1609406917733847.

O'Connor, D. B., Jones, F., Conner, M., McMillan, B., & Ferguson, E. (2008). Effects of daily hassles and eating style on eating behavior. *Health Psychology*, 27(1S), S20.

O'Donovan, J. & Smyth, B. (2005). Trust in recommender systems. *Proceedings of the 10th international conference on Intelligent user interfaces*, 167–174.

Oh, Y., Choi, A., & Woo, W. (2010). u-babsang: a context-aware food recommendation system. *the Journal of Supercomputing*, 54, 61–81.

Oku, K., Nakajima, S., Miyazaki, J., & Uemura, S. (2006). Context-aware svm for context-dependent information recommendation. *7th International Conference on Mobile Data Management (MDM'06)*, 109–109.

Ono, C., Takishima, Y., Motomura, Y., & Asoh, H. (2009). Context-aware preference model based on a study of difference between real and supposed situation data. *User Modeling, Adaptation, and Personalization: 17th International Conference, UMAP 2009, formerly UM and AH, Trento, Italy, June 22-26, 2009. Proceedings 17*, 102–113.

Organization, W. H. (2024). *Obesity and overweight*. `https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight`

Ornab, A. M., Chowdhury, S., & Toa, S. B. (2019). An empirical analysis of collaborative filtering algorithms for building a food recommender system. *Data and Communication Networks: Proceedings of GUCON 2018*, 147–157.

Orne, M. T. (2017). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *Sociological methods*, 279–299. Routledge.

Ouhbi, B., Frikh, B., Zemmouri, E., & Abbad, A. (2018). Deep learning based recommender systems. *2018 ieee 5th international congress on information science and technology (cist)*, 161–166.

Palan, S. & Schitter, C. (2018). Prolific. ac—a subject pool for online experiments. *Journal of Behavioral and Experimental Finance*, 17, 22–27.

Panniello, U., Tuzhilin, A., Gorgoglione, M., Palmisano, C., & Pedone, A. (2009). Experimental comparison of pre-vs. post-filtering approaches in context-aware recommender systems. *Proceedings of the third ACM conference on Recommender systems*, 265–268.

Papagelis, M., Rousidis, I., Plexousakis, D., & Theoharopoulos, E. (2005). Incremental collaborative filtering for highly-scalable recommendation algorithms. *Foundations of Intelligent Systems: 15th International Symposium, ISMIS 2005, Saratoga Springs, NY, USA, May 25-28, 2005. Proceedings 15*, 553–561.

Pariser, E. (2011). *The filter bubble: What the Internet is hiding from you.* penguin UK.

Park, D. H., Kim, H. K., Choi, I. Y., & Kim, J. K. (2012). A literature review and classification of recommender systems research. *Expert systems with applications*, 39(11), 10059–10072.

Park, M.-H., Park, H.-S., & Cho, S.-B. (2008). Restaurant recommendation for group of people in mobile environments using probabilistic multi-criteria decision making. *Computer-Human Interaction: 8th Asia-Pacific Conference, APCHI 2008 Seoul, Korea, July 6-9, 2008 Proceedings 8*, 114–122.

Patra, B. K., Launonen, R., Ollikainen, V., & Nandi, S. (2015). A new similarity measure using bhattacharyya coefficient for collaborative filtering in sparse data. *Knowledge-Based Systems*, 82, 163–177.

Patton, M. Q. (2014). *Qualitative research & evaluation methods: Integrating theory and practice.* Sage publications.

Pazzani, M. J. (1999). A framework for collaborative, content-based and demographic filtering. *Artificial intelligence review*, 13, 393–408.

Pecune, F., Callebert, L., & Marsella, S. (2020). A recommender system for healthy and personalized recipes recommendations. *HealthRecSys@ RecSys*, 15–20.

Peer, E., Brandimarte, L., Samat, S., & Acquisti, A. (2017). Beyond the turk: Alternative platforms for crowdsourcing behavioral research. *Journal of experimental social psychology*, 70, 153–163.

Penney, T. L., Almiron-Roig, E., Shearer, C., McIsaac, J.-L., & Kirk, S. F. (2014). Modifying the food environment for childhood obesity prevention: challenges and opportunities. *Proceedings of the Nutrition Society*, 73(2), 226–236.

Pennington, J., Socher, R., & Manning, C. D. (2014). Glove: Global vectors for word representation. *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, 1532–1543.

Pinho, M., Mackenbach, J., Charreire, H., Oppert, J.-M., Bárdos, H., Glonti, K., Rutter, H., Compernolle, S., De Bourdeaudhuij, I., Beulens, J., et al. (2018). Exploring the relationship between perceived barriers to healthy eating and dietary behaviours in european adults. *European journal of nutrition*, 57, 1761–1770.

Polignano, M., Narducci, F., de Gemmis, M., & Semeraro, G. (2021). Towards emotion-aware recommender systems: an affective coherence model based on emotion-driven behaviors. *Expert Systems with Applications*, 170, 114382.

Popescul, A., Ungar, L. H., Pennock, D. M., & Lawrence, S. (2013). Probabilistic models for unified collaborative and content-based recommendation in sparse-data environments. *arXiv preprint arXiv:1301.2303*.

Poslavskaya, E. & Korolev, A. (2023). Encoding categorical data: Is there yet anything'hotter'than one-hot encoding? *arXiv preprint arXiv:2312.16930*.

Premchaiswadi, W., Poompuang, P., Jongswat, N., & Premchaiswadi, N. (2013). Enhancing diversity-accuracy technique on user-based top-n recommendation algorithms. *2013 IEEE 37th Annual Computer Software and Applications Conference Workshops*, 403–408.

Pu, P., Chen, L., & Hu, R. (2012). Evaluating recommender systems from the user's perspective: survey of the state of the art. *User Modeling and User-Adapted Interaction*, 22, 317–355.

Qomariyah, N. N. (2018). *Pairwise preferences learning for recommender systems*. University of York.

Qomariyah, N. N., Kazakov, D., & Fajar, A. N. (2020). Predicting user preferences with xgboost learning to rank method. *2020 3rd International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*, 123–128.

Raine, K. D. (2005). Determinants of healthy eating in canada: an overview and synthesis. *Canadian journal of public health*, 96(3), S8–S15.

Ramirez-Garcia, X. & García-Valdez, M. (2014). Post-filtering for a restaurant context-aware recommender system. *Recent advances on hybrid approaches for designing intelligent systems*, 695–707.

Ramos, J. et al. (2003). Using tf-idf to determine word relevance in document queries. *Proceedings of the first instructional conference on machine learning*, volume 242, 29–48.

Rashid, A. M., Albert, I., Cosley, D., Lam, S. K., McNee, S. M., Konstan, J. A., & Riedl, J. (2002). Getting to know you: learning new user preferences in recommender systems. *Proceedings of the 7th international conference on Intelligent user interfaces*, 127–134.

Raza, S. & Ding, C. (2019). Progress in context-aware recommender systems—an overview. *Computer Science Review*, 31, 84–97.

Reid, M. & Hammersley, R. (1999). The effects of sucrose and maize oil on subsequent food intake and mood. *British Journal of Nutrition*, 82(6), 447–455.

Reimers, N. & Gurevych, I. (2019). Sentence-bert: Sentence embeddings using siamese bert-networks. *arXiv preprint arXiv:1908.10084*.

Reis, H. T. (2018). Why researchers should think "real-world": A conceptual rationale. *Relationships, well-being and behaviour*, 249–271. Routledge.

Rendle, S., Gantner, Z., Freudenthaler, C., & Schmidt-Thieme, L. (2011). Fast context-aware recommendations with factorization machines. *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*, 635–644.

Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., & Riedl, J. (1994). Grouplens: An open architecture for collaborative filtering of netnews. *Proceedings of the 1994 ACM conference on Computer supported cooperative work*, 175–186.

Resnick, P. & Varian, H. R. (1997). Recommender systems. *Communications of the ACM*, 40(3), 56–58.

Ricci, F. (2022). Recommender systems in tourism. *Handbook of e-Tourism*, 457–474. Springer.

Ricci, F., Rokach, L., & Shapira, B. (2010). Introduction to recommender systems handbook. *Recommender systems handbook*, 1–35. Springer.

Ricci, F., Rokach, L., & Shapira, B. (2015). Recommender systems: introduction and challenges. *Recommender systems handbook*, 1–34. Springer.

Ricci, F. & Werthner, H. (2006). Introduction to the special issue: Recommender systems. *International Journal of Electronic Commerce*, 11(2), 5–9.

Rich, E. (1979). User modeling via stereotypes. *Cognitive science*, 3(4), 329–354.

Robles, V., Larranaga, P., Menasalvas, E., Pérez, M. S., & Herves, V. (2003). Improvement of naive bayes collaborative filtering using interval estimation. *Proceedings IEEE/WIC International Conference on Web Intelligence (WI 2003)*, 168–174.

Robson, C. (2002). *Real world research*, volume 2. Blackwell Oxford.

Rodríguez Sánchez, A., Salmerón Gómez, R., & García, C. (2022). The coefficient of determination in the ridge regression. *Communications in statistics-simulation and computation*, 51(1), 201–219.

Rokicki, M., Herder, E., Kuśmierczyk, T., & Trattner, C. (2016). Plate and prejudice: Gender differences in online cooking. *Proceedings of the 2016 conference on user modeling adaptation and personalization*, 207–215.

Rokicki, M., Herder, E., & Trattner, C. (2017). How editorial, temporal and social biases affect online food popularity and appreciation. *Proceedings of the International AAAI Conference on Web and Social Media*, volume 11, 192–200.

Rokicki, M., Trattner, C., & Herder, E. (2018). The impact of recipe features, social cues and demographics on estimating the healthiness of online recipes. *Proceedings of the international AAAI conference on web and social media*, volume 12.

Rostami, M., Oussalah, M., & Farrahi, V. (2022). A novel time-aware food recommender-system based on deep learning and graph clustering. *IEEE Access*, 10, 52508–52524.

Roy, D. & Dutta, M. (2022). A systematic review and research perspective on recommender systems. *Journal of Big Data*, 9(1), 59.

Roza, A. M. & Shizgal, H. M. (1984). The harris benedict equation reevaluated: resting energy requirements and the body cell mass. *The American journal of clinical nutrition*, 40(1), 168–182.

Rozin, P. (1996). The socio-cultural context of eating and food choice. *Food choice, acceptance and consumption*, 83–104. Springer.

Rozin, P. (2007). Food and eating. *Handbook of cultural psychology*, 2, 447–477.

Rubens, N., Elahi, M., Sugiyama, M., & Kaplan, D. (2015). Active learning in recommender systems. *Recommender systems handbook*, 809–846.

Sagi, O. & Rokach, L. (2021). Approximating xgboost with an interpretable decision tree. *Information sciences*, 572, 522–542.

Salas, C. E., Radovic, D., & Turnbull, O. H. (2012). Inside-out: Comparing internally generated and externally generated basic emotions. *Emotion*, 12(3), 568.

Santana, I. A. P. & Domingues, M. A. (2020). A systematic review on context-aware recommender systems using deep learning and embeddings. *arXiv preprint arXiv:2007.04782*.

Sarwar, B., Karypis, G., Konstan, J., & Riedl, J. (2001). Item-based collaborative filtering recommendation algorithms. *Proceedings of the 10th international conference on World Wide Web*, 285–295.

Sarwar, B., Karypis, G., Konstan, J., & Riedl, J. (2002). Incremental singular value decomposition algorithms for highly scalable recommender systems. *Fifth international conference on computer and information science*, volume 1, 27–8.

Saunders, M., Lewis, P., & Thornhill, A. (2009). *Research methods for business students*. Pearson education.

Scalvedi, M. L., Gennaro, L., Saba, A., & Rossi, L. (2021). Relationship between nutrition knowledge and dietary intake: an assessment among a sample of italian adults. *Frontiers in Nutrition*, 8, 714493.

Schafer, J. B., Frankowski, D., Herlocker, J., & Sen, S. (2007). Collaborative filtering recommender systems. *The adaptive web: methods and strategies of web personalization*, 291–324. Springer.

Schedl, M. (2019). Deep learning in music recommendation systems. *Frontiers in Applied Mathematics and Statistics*, 5, 457883.

Schein, A. I., Popescul, A., Ungar, L. H., & Pennock, D. M. (2002). Methods and metrics for cold-start recommendations. *Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval*, 253–260.

Scheiner, S. M. (2020). Manova: multiple response variables and multispecies interactions. *Design and analysis of ecological experiments*, 94–112. Chapman and Hall/CRC.

Schwarzer, R. (2008). Modeling health behavior change: How to predict and modify the adoption and maintenance of health behaviors. *Applied psychology*, 57(1), 1–29.

Seifu, C. N., Fahey, P. P., Hailemariam, T. G., Frost, S. A., & Atlantis, E. (2021). Dietary patterns associated with obesity outcomes in adults: An umbrella review of systematic reviews. *Public Health Nutrition*, 24(18), 6390–6414.

Semeraro, G., Basile, P., de Gemmis, M., & Lops, P. (2009). User profiles for personalizing digital libraries. *Handbook of Research on Digital Libraries: Design, Development, and Impact*, 149–158. IGI Global.

Seo, S., Huang, J., Yang, H., & Liu, Y. (2017). Interpretable convolutional neural networks with dual local and global attention for review rating prediction. *Proceedings of the eleventh ACM conference on recommender systems*, 297–305.

Shahbazi, Z. & Byun, Y.-C. (2019). Product recommendation based on content-based filtering using xgboost classifier. *Int. J. Adv. Sci. Technol*, 29, 6979–6988.

Shahbazi, Z., Hazra, D., Park, S., & Byun, Y. C. (2020). Toward improving the prediction accuracy of product recommendation system using extreme gradient boosting and encoding approaches. *Symmetry*, 12(9), 1566.

Shakirova, E. (2017). Collaborative filtering for music recommender system. *2017 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus)*, 548–550.

Shang, M.-S., Zhang, Z.-K., Zhou, T., & Zhang, Y.-C. (2010). Collaborative filtering with diffusion-based similarity on tripartite graphs. *Physica A: Statistical Mechanics and its Applications*, 389(6), 1259–1264.

Shani, G. & Gunawardana, A. (2011). Evaluating recommendation systems. *Recommender systems handbook*, 257–297.

Shani, G., Heckerman, D., Brafman, R. I., & Boutilier, C. (2005). An mdp-based recommender system. *Journal of machine Learning research*, 6(9).

Shardanand, U. & Maes, P. (1995). Social information filtering: Algorithms for automating "word of mouth". *Proceedings of the SIGCHI conference on Human factors in computing systems*, 210–217.

Sharma, L. & Gera, A. (2013a). A survey of recommendation system: Research challenges. *International Journal of Engineering Trends and Technology (IJETT)*, 4(5), 1989–1992.

Sharma, L. & Gera, A. (2013b). A survey of recommendation system: Research challenges. *International Journal of Engineering Trends and Technology (IJETT)*, 4(5), 1989–1992.

Sharma, M. & Mann, S. (2013). A survey of recommender systems: approaches and limitations. *International Journal of Innovations in Engineering and Technology*, 2(2), 8–14.

Shepherd, R. & Raats, M. (2006). *The psychology of food choice*. Cabi.

Shi, C., Zhang, Z., Luo, P., Yu, P. S., Yue, Y., & Wu, B. (2015). Semantic path based personalized recommendation on weighted heterogeneous information networks. *Proceedings of the 24th ACM international on conference on information and knowledge management*, 453–462.

Shorten, A. & Smith, J. (2017). *Mixed methods research: expanding the evidence base*.

Siddik, M. B. S. & Wibowo, A. T. (2023). Collaborative filtering based food recommendation system using matrix factorization. *Jurnal Media Informatika Budidarma*, 7(3), 1041–1049.

Siddiqui, S. A., Erol, Z., Rugji, J., Taşçı, F., Kahraman, H. A., Toppi, V., Musa, L., Di Giacinto, G., Bahmid, N. A., Mehdizadeh, M., et al. (2023). An overview of fermentation in the food industry-looking back from a new perspective. *Bioresources and Bioprocessing*, 10(1), 85.

Simkus, J. (2023). Between-subjects design: Overview & examples. *Simply Psychology*.

Singh, D. & Singh, B. (2020). Investigating the impact of data normalization on classification performance. *Applied Soft Computing*, 97, 105524.

Sobal, J., Bisogni, C. A., Devine, C. M., & Jastran, M. (2006). A conceptual model of the food choice process over the life course. *The psychology of food choice*, 1–18. Cabi Wallingford UK.

Sobal, J., Bisogni, C. A., & Jastran, M. (2014). Food choice is multifaceted, contextual, dynamic, multilevel, integrated, and diverse. *Mind, Brain, and Education*, 8(1), 6–12.

Sobal, J. & Nelson, M. K. (2003). Commensal eating patterns: a community study. *Appetite*, 41(2), 181–190.

Sobecki, J., Babiak, E., & Słanina, M. (2006). Application of hybrid recommendation in web-based cooking assistant. *Knowledge-Based Intelligent Information and Engineering Systems: 10th International Conference, KES 2006, Bournemouth, UK, October 9-11, 2006. Proceedings, Part III 10*, 797–804.

Soboroff, I. & Nicholas, C. (1999). Combining content and collaboration in text filtering. *Proceedings of the IJCAI*, volume 99, 86–91.

Somarathna, R., Bednarz, T., & Mohammadi, G. (2022). Virtual reality for emotion elicitation–a review. *IEEE Transactions on Affective Computing*.

Song, Y., Elkahky, A. M., & He, X. (2016). Multi-rate deep learning for temporal recommendation. *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*, 909–912.

Sookrah, R., Dhowtal, J. D., & Nagowah, S. D. (2019). A dash diet recommendation system for hypertensive patients using machine learning. *2019 7th international conference on information and communication technology (ICoICT)*, 1–6.

Spence, C. (2021). Explaining seasonal patterns of food consumption. *International journal of gastronomy and food science*, 24, 100332.

Starke, A. (2019). Recsys challenges in achieving sustainable eating habits. *HealthRecSys@RecSys*, 29–30.

Starke, A. D. & Trattner, C. (2021). Promoting healthy food choices online: a case for multi-list recommender systems. *Proceedings of the ACM IUI 2021 Workshops*.

Starke, A. D., Willemsen, M. C., & Snijders, C. (2020). With a little help from my peers: Depicting social norms in a recommender interface to promote energy conservation. *Proceedings of the 25th international conference on intelligent user interfaces*, 568–578.

Starke, A. D., Willemsen, M. C., & Trattner, C. (2021). Nudging healthy choices in food search through visual attractiveness. *Frontiers in Artificial Intelligence*, 4, 621743.

Stebbins, R. A. (2001). *Exploratory research in the social sciences*, volume 48. Sage.

Story, M., Kaphingst, K. M., Robinson-O'Brien, R., & Glanz, K. (2008). Creating healthy food and eating environments: policy and environmental approaches. *Annu. Rev. Public Health*, 29(1), 253–272.

Su, X. & Khoshgoftaar, T. M. (2009). A survey of collaborative filtering techniques. *Advances in artificial intelligence*, 2009(1), 421425.

Su, X., Kubat, M., & Tapia, M. A. (2005). Query size estimation using clustering techniques. *17th IEEE International Conference on Tools with Artificial Intelligence (ICTAI'05)*, 5–pp.

Sun, Y., Han, J., Yan, X., Yu, P. S., & Wu, T. (2011). Pathsim: Meta path-based top-k similarity search in heterogeneous information networks. *Proceedings of the VLDB Endowment*, 4(11), 992–1003.

Symons, M. (1991). *Eating into Thinking: Explorations in the sociology of cuisine*. Flinders University of South Australia.

Tarus, J. K., Niu, Z., & Mustafa, G. (2018). Knowledge-based recommendation: a review of ontology-based recommender systems for e-learning. *Artificial intelligence review*, 50, 21–48.

Teng, C.-Y., Lin, Y.-R., & Adamic, L. A. (2012). Recipe recommendation using ingredient networks. *Proceedings of the 4th annual ACM web science conference*, 298–307.

Tennyson, R. D. & Breuer, K. (2002). Improving problem solving and creativity through use of complex-dynamic simulations. *Computers in Human Behavior*, 18(6), 650–668.

Terveen, L., Hill, W., Amento, B., McDonald, D., & Creter, J. (1997). Phoaks: A system for sharing recommendations. *Communications of the ACM*, 40(3), 59–62.

Tintarev, N. & Masthoff, J. (2015). Explaining recommendations: Design and evaluation. *Recommender systems handbook*, 353–382. Springer.

Tiu Wright, L., Nancarrow, C., & Kwok, P. M. (2001). Food taste preferences and cultural influences on consumption. *British food journal*, 103(5), 348–357.

Tkalčič, M., Burnik, U., Odić, A., Košir, A., & Tasič, J. (2013). Emotion-aware recommender systems–a framework and a case study. *ICT Innovations 2012: Secure and Intelligent Systems*, 141–150.

Torres, S. J. & Nowson, C. A. (2007). Relationship between stress, eating behavior, and obesity. *Nutrition*, 23(11-12), 887–894.

Tran, T. & Cohen, R. (2000). Hybrid recommender systems for electronic commerce. *Proc. Knowledge-Based Electronic Markets, Papers from the AAAI Workshop, Technical Report WS-00-04, AAAI Press*, volume 40.

Tran, T. N. T., Felfernig, A., Trattner, C., & Holzinger, A. (2021). Recommender systems in the healthcare domain: state-of-the-art and research issues. *Journal of Intelligent Information Systems*, 57, 171–201.

Trang Tran, T. N., Atas, M., Felfernig, A., & Stettinger, M. (2018). An overview of recommender systems in the healthy food domain. *Journal of Intelligent Information Systems*, 50, 501–526.

Trattner, C. & Elsweiler, D. (2017a). Food recommender systems: important contributions, challenges and future research directions. *arXiv preprint arXiv:1711.02760*.

Trattner, C. & Elsweiler, D. (2017b). Investigating the healthiness of internet-sourced recipes: implications for meal planning and recommender systems. *Proceedings of the 26th international conference on world wide web*, 489–498.

Trattner, C., Elsweiler, D., & Howard, S. (2017a). Estimating the healthiness of internet recipes: a cross-sectional study. *Frontiers in public health*, 5, 16.

Trattner, C., Moesslang, D., & Elsweiler, D. (2018). On the predictability of the popularity of online recipes. *EPJ Data Science*, 7(1), 1–39.

Trattner, C., Rokicki, M., & Herder, E. (2017b). On the relations between cooking interests, hobbies and nutritional values of online recipes: Implications for health-aware recipe recommender systems. *Adjunct publication of the 25th conference on user modeling, adaptation and personalization*, 59–64.

Trochim, W. M. & Donnelly, J. P. (2001). *Research methods knowledge base*, volume 2. Atomic dog publishing Cincinnati, OH.

TTCI, C. I. (2013). The travel & tourism competitiveness index 2013: Contributing to national growth and employment. *The Travel & Tourism Competitiveness Report 2013*, 3.

Turner III, D. W. (2010). Qualitative interview design: A practical guide for novice investigators. *The qualitative report*, 15(3), 754.

Udilă, A. (2023). Encoding methods for categorical data: A comparative analysis for linear models, decision trees, and support vector machines.

Ueta, T., Iwakami, M., & Ito, T. (2011). A recipe recommendation system based on automatic nutrition information extraction. *Knowledge Science, Engineering and Management: 5th International Conference, KSEM 2011, Irvine, CA, USA, December 12-14, 2011. Proceedings 5*, 79–90.

Ungar, L. H. & Foster, D. P. (1998). Clustering methods for collaborative filtering. *AAAI workshop on recommendation systems*, volume 1, 114–129.

Vairale, V. S. & Shukla, S. (2021). Recommendation of food items for thyroid patients using content-based knn method. *Data Science and Security: Proceedings of IDSCS 2020*, 71–77.

Vallejo-Correa, P., Monsalve-Pulido, J., & Tabares-Betancur, M. (2021). A systematic mapping review of context-aware analysis and its approach to mobile learning and ubiquitous learning processes. *Computer Science Review*, 39, 100335.

Van den Oord, A., Dieleman, S., & Schrauwen, B. (2013). Deep content-based music recommendation. *Advances in neural information processing systems*, 26.

Van Rijsbergen, C. (1979). Information retrieval: theory and practice. *Proceedings of the joint IBM/University of Newcastle upon tyne seminar on data base systems*, volume 79, 1–14.

van Strien, T., Cebolla, A., Etchemendy, E., Gutierrez-Maldonado, J., Ferrer-Garcia, M., Botella, C., & Baños, R. (2013). Emotional eating and food intake after sadness and joy. *Appetite*, 66, 20–25.

Vasile, F., Smirnova, E., & Conneau, A. (2016). Meta-prod2vec: Product embeddings using side-information for recommendation. *Proceedings of the 10th ACM conference on recommender systems*, 225–232.

Vigni, M. L., Durante, C., & Cocchi, M. (2013). Exploratory data analysis. *Data handling in science and technology*, volume 28, 55–126. Elsevier.

Vijayakumar, V., Vairavasundaram, S., Logesh, R., & Sivapathi, A. (2019). Effective knowledge based recommender system for tailored multiple point of interest recommendation. *International Journal of Web Portals (IJWP)*, 11(1), 1–18.

Walker, E. B. & Boyer, D. M. (2018). Research as storytelling: The use of video for mixed methods research. *Video Journal of Education and Pedagogy*, 3, 1–12.

Wang, W., Duan, L.-Y., Jiang, H., Jing, P., Song, X., & Nie, L. (2021). Market2dish: Health-aware food recommendation. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 17(1), 1–19.

Wansink, B., Painter, J. E., & Lee, Y.-K. (2006). The office candy dish: proximity's influence on estimated and actual consumption. *International journal of obesity*, 30(5), 871–875.

Werner, D., Cruz, C., & Nicolle, C. (2013). Ontology-based recommender system of economic articles. *arXiv preprint arXiv:1301.4781*.

Who, J. & Consultation, F. E. (2003). Diet, nutrition and the prevention of chronic diseases. *World Health Organ Tech Rep Ser*, 916(i-viii), 1–149.

Wickens, T. D. & Keppel, G. (2004). *Design and analysis: A researcher's handbook*. Pearson Prentice-Hall Upper Saddle River, NJ.

Wold, S., Esbensen, K., & Geladi, P. (1987). Principal component analysis. *Chemometrics and intelligent laboratory systems*, 2(1-3), 37–52.

Wolstenholme, H., Kelly, C., Hennessy, M., & Heary, C. (2020). Childhood fussy/picky eating behaviours: A systematic review and synthesis of qualitative studies. *International Journal of Behavioral Nutrition and Physical Activity*, 17(1), 1–22.

Wong, T.-T. & Yeh, P.-Y. (2019). Reliable accuracy estimates from k-fold cross validation. *IEEE Transactions on Knowledge and Data Engineering*, 32(8), 1586–1594.

Wu, W., He, L., & Yang, J. (2012). Evaluating recommender systems. *Seventh International Conference on Digital Information Management (ICDIM 2012)*, 56–61.

Wu, X. & Sun, G. (2024). Context-aware music recommendation algorithm combining classification and collaborative filtering. *Scalable Computing: Practice and Experience*, 25(3), 1923–1931.

Xiong, R., Wang, J., Zhang, N., & Ma, Y. (2018). Deep hybrid collaborative filtering for web service recommendation. *Expert systems with Applications*, 110, 191–205.

Yakhchi, S. (2021). Learning complex users' preferences for recommender systems. *arXiv preprint arXiv:2107.01529*.

Yang, L., Hsieh, C.-K., Yang, H., Pollak, J. P., Dell, N., Belongie, S., Cole, C., & Estrin, D. (2017). Yum-me: a personalized nutrient-based meal recommender system. *ACM Transactions on Information Systems (TOIS)*, 36(1), 1–31.

Yao, L., Sheng, Q. Z., Ngu, A. H., Yu, J., & Segev, A. (2014). Unified collaborative and content-based web service recommendation. *IEEE Transactions on Services Computing*, 8(3), 453–466.

Yin, R. K. (2009). *Case study research: Design and methods*, volume 5. sage.

Yoshii, K., Goto, M., Komatani, K., Ogata, T., & Okuno, H. G. (2008). An efficient hybrid music recommender system using an incrementally trainable probabilistic generative model. *IEEE Transactions on Audio, Speech, and Language Processing*, 16(2), 435–447.

Yu, X., Ren, X., Sun, Y., Gu, Q., Sturt, B., Khandelwal, U., Norick, B., & Han, J. (2014). Personalized entity recommendation: A heterogeneous information network approach. *Proceedings of the 7th ACM international conference on Web search and data mining*, 283–292.

Zang, T., Zhu, Y., Liu, H., Zhang, R., & Yu, J. (2022). A survey on cross-domain recommendation: taxonomies, methods, and future directions. *ACM Transactions on Information Systems*, 41(2), 1–39.

Zarzour, H., Al-Sharif, Z., Al-Ayyoub, M., & Jararweh, Y. (2018). A new collaborative filtering recommendation algorithm based on dimensionality reduction and clustering techniques. *2018 9th international conference on information and communication systems (ICICS)*, 102–106.

Zhang, Q., Elsweiler, D., & Trattner, C. (2020a). Visual cultural biases in food classification. *Foods*, 9(6), 823.

Zhang, Q., Elsweiler, D., & Trattner, C. (2023). Understanding and predicting cross-cultural food preferences with online recipe images. *Information Processing & Management*, 60(5), 103443.

Zhang, S., Yao, L., Sun, A., & Tay, Y. (2019). Deep learning based recommender system: A survey and new perspectives. *ACM computing surveys (CSUR)*, 52(1), 1–38.

Zhang, T. & Iyengar, V. S. (2002). Recommender systems using linear classifiers. *The Journal of Machine Learning Research*, 2, 313–334.

Zhang, Y., Ai, Q., Chen, X., & Croft, W. B. (2017). Joint representation learning for top-n recommendation with heterogeneous information sources. *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, 1449–1458.

Zhang, Y., Chen, X., et al. (2020b). Explainable recommendation: A survey and new perspectives. *Foundations and Trends® in Information Retrieval*, 14(1), 1–101.

Zhang, Z., Wang, Z., Ma, T., Taneja, V. S., Nelson, S., Le, N. H. L., Murugesan, K., Ju, M., Chawla, N. V., Zhang, C., et al. (2024). Mopi-hfrs: A multi-objective personalized health-aware food recommendation system with llm-enhanced interpretation. *arXiv preprint arXiv:2412.08847*.

Zheng, L. & Zhu, F. (2017). Preference integration in context-aware recommendation. *Database Systems for Advanced Applications: 22nd International Conference, DASFAA 2017, Suzhou, China, March 27-30, 2017, Proceedings, Part I 22*, 475–489.

Zheng, Y., Burke, R., & Mobasher, B. (2013). Recommendation with differential context weighting. *User Modeling, Adaptation, and Personalization: 21th International Conference, UMAP 2013, Rome, Italy, June 10-14, 2013 Proceedings 21*, 152–164.

Zhou, X., He, J., Huang, G., & Zhang, Y. (2015). Svd-based incremental approaches for recommender systems. *Journal of Computer and System Sciences*, 81(4), 717–733.

Zhu, Y.-X., Huang, J., Zhang, Z.-K., Zhang, Q.-M., Zhou, T., & Ahn, Y.-Y. (2013). Geography and similarity of regional cuisines in china. *PloS one*, 8(11), e79161.

Zhuhadar, L., Nasraoui, O., Wyatt, R., & Romero, E. (2009). Multi-model ontology-based hybrid recommender system in e-learning domain. *2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology*, volume 3, 91–95.

Zikmund, W. G., Babin, B. J., Carr, J. C., & Griffin, M. (2013). *Business research methods*. Cengage learning.

Zioutos, K., Kondylakis, H., & Stefanidis, K. (2023). Healthy personalized recipe recommendations for weekly meal planning. *Computers*, 13(1), 1.

Appendices

Appendix A

Research Stage 1: Ethics Review Documentation and Approval

Downloaded: 09/10/2024
Approved: 11/04/2022

Mengyisong Zhao
Registration number: 200296416
Information School [a.k.a iSchool]
Programme: PhD information studies

Dear Mengyisong

PROJECT TITLE: Understanding food perceptions and needs under various contextual situation
APPLICATION: Reference Number 045474

On behalf of the University ethics reviewers who reviewed your project, I am pleased to inform you that on 11/04/2022 the above-named project was **approved** on ethics grounds, on the basis that you will adhere to the following documentation that you submitted for ethics review:

- University research ethics application form 045474 (form submission date: 07/04/2022); (expected project end date: 01/07/2022).
- Participant information sheet 1103875 version 4 (07/04/2022).
- Participant consent form 1103876 version 1 (23/03/2022).

The following amendments to this application have been approved:

- Amendment approved: 17/06/2022

If during the course of the project you need to deviate significantly from the above-approved documentation please inform me since written approval will be required.

Your responsibilities in delivering this research project are set out at the end of this letter.

Yours sincerely

Peter Bath
Ethics Admin
Information School [a.k.a iSchool]

Please note the following responsibilities of the researcher in delivering the research project:

- The project must abide by the University's Research Ethics Policy: https://www.sheffield.ac.uk/research-services/ethics-integrity/policy
- The project must abide by the University's Good Research & Innovation Practices Policy: https://www.sheffield.ac.uk/polopoly_fs/1.671066!/file/GRIPPolicy.pdf
- The researcher must inform their supervisor (in the case of a student) or Ethics Admin (in the case of a member of staff) of any significant changes to the project or the approved documentation.
- The researcher must comply with the requirements of the law and relevant guidelines relating to security and confidentiality of personal data.
- The researcher is responsible for effectively managing the data collected both during and after the end of the project in line with best practice, and any relevant legislative, regulatory or contractual requirements.

Application 045474

Amendment - Complete (Submitted on 17/06/2022)

Delete

Description of changes

This study will be advertised using the volunteer lists for staff or students maintained by CiCS, and a small amount of reward will be considered to provide to participants. The reasons to make these changes are to ensure that sufficient participants can be recruited in diverse groups.

Additional ethical considerations

Do the proposed changes pose any additional ethical considerations?
No

Additional risks

Do any of the proposed amendments to the research potentially change the risk for any of the researchers?
No

Supporting documentation revisions

Do the proposed amendments require revisions to any of the supporting documentation? Please note that when uploading new versions of documents which you have previously provided, you should give a description of the document which clearly indicates that this is a new version, e.g. by providing an appropriate version number. It is also helpful to the reviewers if you clearly mark the changes you have made in the document itself (e.g. by highlighting new text or using tracked changes in Word).
No

Other relevant information

Decision

should be approved

Comments on approval:

Please ensure that the "small amount of reward will be considered to provide to participants" will be small enough so that participants are not persuaded to take part against their better judgement just for the sake of the money. This is not likely to happen with this project, but I suggest you discuss the amount with your supervisor and email Peter Bath (p.a.bath@sheffield.ac.uk) and cc. to ischool_ethics@sheffield.ac.uk. Thank you, Peter Bath (Ethics Co-ordinator).

Original application

Section A: Applicant details

Date application started:
Tue 1 March 2022 at 16:49

First name:
Mengyisong

Last name:
Zhao

Email:
mzhao18@sheffield.ac.uk

Programme name:

PhD information studies

Module name:
PhD information studies
Last updated:
06/07/2022

Department:
Information School [a.k.a iSchool]

Applying as:
Postgraduate research

Research project title:
Understanding food perceptions and needs under various contextual situation

Has your research project undergone academic review, in accordance with the appropriate process?
Yes

Similar applications:
- not entered -

Section B: Basic information

Supervisor

Name	Email
Frank Hopfgartner	f.hopfgartner@sheffield.ac.uk

Proposed project duration

Start date (of data collection):
Sun 1 May 2022

Anticipated end date (of project)
Fri 1 July 2022

3: Project code (where applicable)

Project externally funded?
No

Project code
- not entered -

Suitability

Takes place outside UK?
No

Involves NHS?
No

Health and/or social care human-interventional study?
No

ESRC funded?
No

Likely to lead to publication in a peer-reviewed journal?
No

Led by another UK institution?
No

Involves human tissue?
No

Clinical trial or a medical device study?
No

Involves social care services provided by a local authority?
No

Is social care research requiring review via the University Research Ethics Procedure
No

Involves adults who lack the capacity to consent?
No

Involves research on groups that are on the Home Office list of 'Proscribed terrorist groups or organisations?
No

Indicators of risk

Involves potentially vulnerable participants?
No
Involves potentially highly sensitive topics?
No

Section C: Summary of research

1. Aims & Objectives

The aim of this study is to understand people’s eating perceptions and needs, as well as to explore how their eating preferences change under different contextual situations (i.e., age, gender, professional, cultural background, etc). The research objectives are as follows:
1. To investigate peoples’ perceptions towards food choice and needs under different contextual situations.
2. To identify the most influential (novel) contextual features that might influence food choice.
3. To explore how these contextual features affect nutritional intake.

The study intends to seek answers to the following research questions:
1. Do eating perceptions and preferences differ between people under different contextual situations?
2. How do contextual factors affect people’s food choice and preferences?
3. To what extent does the nutritional content of food affect their choices?
4. Are people aware of the nutrition information of their chosen food and can they acquire enough support to gather this information?

2. Methodology

The purpose of this study is to gain new insight from interviewees about what contextual factors impact their food choice and nutritional intake as well as to understand their eating perceptions and needs. Semi-structured interviews are considered as an appropriate method, this type of interview provides more space of expression for participants. The interview will be divided into four parts: demographic information, personal factors, food-based factors and online recipes. Each interview will take around 20 to 30 minutes. As part of consenting to taking part in this study, interviews will be recorded, and then transcribed by the researcher. The interview recording will not be labelled with any personal information such as name or workplace of interviewees’ and it will be listened to by only the researcher and her supervisors for the purposes of transcription. The data will be used to identify the most contributing factors that affect peoples’ food choice and guide for the second and third stage of the research (what contextual factors would impact the user’s recipe rating behaviour and whether integrating these factors can improve the recommender system performance. The data collected in this stage will be used for publications, but the personal information of interviewees will be completely anonymized. This study will not directly quote the opinions of interviewees. All the personal data will be anonymised and assigned to a specific code (e.g., the name of the participant will be assigned to identified participant ID:0001).

Data Collection:
As this research aims to understand the human beings' eating perceptions and needs, we aim to recruit participants of diverse backgrounds. For example, participants associated with the university and those not associated with the university, as well as participants with different cultural backgrounds and in different life stages. No particular sampling strategy will be adopted in this study. We intend to reach out to a minimum of 12 participants but consider a larger sample size as well.

Data Analysis:
A thematic analysis will be adopted to analyse the interview data. Braun and Clarke (2006) suggested that certain steps be followed in thematic analysis, including being familiar with the data, creating initial codes, looking for potential themes and identifying themes, indicating what themes means and what data is expected within each theme. This research will go through these steps during the thematic

analysis for its qualitative investigation.

3. Personal Safety

Have you completed your departmental risk assessment procedures, if appropriate?

Yes

Raises personal safety issues?

No

This study poses minimal risks to the researcher given that the location of the setting of data collection will take place either virtually or in a public area (i.e., the iSchool iSpace). There will be no restriction for participants, the interviews will take place based on their choice. If held virtually, the video chat software such as Microsoft Teams, Skype, Google meet or Zoom will be used to conduct the interviews. If held in person, the interview will take place in a public area, for example, iSpace in Information School, there are glass walls so anyone can see what is going on there. The researcher will not be exposed to risks or harms greater than those that encounters in her normal life.

Section D: About the participants

1. Potential Participants

To achieve the aforementioned research questions, and to understand the human's eating perceptions and needs, their attitude about nutrition information, as well as to explore how their eating preferences changed under various contextual situations. The researcher should aim to recruit participants in a diverse group of people (including but not limited to different age, gender, professional, culture background), which provide a rich demographic contextual background for researchers to analyse and seek the patterns. Only participants aged over 18 will be included in the selection of this study.

2. Recruiting Potential Participants

Researcher has plan to advertise this study on various platforms (i.e., mailing lists) to reach a diverse audience. Potential interviewees will be sent an email with pre-interview information, which will include:
1- Background information about the researcher (including the name of University and research group), and the aims of research.
2- A brief summary of the research aims and expected outcomes.
3- A copy of the ethical approval. Including the names of supervisors and their emails.
4- A copy of the consent form.
5- A clear statement about the participants' rights, including confidentiality and anonymity of information.

Once the participants have shown interest in this study, they will be asked to reply to the email as well as answer the demographic questions, including the gender groups (males and females), age groups (18-24, 25-34, 35-44, 45-54, 55-64 and 65 and over) and race group (White, Black or African American, American Indian or Alaska Native, Asian, Native Hawaiian or Other Pacific Islander). Then, the researchers will reply back to them and invite the potential participants to choose a time, place and platform for the interview that is most convenient for them. The first strategy of recruiting participants will be done by sending emails to potential participants. In case an insufficient number of participants responds to the initial call for participation, the researcher will call for participants on other platforms.

Regardless of the platform used to conduct the interviews, all interviews will be audio-recorded with an external recorder after obtaining necessary permissions from the participants. If needed, a post-interview follow-up will be arranged between the researcher and the interviewee to clarify or validate uncertain points in the transcript.

2.1. Advertising methods

Will the study be advertised using the volunteer lists for staff or students maintained by IT Services? No

- not entered -

3. Consent

Will informed consent be obtained from the participants? (i.e. the proposed process) Yes

Before conducting the interview, the participants will be asked for permission to audio-record the interview, and give their agreement to use their data for analytics and publication purposes. The researchers have prepared a consent form and an information sheet. These documents will provide prospective subjects with detailed information about the purpose of the study, procedures, data protection, voluntary participation, and any potential risks or benefits that may arise as a result of their participation. They can also give the subjects information on ways to withdraw from the study, along with information on when it may no longer be possible for their data to be removed (for example, after publication). In addition, the participant will be assured that even after signing the consent form, it is still possible to withdraw from the project/ interview at any time.

Prior to the interview, both the informed consent form and information sheet will be emailed to potential participants in order to obtain their initial consent. Then, the participant will be given 14 days to study the information sheet and to ask inquiries as needed. If the participant agrees to take part in this study, he/she will be asked to sign the consent form electronically, while both the researcher and the participant will retain an electronic copy of it.

4. Payment

Will financial/in kind payments be offered to participants? No

5. Potential Harm to Participants

What is the potential for physical and/or psychological harm/distress to the participants?

The research process has little to no risk of physical harm. Participants' personal identities will be anonymised and pseudonyms assigned prior to transcription. Identities and any distinguishing characteristics indicated in the interview will be omitted from the interview transcripts to ensure that participants cannot be identified from the text. In addition to previous measures, the participants will be informed about their right to refuse to answer any sensitive questions that may increase their anxiety/ or distress during the interview. Also, they will be given the opportunity to terminate the interview at any point of time, and request the researcher to destroy their data permanently. Moreover, the interviewees will be protected through the signed informed consent form which will allow him/her to raise any complaints about the researcher if he/or she feels that their data is compromised or made public without formally obtaining the necessary permissions. All the personal data will be anonymised and assigned to a specific code (e.g., the name of the participant will be assigned to identified participant ID:0001). The interview won't go into detail about participants who suffer from eating disorders, diabetes or obsities, because the research focuses on understanding how people make their food choice decision, only if the participant is willing to share this information. If any mental depression emerges during the interview, the interview will stop immediately. If necessary, the University mental health service will be contacted for support.

How will this be managed to ensure appropriate protection and well-being of the participants?

All research participants will be assured that their participation in this study is voluntary and free of pressure, and they can withdraw from the study at any time without any stressful consequences. To ensure that the participants have read and understood their rights and responsibilities, they will be asked to provide a signature in both the consent form and information sheet.

6. Potential harm to others who may be affected by the research activities

Which other people, if any, may be affected by the research activities, beyond the participants and the research team?

A possible potential harm might occur if the interviewees are mentioning information about their family members or friends.

What is the potential for harm to these people?

A harm of sharing information about the experience and opinions of their family members or friends from the participants without their consents.

How will this be managed to ensure appropriate safeguarding of these people?

By reminding the participants of the privacy of others and not to mention any sensitive information that is not needed by this interview. This issue will also be considered when writing up findings for the dissertation and subsequent publications, checking that third parties can not be identified directly or through the context of discussion or quotation.

7. Reporting of safeguarding concerns or incidents

What arrangements will be in place for participants, and any other people external to the University who are involved in, or affected by, the research, to enable reporting of incidents or concerns?

I mentioned in the participants' sheet the following statement, in case something went wrong during the interview.
" If you feel that your personal data has not dealt correctly as per information provided in this sheet, or wish to raise any concerns/ or complaint about the research, you can first discuss this with the principal researcher via this email address (mzhao18@sheffield.ac.uk) or her supervisors Dr. Frank Hopgartner f.hopfgartner@sheffield.ac.uk and Dr. Morgan Harvey m.harvey@sheffield.ac.uk. Your complaint will be dealt with respectfully, and we will respond appropriately and as soon as possible. However, if you feel that your complaint has not been dealt with appropriately, then you can email the research supervisor via their email address. In addition, if you wish to complain about any other serious problems that may arise during or following your participation in the research, you can contact the University's 'Registrar and Secretary'."

Who will be the Designated Safeguarding Contact(s)?

Dr. Morgan Harvey m.harvey@sheffield.ac.uk and Dr. Frank Hopgartner f.hopfgartner@sheffield.ac.uk.

How will reported incidents or concerns be handled and escalated?

By discussing the reported issues internally with my supervisors. And, if needed, the reported issues or concerns can be raised and escalated to the university's ethics committee.

Section E: About the data

1. Data Processing

Will you be processing (i.e. collecting, recording, storing, or otherwise using) personal data as part of this project? (Personal data is any information relating to an identified or identifiable living person).
Yes

Which organisation(s) will act as Data Controller?

University of Sheffield only

2. Legal basis for processing of personal data

The University considers that for the vast majority of research, 'a task in the public interest' (6(1)(e)) will be the most appropriate legal basis. If, following discussion with the UREC, you wish to use an alternative legal basis, please provide details of the legal basis, and the reasons for applying it, below:

- not entered -

Will you be processing (i.e. collecting, recording, storing, or otherwise using) 'Special Category' personal data?
Yes

The University considers the most appropriate condition to be that 'processing is necessary for archiving purposes in the public interest, scientific research purposes or statistical purposes' (9(2)(j)) . If, following discussion with the UREC, you wish to use an alternative condition, please provide details of the condition, and the reasons for applying it, below:

- not entered -

3. Data Confidentiality

What measures will be put in place to ensure confidentiality of personal data, where appropriate?

All the data will only be saved and backed up at Google Drive and shared with the supervisors of researchers. Personal data of the interviewee will be pseudonymised by storing personal details (e.g., name) separately and creating a 'key' or 'code' to enable re-identification. To be specific, the participant name will be replaced immediately after the interview by assigning a random code/or number which makes it difficult for the stranger to identify the true identity of the participants. The data will be generalised to remove certain identifiers without compromising the data's accuracy. Likewise, all data will be anonymized, and the researcher will ensure that the participants are aware of this and that the dissertation and related publications do not reveal any name, job title, organisation or identifiable data that could lead to the identification of any participant.

4. Data Storage and Security

In general terms, who will have access to the data generated at each stage of the research, and in what form

In general, the researcher and her supervisors will have access to the anonymised research data at all stages of the research from the data collection to the archival/or deletion of research data. The researcher only will have access to data without anonymisation and be responsible to store/or move the relevant research data into the university cloud drive, and ensure that it is backed up on a regular basis. Also, the researcher will be responsible to move the research data to the archive if required by the university and agreed by the participants.

What steps will be taken to ensure the security of data processed during the project, including any identifiable personal data, other than those already described earlier in this form?

Throughout the project lifecycle, all research data will be saved anonymously in Google Drive which can only be accessed by the main researcher and her supervisors. When working off-campus, the researcher will make sure to use the University of Sheffield VPN.

Will all identifiable personal data be destroyed once the project has ended?
Yes

Please outline when this will take place (this should take into account regulatory and funder requirements).

After the award of PhD has been granted, all research data will be destroyed.

Section F: Supporting documentation

Information & Consent

Participant information sheets relevant to project?

Yes

Document 1103875 (Version 4) All versions

Consent forms relevant to project?
Yes

Document 1103876 (Version 1) All versions

Additional Documentation

External Documentation

- not entered -

Section G: Declaration

Signed by:
Mengyisong Zhao
Date signed:
Thu 7 April 2022 at 16:52

Offical notes

- not entered -

Understanding Food Perceptions and Needs Under Various Contextual Situation Participant Consent Form

Please fill out this consent form if you are willing to participate in our research study. If you have any questions about the study, please refer to the information sheet provided or contact Mengyisong Zhao via mzhao18@sheffield.ac.uk.

* Indicates required question

1. Email *

____

Taking part in the research project

Please tick the appropriate boxes.

2. I have read and understood the project information sheet dated or the project has been fully explained to me. (If you will answer No to this question, please do not proceed with this consent form until you are fully aware of what your participation in the project will mean.) *

   Mark only one oval.

   ◯ Yes

   ◯ No

3. I have been given the opportunity to ask questions about the project. *

   Mark only one oval.

   ◯ Yes

   ◯ No

4. I agree to take part in the project. I understand that taking part in the project will include being interviewed AND recorded (audio as agreed upon before the interview). *

   Mark only one oval.

   - Yes
   - No

5. I understand that by choosing to participate as a volunteer in this research, this does not create a legally binding agreement nor is it intended to create an employment relationship with the University of Sheffield. *

   Mark only one oval.

   - Yes
   - No

6. I understand that my taking part is voluntary and that I can withdraw from the study at any time/before data has been anonymised, analysed or published. I do not have to give any reasons for why I no longer want to take part and there will be no adverse consequences if I choose to withdraw. *

   Mark only one oval.

   - Yes
   - No

How my information will be used during and after the project

7. I understand my personal details such as name, phone number, address and email address etc. will not be revealed to people outside the project. *

   Mark only one oval.

   - Yes
   - No

8. I understand and agree that my words may be quoted in publications, reports, web pages, and other research outputs. I understand that I will not be named in these outputs unless I specifically request this. *

Mark only one oval.

- Yes
- No

9. I understand and agree that other authorised researchers will have access to this data only if they agree to preserve the confidentiality of the information as requested in this form. *

Mark only one oval.

- Yes
- No

10. I understand and agree that other authorised researchers may use my data in publications, reports, web pages, and other research outputs, only if they agree to preserve the confidentiality of the information as requested in this form. *

Mark only one oval.

- Yes
- No

11. I give permission for the anonymized interview transcript that I provide to be deposited in the White Rose Thesis Repository so it can be used for future research and learning *

Mark only one oval.

- Yes
- No

So that the information you provide can be used legally by the researchers

12. I agree to assign the copyright I hold in any materials generated as part of this project to The University of Sheffield. *

Mark only one oval.

- Yes
- No

13. Please type in your name to sign this consent form. *

The University Of Sheffield.

Information School

Participant Information Sheet

Project Title: Understanding food perceptions and needs under various contextual situations

You are being invited to take part in a research project. Before you decide whether or not to participate, it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part. Thank you for reading this.

1. *What is the purpose of the study?*

The purpose of this study is to understand your eating perceptions and needs, how you normally make food choice decisions, whether you are aware of nutritional information before you make food choices, as well as to explore what contextual factors (i.e., different age, gender, professional, seasonal change or emotional change) affect your choice.

2. *Why have I been chosen?*

You have been invited to take part in this research because you are an adult, speak English, and are willing to share with us your eating preferences.

3. *Do I have to take part?*

Taking part in this research is entirely voluntary. If you decide to take part, you will be provided with a separate Google consent form and asked to sign it and submit it to the researcher. Before conducting the interview, you will be asked for permission to audio-record the interview. In addition, you will be assured that, even after signing the consent form, it is still possible to withdraw from the interview at any time and you do not have to give a reason. If you choose to withdraw your participation then any and all data collected about you and during any studies you have participated in will be deleted. Please note that by choosing to participate in this research, this will not create a legally binding agreement, nor is it intended to create an employment relationship between you and the University of Sheffield.

4. *What will happen to me if I take part?*

If you agree to participate, you will be asked to confirm your participation by signing a consent form. You will be then invited to participate in a one-to-one interview. The one-to-one interview will take place at a convenient time/place for you, over the Internet (using, e.g., Skype or Google meet platforms) or in person in the Sheffield Information School iSpace. The interview will involve questions about your daily food choices, the reason why you normally make such decisions, and what factors you think would most affect your food choice. Further details about the key themes and interview questions will be sent to you prior to the interview date via email. The interview should take approximately 30 minutes to 40 minutes. As part of consenting to taking part in this study, interviews will be recorded, and then transcribed into text by the researcher. The interview recording will not be labelled with any personal information such as your name or your workplace, and it will be listened to by only the researcher and her supervisors for the purposes of transcription. All interview recordings and the transcripts will be destroyed after awarding the PhD degree or the end of this study's duration (i.e., after April 2024).

5. What are the possible disadvantages and risks of taking part?

There are no physical risks involved in taking part in this study, and you can take part either virtually or in person. No personal information will be recorded that could be used to individually identify you and the recorded data and transcripts will be kept in secure, encrypted, password-protected storage media at all times.

6. What are the possible benefits of taking part?

Your perceptions are valuable to this research. Although there are no immediate benefits for those people participating in the project, by sharing your knowledge, you will help in expanding knowledge and empirical understanding in the field of recommender systems and benefit the next generation of food recommender systems to be more intelligent and humanistic.

7. What if something goes wrong?

If you feel that your personal data has not been dealt with correctly as per information provided in this sheet, or wish to raise any other concerns/ or complaint about the research, you can first discuss this with the principal researcher via this email address (mzhao18@sheffield.ac.uk) or her supervisors Dr. Frank Hopfgartner f.hopfgartner@sheffield.ac.uk and Dr. Morgan Harvey m.harvey@sheffield.ac.uk Your complaint will be dealt with respectfully, and we will respond appropriately and as soon as possible. However, if you feel that your complaint has not been dealt with appropriately, then you can email the research supervisor via their email address. In addition, if you wish to complain about any other serious problems that may arise during or following your participation in the research, you can contact the University's 'Registrar and Secretary'. If needed, a post-interview follow-up will be arranged between the researcher and the interviewee to clarify or validate uncertain points in the transcript.

8. Will my taking part in this project be kept confidential?

All the information that researchers will collect from you and through the interviews will be kept strictly confidential, and by default your contributions will be anonymized.

9. What is the legal basis for processing my personal data?

According to data protection legislation, we are required to inform you that the legal basis we are applying in order to process your personal data is that 'processing is necessary for the performance of a task carried out in the public interest' (Article 6(1)(e)) and 'processing is necessary for archiving purposes in the public interest, scientific research purposes or statistical purposes' (9(2)(j)). Further information can be found in the University's Privacy Notice https://www.sheffield.ac.uk/govern/data-protection/privacy/general.

10. What will happen to the data collected, and the results of the research project?

The final dissertation will be available on the White Rose thesis repository[1]. In addition, the researchers are expecting to present the data gathered in this study in various formats; journal publication or conference presentations like ACM CHIIR, ECIR and other regional and international conferences.

11. Who is organising and funding the research?

This study is not being funded by any party. The researcher is self-funded.

12. Who is the Data Controller?

The University of Sheffield will act as the Data Controller for this study. This means that the University is responsible for looking after your information and using it properly.

13. Who has ethically reviewed the project?

This project has been ethically approved through the Information School ethics review procedure.

14. Will I be recorded, and how will the recorded media be used?

---

[1] You can browse the collection of the White Rose Thesis Repository using this link https://etheses.whiterose.ac.uk/ .

The audio recordings of your activities made during this research will be used only for analysis and for illustration in this thesis. No other use will be made of them without your written permission, and only the principal researcher of this project will be allowed to access the original recordings.

15. Who can I contact for further information?

Project team:

Mengyisong Zhao
Information School
Regent Court (IS)
211 Portobello
Sheffield
S1 4DP
16. Email: mzhao18@sheffield.ac.uk

Dr. Frank Hopfgartner
Senior Lecturer in Data Science
Information School
Regent Court (IS)
211 Portobello
Sheffield
S1 4DP
Email: f.hopfgartner@sheffield.ac.uk

Dr. Morgan Harvey
Lecturer in Data Science
Information School
Regent Court (IS)
211 Portobello
Sheffield
S1 4DP
Email: m.harvey@sheffield.ac.uk

Departmental ethical team contact:
ischool_ethics@sheffield.ac.uk

Appendix B

Research Stage 2: Ethics Review Documentation and Approval

Approved: 09/02/2023

Mengyisong Zhao
Registration number: 200296416
Information School [a.k.a iSchool]
Programme: PhD in information studies

Dear Mengyisong

PROJECT TITLE: Recipe rating behaviour under different simulated contextual situations
APPLICATION: Reference Number 050555

On behalf of the University ethics reviewers who reviewed your project, I am pleased to inform you that on 09/02/2023 the above-named project was **approved** on ethics grounds, on the basis that you will adhere to the following documentation that you submitted for ethics review:

- University research ethics application form 050555 (form submission date: 06/02/2023); (expected project end date: 09/05/2023).
- Participant information sheet 1115799 version 2 (06/01/2023).
- Participant information sheet 1116800 version 1 (31/01/2023).
- Participant information sheet 1117038 version 1 (06/02/2023).
- Participant consent form 1115800 version 2 (06/01/2023).
- Participant consent form 1116801 version 1 (31/01/2023).
- Participant consent form 1117040 version 1 (06/02/2023).

If during the course of the project you need to deviate significantly from the above-approved documentation please inform me since written approval will be required.

Your responsibilities in delivering this research project are set out at the end of this letter.

Yours sincerely

Claire Du Puget
Ethics Admin
Information School [a.k.a iSchool]

Please note the following responsibilities of the researcher in delivering the research project:

- The project must abide by the University's Research Ethics Policy: https://www.sheffield.ac.uk/research-services/ethics-integrity/policy
- The project must abide by the University's Good Research & Innovation Practices Policy: https://www.sheffield.ac.uk/polopoly_fs/1.671066!/file/GRIPPolicy.pdf
- The researcher must inform their supervisor (in the case of a student) or Ethics Admin (in the case of a member of staff) of any significant changes to the project or the approved documentation.
- The researcher must comply with the requirements of the law and relevant guidelines relating to security and confidentiality of personal data.
- The researcher is responsible for effectively managing the data collected both during and after the end of the project in line with best practice, and any relevant legislative, regulatory or contractual requirements.

The University Of Sheffield.

Application 050555

Section A: Applicant details

Date application started:
Wed 9 November 2022 at 16:32

First name:
Mengyisong

Last name:
Zhao

Email:
mzhao18@sheffield.ac.uk

Programme name:
PhD in information studies

Module name:
PhD in information studies
Last updated:
09/02/2023

Department:
Information School [a.k.a iSchool]

Applying as:
Postgraduate research

Research project title:
Recipe rating behaviour under different simulated contextual situations

Has your research project undergone academic review, in accordance with the appropriate process?
Yes

Similar applications:
- not entered -

Section B: Basic information

Supervisor

Name	Email
Morgan Harvey	m.harvey@sheffield.ac.uk

Proposed project duration

Start date (of data collection):
Thu 9 February 2023

Anticipated end date (of project)
Tue 9 May 2023

3: Project code (where applicable)

Project externally funded?
No

Project code
- not entered -

Suitability

Takes place outside UK?
No

Involves NHS?
No

Health and/or social care human-interventional study?
No

ESRC funded?
No

Likely to lead to publication in a peer-reviewed journal?
No

Led by another UK institution?
No

Involves human tissue?
No

Clinical trial or a medical device study?
No

Involves social care services provided by a local authority?
No

Is social care research requiring review via the University Research Ethics Procedure
No

Involves adults who lack the capacity to consent?
No

Involves research on groups that are on the Home Office list of 'Proscribed terrorist groups or organisations?
No

Indicators of risk

Involves potentially vulnerable participants?
No
Involves potentially highly sensitive topics?
No

Section C: Summary of research

1. Aims & Objectives

The aim of this project is to collect user ratings for a collection of recipes under different simulated contextual situations, in order to explore how user rating behaviour and eating preferences change under both static and dynamic combinations of contextual scenarios. In order to simulate various contextual situations, we have prepared short video clips (20 seconds each) and associated textual environmental descriptions, which participants will view and read prior to rating recipes. For example, to simulate a relaxing weekend, the video shows calming images and plays relaxing, ambient music. Participants will then be asked to rate recipes, which will also feature further hints in their design related to the contextual scenario (e.g.,a stylised image of the sun for the summer context). Data collection will be facilitated via the online crowdsourcing platform Prolific. The research objectives are as follows:
1. Statistically identify the most influential (novel) contextual features that might influence people's food choice and nutritional intake behaviour through users recipe ratings.
2. To accurately predict the user's preferred recipe under each influential contextual situation.
3. To balance the user’s nutritional intake with past preference under each influential contextual situation.

The study intends to answer the following research questions:

1. How do contextual factors affect people’s food choice and nutritional intake behaviours?
2. Whether integrated contextual factors would potentially improve the rating prediction results.
3. Which dynamic contextual factor is most influential for the rating prediction model?

2. Methodology

This study will collect quantitative user recipe ratings and analyse differences in rating distribution across six simulated contextual scenarios, which include: summer, winter, happy, sad, busy, and relaxed, to explore the most influential contextual factors impacting food choice and, subsequently, nutritional intake. We will also ask participants to give qualitative feedback on the reasons behind their choice of ratings.

Data Collection:
The survey will be conducted using Qualtrics and will be managed through an online crowdsourcing platform. The survey will include 3 parts: 1) demographic information collection; 2) rating of recipes under different contextual situations; 3) qualitative feedback on reasons behind rating choices. Before participants start to rate recipes, a short video (around 20 seconds) will be shown to better trigger particular concerns or emotions being in a particular contextual condition. Each participant will be allocated to a single simulated contextual scenario (one of six contextual situations mentioned above), which will not change as they rate recipes. The contextual scenario video will only be played once prior to the rating process. Participants will be asked to rate 25 recipes and, after 15 recipes have been rated, a text and image reminder will be displayed, to ensure participants are still rating under the desired simulated contextual situation. Participants can complete the task at their leisure, as there are no time limits imposed. After participants have finished rating recipes, two questions will be asked, to understand the possible reasons why they are giving high or low ratings to the recipes.

Data Analysis:
Statistical analysis will be used to analyse the user rating data, particularly to explore the relationships between recipe ingredients, cooking times, cooking complexity, nutritional intake with various contextual factors. Commonly used machine learning techniques such as XGBoost, Logistic Regression can be used to predict the user rating.

3. Personal Safety

Have you completed your departmental risk assessment procedures, if appropriate?

Not Applicable

Raises personal safety issues?

No

This study poses minimal to no risks to the researcher given that the location of the setting of data collection will take place online only. There will be no restrictions for participants, the survey questions can be easily answered by logging in to their Prolific account.

Section D: About the participants

1. Potential Participants

No restrictions will be applied other than that the participants be 18 years of age or older, and that they are fluent in English. We aim to recruit a diverse group of participants and will use the crowdsourcing platform’s in-built tools to help achieve this goal. We aim to recruit ~300 participants (the precise number will be determined based on power analysis). This figure was derived via power analysis, as there are in total seven contextual scenarios, one control group, and 50 recipes to be rated. Each participant will rate 25 recipes under a single contextual scenario, and we wish to achieve 20 ratings/recipes. Assuming a medium effect size due to contextual scenario (i.e., d=0.5), power analysis suggests that 300 participants will be needed to reveal significant differences, if they exist. Effect size, and therefore participant numbers, will be adjusted based on the results of a pilot study.

2. Recruiting Potential Participants

Participants will be recruited through the crowdsourcing platform, which has a pool of over 130,000 vetted participants. Conducting recruitment entirely through the platform simplifies the process, reduces the scope for human error, and allows us to predefine the desired demographic profile of our participant pool.

2.1. Advertising methods

Will the study be advertised using the volunteer lists for staff or students maintained by IT Services? No

- not entered -

3. Consent

Will informed consent be obtained from the participants? (i.e. the proposed process) Yes

The researchers have prepared a consent form and brief introduction of to be presented to participants prior to their participation. This information will include the purpose of the study, procedures, data protection, voluntary participation, and any potential risks or benefits that may arise as a result of their participation. In addition, the participant will be assured that even after signing the consent form, it is still possible to withdraw from the project at any time. Consent will be obtained via a checkbox in the Qualtrics form, which participants will have to click on before they can proceed.

4. Payment

Will financial/in kind payments be offered to participants? Yes

In order to recruit a large number of participants on the crowdsourcing platform and ensure that they are adequately and fairly compensated for their time, we will pay participants at the rate of the current UK living wage, according to the Living Wage Foundation. As the current UK living wage is £10.90 per hour, and the whole research process is expected to take around 15 minutes, participants will be paid £2.73 for their time. Only participants that complete the survey will be remunerated.

5. Potential Harm to Participants

What is the potential for physical and/or psychological harm/distress to the participants?

One of the contextual scenarios intends to trigger feelings of sadness in participants, which might lead to participants falling into negativity. However, the video content only reminds participants to imagine or recall a moment when they felt sad, not actually make them feel sad in the present. To counter this, participants in the sad contextual scenario group, after they finish the study, will be shown a happy video. The research process carries little to no risk of physical harm to the participants as all the research process will conduct online.

How will this be managed to ensure appropriate protection and well-being of the participants?

All research participants will be assured that their participation in this study is voluntary and free of pressure, and they can withdraw from the study at any time without further consequences. To ensure that the participants have read and understood their rights and responsibilities, they will be asked to provide a digital signature on the consent form. Participants will be encouraged to contact their GP if they experience strong negative emotions. Any concern of the study, participants will be informed to contact with the principal researcher via this email address (mzhao18@sheffield.ac.uk) or her supervisors Dr. Morgan Harvey m.harvey@sheffield.ac.uk, Dr. David Cameron d.s.cameron@sheffield.ac.uk and Prof. Frank Hopfgartner hopfgartner@uni-koblenz.de at any time. In addition to previous measures, the participants will be informed about their right to refuse to answer any sensitive questions that may increase their anxiety/ or distress during the rating process. Although the Prolific ID will be collected during the study, this information cannot be used to identify individuals and is anonymised.

6. Potential harm to others who may be affected by the research activities

Which other people, if any, may be affected by the research activities, beyond the participants and the research team?

As all participants will work independently at their own computers, there is no real risk of harm to others.

What is the potential for harm to these people?

Little to no potential harm to other people.

How will this be managed to ensure appropriate safeguarding of these people?

N/A

7. Reporting of safeguarding concerns or incidents

What arrangements will be in place for participants, and any other people external to the University who are involved in, or affected by, the research, to enable reporting of incidents or concerns?

In the first instance, participants can contact either the lead researchers, Mengyisong Zhao, or one of the researcher’s supervisors (Dr. Morgan Harvey, Dr. David Cameron, or Prof. Frank Hopfgartner). The information sheet contains contact details for all of the above as well as for the departmental ethics team.

Who will be the Designated Safeguarding Contact(s)?

Dr. Morgan Harvey m.harvey@sheffield.ac.uk.

How will reported incidents or concerns be handled and escalated?

By discussing the reported issues internally with my supervisors. And, if needed, the reported issues or concerns can be raised and escalated to the university’s ethics committee.

Section E: About the data

1. Data Processing

Will you be processing (i.e. collecting, recording, storing, or otherwise using) personal data as part of this project? (Personal data is any information relating to an identified or identifiable living person).
No

Please outline how your data will be managed and stored securely, in line with good practice and relevant funder requirements

Data will be only be accessible to the project researchers and all data collected will be anonymised before analysis with any references to individuals being omitted. All data will be held on encrypted and password-protected environments at all times (the Prolific platform initially, then the University Google Drive).

Throughout the project lifecycle, all research data will always be held on encrypted and password-protected environments, the Prolific platform initially, then the University Google Drive, which can only be accessed by the main researcher and her supervisors. When working off-campus, the researcher will make sure to use the University of Sheffield VPN.

Section F: Supporting documentation

Information & Consent

Participant information sheets relevant to project?
Yes

Document 1115799 (Version 2) All versions

Document 1116800 (Version 1) All versions

Document 1117038 (Version 1) All versions
Could you please check this newest version of information sheet, I have already made changes based on your comments

Consent forms relevant to project?
Yes

Document 1115800 (Version 2) All versions

Document 1116801 (Version 1) All versions

Document 1117040 (Version 1) All versions
Could you please check this newest version of consent form, I have already made changes based on your comments

Additional Documentation

Document 1116802 (Version 1) All versions
Update in new versions of documents

Document 1116803 (Version 1) All versions
Detailed mark of changes in new versions of the decuments

External Documentation

- not entered -

Section G: Declaration

Signed by:
Mengyisong Zhao
Date signed:
Mon 6 February 2023 at 19:37

Offical notes

- not entered -

Recipe rating behaviour under different simulated contextual situations

Please fill out this consent form if you are willing to participate in our research study. If you have any questions about the study, please refer to the information sheet provided or contact Mengyisong Zhao via mzhao18@sheffield.ac.uk.

Taking Part in the Project
I have read and understood the project information sheet dated, or the project has been fully explained to me.

I agree to take part in the project. I understand that taking part in the project will include answering the survey questions and agreement to collecting personal ID.

I understand that by choosing to participate as a volunteer in this research, this does not create a legally binding agreement nor is it intended to create an employment relationship with the University of Sheffield.

I understand that my taking part is voluntary and that I can withdraw from the study at any time/before data has been anonymised, analysed or published. I do not have to give any reasons for why I no longer want to take part and there will be no adverse consequences if I choose to withdraw.

How my information will be used during and after the project
I understand my Prolific ID will not be revealed to people outside the project.

I understand and agree that my answer may be used in publications, reports, web pages, and other research outputs. I understand that I will not be named in these outputs.

I understand and agree that other authorised researchers will have access to this data only if they agree to preserve the confidentiality of the information as requested in this form.

I understand and agree that other authorised researchers may use my data in publications, reports, web pages, and other research outputs, only if they agree to preserve the confidentiality of the information as requested in this form.

So that the information you provide can be used legally by the researchers
I agree to assign the copyright I hold in any materials generated as part of this project to The University of Sheffield.

1. Click to agree the consent form

   Mark only one oval.

   - Agree
   - Disagree

Participant Information Sheet

Project Title: Recipe rating behaviour under different simulated contextual situations

You are being invited to take part in a research project. Before you decide whether or not to participate, it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part. Thank you for reading this.

1. What is the purpose of the study?

The purpose of this study is to understand people's eating patterns under different contextual situations through rating recipes. Based on the results of recipe rating from many individuals under different contextual scenarios, we will be able to identify significance differences caused by these contexts and explore how they affect people's eating and recipe rating behaviour. The ultimate goal is to improve the design of context-aware healthy food recommender systems.

2. Why have I been chosen?

You have been invited to take part in this research because you are an adult, are able to read and understand English, and are willing to share with us your eating preferences.

3. Do I have to take part?

Taking part in this research is entirely voluntary. If you decide to take part, you will be asked to give your consent digitally by ticking a checkbox. In addition, you will be assured that, even after signing the consent form, it is still possible to withdraw from the project at any time, and you do not have to give a reason. If you choose to withdraw your participation, then any and all data collected about you and during any studies you have participated in will be deleted. Please note that by choosing to participate in this research, this will not create a legally binding agreement, nor is it intended to create an employment relationship between you and the University of Sheffield.

4. What will happen to me if I take part?

If you agree to participate, you will be asked to confirm your participation by signing a consent form digitally. You will then be invited to answer the survey, which comprises several parts. First you will be asked some basic demographic questions, e.g., age, gender, and home country. You will then be shown a short video (around 20 seconds) to introduce you to and simulate a specific contextual scenario (this could be, for example, a beautiful spring day). Then you will be asked to rate 20 to 30 recipes, keeping in mind the simulated context. You can complete the survey at your leisure, as there are no time limits imposed. Once you have finished rating recipes, two final questions will be asked, to understand the possible reasons why you might have given high or low ratings to the recipes.

5. What are the possible disadvantages and risks of taking part?

There are no physical risks involved in taking part in this study, as you take part in this research online only. No personal information will be recorded that could be used to individually identify you and the recorded data will be kept in secure, encrypted, password-protected storage media at all time.

6. What are the possible benefits of taking part?

You will be remunerated for your time based on the current UK living wage, as defined by the UK living wage foundation. More importantly, your perceptions are very valuable to this research. For those people participating in the project, by sharing your knowledge, you will help in expanding knowledge and empirical understanding in the field of recommender systems and benefit the next generation of food recommender systems to be more intelligent and humanistic.

7. What if something goes wrong?

If you feel that your personal data has not dealt correctly as per information provided in this sheet, or wish to raise any concerns or complaint about the research, you can first discuss this with the principal researcher via this email address (mzhao18@sheffield.ac.uk) or her supervisors Dr. Morgan Harvey m.harvey@sheffield.ac.uk, Dr. David Cameron d.s.cameron@sheffield.ac.uk and Prof. Frank Hopfgartner hopfgartner@uni-koblenz.de. Your complaint will be dealt with respectfully, and we will respond appropriately and as soon as possible. However, if you feel that your complaint has not been dealt with appropriately, then you can email the research supervisor via their email address. In addition, if you wish to complain about any other serious problems that may arise during or following your participation in the research, you can contact the University's Registrar and Secretary.

8. Will my taking part in this project be kept confidential?

All the information that researchers will collect from you and through this project will be kept strictly confidential, and by default your contributions will be anonymized.

9. What is the legal basis for processing my personal data?

According to data protection legislation, we are required to inform you that the legal basis we are applying in order to process your personal data is that 'processing is necessary for the performance of a task carried out in the public interest' (Article 6(1)(e)) and 'processing is necessary for archiving purposes in the public interest, scientific research purposes or statistical purposes' (9(2)(j)). Further information can be found in the University's Privacy Notice https://www.sheffield.ac.uk/govern/data-protection/privacy/general.

10. What will happen to the data collected, and the results of the research project?

The final dissertation will be available on the White Rose thesis repository[1]. In addition, the researchers are expecting to present the data gathered in this study in various formats; journal publication or conference presentations like ACM CHIIR, ECIR and other regional and international conferences.

11. Who is organising and funding the research?

This study is not being funded by any party. The researcher is self-funded.

12. Who is the Data Controller?

The University of Sheffield will act as the Data Controller for this study. This means that the University is responsible for looking after your information and using it properly.

13. Who has ethically reviewed the project?

This project has been ethically approved through the Information School ethics review procedure.

14. Who can I contact for further information?

Project team:
Mengyisong Zhao
Information School
Regent Court (IS)
211 Portobello

[1] You can browse the collection of the White Rose Thesis Repository using this link https://etheses.whiterose.ac.uk/ .

Sheffield
S1 4DP
15. Email: mzhao18@sheffield.ac.uk

Dr. Morgan Harvey
Lecturer in Data Science
Information School
Regent Court (IS)
211 Portobello
Sheffield
S1 4DP
Email: m.harvey@sheffield.ac.uk

Dr. David Cameron
Lecturer in Data Science
Information School
Regent Court (IS)
211 Portobello
Sheffield
S1 4DP
Email: d.s.cameron@sheffield.ac.uk

Prof. Frank Hopfgartner
Professor for Data Science
Head of the Institute for Web Science & Technologies
UNIVERSITÄT KOBLENZ · LANDAU
Germany
Email: hopfgartner@uni-koblenz.de

Departmental ethical team contact:
ischool_ethics@sheffield.ac.uk

Appendix C

First-Stage Research: Pre-Interview Participant Background Questionnaire

Table C.1: *First-Stage Research: Pre-Interview Participant Background Questionnaire*

Questions	**Answer**
Demographic Information	
How old are you?	
How would you describe your gender? (i.e. Male, Female, Other..., Prefer not to answer)	
What is your ethnic origin?	
What is your home country?	
What is your current country of residence	
How long have you been living at the current country of residence? (i.e. Less than 1 year, 1 to 2 years, 2 to 3 years, 3 to 4 years, 4 to 5 years, More than 5 years)	
What is your occupation?	
On a scale from 1 to 5, how physically active do you consider yourself to be? (i.e. 1-Very inactive, 2-inactive, 3-medium level, 4-active, 5-very active	
What are your hobbies, could you please list some below?	
Food preferences and cooking and recipe searching experiences	
Do you have any specific dietary requirement? (i.e. None, Vegetarian, Food allergies, Other...)	
How often do you cook? (i.e. Never, Yearly, Quarterly, Monthly, Weekly, Daily)	
How often do you use cook books? (i.e. Never, Yearly, Quarterly, Monthly, Weekly, Daily)	
How often do you search for recipes online? (i.e. Never, Yearly, Quarterly, Monthly, Weekly, Daily)	
What kind of recipe presentation do you prefer? (i.e. Texts, Textsa and images, Videos)	
Please list the websites or Apps that you frequently visit for online recipes.	

Appendix D

First-Stage Research: Interview Guide

Interview Guide

Warm-up questions:

Can you recall what you ate for your most recent meal? Did you cook it by yourself or did you eat it somewhere else?

May I ask why you made this food choice?

What kind of food have you eaten most often in the past month? Is there any specific reason why you frequently eat this?

When you have to decide what to have for your lunch or dinner, do you decide quickly what you want to eat? Could you please explain your decision process?

Part 1: Personal and social factors

If your home country and current residence country are different then which country's food do you prefer to eat? And Why?

Can you think of any family eating habits when you were a child and if and how these habits impact your present eating habits as an adult?

Can you think of any social factors that may influence your food choice? (if they struggle with this then give them an example.
- For example, when you are on your own or with your family or with friends or with colleagues?)

Can you describe a situation where your mood or emotion affected your food choice?
- For example, do you have a favourite food you eat to cheer you up?
- How does this affect your food choice?
- Or do you feel bad-tempered or irritable if you haven't eaten anything for a while? (Mention stress as factor as well)

Can you think of any other personal reasons or factors that may influence your food choice? How does this impact your choice?

Which personal factors do you think have the strongest impact on your food choice?

Part 2: Nutritional attitude and food-based factors

How important is nutritional information to you when making food choices?
- Are there instances where it troubles you that you cannot find out the nutritional information of the food you are eating?
- Are you able to acquire enough guidance about nutrition information on each food you choose? How?
- Can you think of a situation in which you would reject a recipe or food product that you would otherwise like to eat on the basis of its nutritional profile? Why?

Do you have favoured ingredients or cooking methods that frequently influence your food choice?

What other factors about a food product or recipe would influence your choice to cook/eat it? (e.g., complexity of cooking, preparation time?)

In which situations would you reject a recipe? (For example, could it be that you don't like the ingredients or consider the cooking methods to be unhealthy?)

Part 3: Online recipes searching and cooking behaviours

Where do you search for recipes online? Websites or apps?

In which situations would you search for recipes online?
- Do you first choose recipes you like, and then select based on the corresponding ingredients or the other way around?
- What do you consider to be a stronger factor when choosing a recipe or food to eat: The way it is presented (e.g., in a picture) or how healthy it is? Why?
- How do you come across these recipes? Do you actively search for them or where they recommended to you (e.g., on a social media platform)?

What type of information would you need for an online recipe to contain? (For example, preparation time, cooking time, servings, ingredients, directions, nutritional information, cooking equipment, all of them, other)

Part 4: Simulated scenarios

Seasons (Summer and Winter)

Scenario 1a: Imagine you are at home on a hot summer’s day with the sun beating down. Can you picture yourself in this scenario? You decide to eat something. If you were to order food, what would you choose? If you decided to cook, what kind of dish would you prepare? Is there any specific reason why you would make these decisions?

Scenario 1b: Imagine you are at home on a cold winter’s day, it's snowing outside. Can you picture yourself in this scenario? Then you decide to eat something. If you were to order food, what would you choose? If you decided to cook, what kind of dish would you prepare? Is there any specific reason why you would make these decisions?

Busy weekday and Relaxing weekend

Scenario 2a: “Imagine you have a busy day at work, like after a whole day of meetings, and you still need to work on a big project. Under these circumstances, what dish are you most likely going to eat? If you were to order food, what would you choose? If you decided to cook, what kind of dish would you prepare? Is there any specific reason why you would make these decisions?

Scenario 2b. Imagine you are having a relaxing weekend. What kind of activities would you like to do? What dish are you most likely going to eat? If you were to order food, what would you choose? If you decided to cook, what kind of dish would you prepare? Is there any specific reason why you would make these decisions?

Emotions (Happy and Sad)

Scenario 3a: “Now imagine you have won the prize you have always desired and you feel very happy, with high energy and are in a celebratory mood. If you were to order food, what would you choose? If you decided to cook, what kind of dish would you prepare? Is there any specific reason why you would make these decisions?

Scenario 3b: “Now imagine you have received some bad news and feel sad about it. If you were to order food, what would you choose? If you decided to cook, what kind of dish would you prepare? Is there any specific reason why you would make these decisions?

Physical activities and Casual time

Scenario 4a: “ Imagine you've just completed a fairly exhausting physical activity, for example, running outside, or working out at gym, or doing yoga. If you were to order food, what would you choose? If you decided to cook, what kind of dish would you prepare? Is there any specific reason why you would make these decisions?

Scenario 4b: “ Imagine you've just had a very chilled and relaxing time, you were there listening to music, playing video games or watching your favourite TV show. If you were to order food, what would you choose? If you decided to cook, what kind of dish would you prepare? Is there any specific reason why you would make these decisions?

Appendix E

Second-Stage research: Participants' demographic features characterisation

Table E.1: *Demographic features characterisation*

Demographic data	Elements	Frequency (N)	Percentage (%)
Gender	Male	212	53.4%
	Female	177	44.6%
	Non-binary or gender diverse	5	1.3%
	Others	3	0.7%
Age	18-24	43	10.8%
	25-34	122	30.7%
	35-44	100	25.2%
	45-54	77	19.4%
	55-64	41	10.3%
	65-74	13	3.3%
	Above 75	1	0.3%
Ethnic origin	White	329	82.9%
	Asian or Asian British	34	8.6%
	Black, Black British, Caribbean or African	16	4.0%
	Mixed or Multiple Ethnic Groups	10	2.5%
	Others	3	0.7%
	Prefer not to say	5	1.3%
Physical activities level	1 (very inactive)	15	3.8%
	2 (inactive)	54	13.6%
	3 (Medium)	188	47.4%
	4 (Active)	104	26.2%
	5 (Very active)	36	9.1%
Cooking frequency	Never	8	2.0%
	Yearly	12	3.0%
	Quarterly	23	5.8%
	Monthly	50	12.6%
	Weekly	120	30.2%
	Daily	184	46.3%
Cooking skill level	No experience	5	1.3%
	Beginner	67	16.9%
	Intermediate	305	76.8%
	Expert	20	5.0%
Cook book using frequency	Never	89	22.4%
	Yearly	73	18.4%
	Quarterly	87	21.9%
	Monthly	102	25.7%
	Weekly	40	10.1%
	Daily	6	1.5%
Online recipe searching frequency	Never	14	3.5%
	Yearly	35	8.8%
	Quarterly	71	17.9%
	Monthly	127	32.0%
	Weekly	135	34.0%
	Daily	15	3.8%
Total number of participants		397	

Appendix F

Recipe Database Feature Demonstration

Table F.1: *Recipe database feature demonstration*

Recipe name	Category	Vegan and vegetarian	FSA health level	Fat/100g	Saturates/100g	Sugar/100g	Salt/100g	(WHO) Health level	(WHO)_adj Health leval	(FDA) Health level	(FDA)_adj Health level
Apple Pie by Grandma Ople	Dessert/Snack	Vegetarian	9	55.73	26.40	0.00	0.36	4	3	3	2
Restaurant-Style Buffalo Chicken Wings	Main dish	None	10	339.73	59.54	0.00	1.61	4	4	3	2
Slow Cooker Chicken Taco Soup	Soup	None	11	26.85	14.13	8.48	2.26	4	3	3	1
Broiled Tilapia Parmesan	Main dish	None	6	9.63	3.70	0.00	0.16	4	3	3	3
Slow Cooker Chicken and Dumplings	Main dish	None	7	14.05	4.68	4.68	0.97	4	3	3	2
Chef John's Macaroni and Cheese	Main dish	None	10	42.40	26.50	10.60	1.23	4	3	2	2
Delicious Ham and Potato Soup	Soup	None	6	6.03	3.29	2.19	0.22	4	3	3	3
Homemade Beef Stew	Main dish	None	6	12.09	4.03	1.73	0.25	4	3	3	2
Gourmet Mushroom Risotto	Main dish	Vegetarian	7	4.86	2.00	1.14	0.32	4	2	3	1
Bread Pudding	Dessert/Snack	Vegetarian	8	8.69	3.48	33.03	0.24	4	2	3	2
Guacamole	Salad	Vegetarian	6	11.03	1.50	1.50	0.30	4	5	3	3
Chantal's New York Cheesecake	Dessert/Snack	Vegetarian	10	25.67	14.97	22.81	0.27	4	3	2	1
Baked Teriyaki Chicken	Main dish	None	12	29.56	8.87	50.24	3.79	4	1	3	1
Red Skinned Potato Salad	Salad	None	8	25.85	4.89	1.40	0.37	4	4	3	2
To Die For Blueberry Muffins	Dessert/Snack	Vegetarian	11	22.37	6.99	47.54	0.45	4	2	3	2
Tiramisu Layer Cake	Dessert/Snack	Vegetarian	11	38.23	19.77	44.82	0.41	4	2	3	1
Mom's Zucchini Bread	Dessert/Snack	Vegetarian	10	22.11	3.40	32.31	0.31	4	2	3	2
Vietnamese Fresh Spring Rolls	Main dish	None	6	2.29	0.00	11.45	0.70	4	2	3	3
Stuffed Green Peppers	Main dish	None	10	18.97	10.27	9.48	0.77	4	3	3	1
Perfect Summer Fruit Salad	Salad	Vegetarian	5	0.74	0.00	21.36	0.00	3	3	3	2
Simple BBQ Ribs	Main dish	None	8	6.72	2.44	5.19	1.30	3	2	2	0
Fruit and Yogurt Smoothie	Dessert/Snack	Vegetarian	5	0.84	0.84	18.55	0.04	3	3	3	2
Chef John's Perfect Prime Rib	Main dish	None	8	34.52	15.01	0.00	0.07	4	4	2	3
World's Best Lasagna	Main dish	None	7	8.03	3.82	3.44	0.54	4	2	3	2
Marinated Flank Steak	Main dish	None	10	80.60	17.27	0.00	1.84	4	4	3	3
Grilled Rock Lobster Tails	Main dish	None	8	18.63	3.36	0.00	0.62	3	3	3	1
Seafood Chowder	Main dish	None	5	1.65	0.33	1.98	0.41	4	2	3	3
Old Charleston-Style Shrimp and Grits	Main dish	None	8	11.21	5.10	0.76	0.42	4	3	2	1
Back-of-the-Box Hershey's Chocolate Cake	Dessert/Snack	Vegetarian	10	15.44	6.79	54.34	0.32	4	2	2	1
Curry Stand Chicken Tikka Masala Sauce	Main dish	None	8	12.64	7.15	3.85	0.54	4	4	3	1
Panna Cotta	Dessert/Snack	Vegetarian	10	52.21	32.46	25.40	0.06	4	2	2	1
General Tso's Chicken	Main dish	None	8	12.43	2.01	6.04	0.40	4	2	3	1
Pecan Snack	Dessert/Snack	Vegetarian	11	55.81	5.58	27.91	0.41	4	4	3	3
Homemade Rock Candy	Dessert/Snack	Vegetarian	6	0.00	0.00	109.42	0.03	3	3	3	2
Chocolate Chocolate Chip Cookies	Dessert/Snack	Vegetarian	10	30.18	17.25	43.11	0.27	4	2	3	3
Churros	Dessert/Snack	Vegetarian	6	8.95	1.23	5.79	0.05	4	4	3	1
Sloppy Joes	Main dish	None	8	15.04	5.64	0.00	0.37	4	5	3	3
Szechwan Shrimp	Main dish	None	6	3.12	0.78	3.12	0.39	5	3	3	3
Beef Bulgogi	Main dish	None	7	7.54	1.74	4.64	0.67	4	2	3	2
Instant Pot Salsa Chicken	Main dish	None	5	1.65	0.33	1.65	0.51	5	2	3	2
Grilled Salmon	Main dish	None	8	14.34	2.15	8.61	0.78	4	3	3	2
Buffalo-Style Chicken Pizza	Main dish	None	10	42.98	12.54	0.00	1.94	4	2	3	1
Delicious Egg Salad for Sandwiches	Salad	Vegetarian	7	19.39	3.64	0.61	0.21	4	4	3	3
Quick and Easy Chicken Noodle Soup	Soup	None	7	5.22	1.74	3.48	1.18	4	2	3	2
Burrito Pie	Main dish	None	8	12.08	5.03	0.50	0.44	4	3	3	1
Simple Sweet and Spicy Chicken Wraps	Main dish	None	7	13.86	2.41	4.22	0.48	4	3	3	2
Ultimate Twice-Baked Potatoes	Main dish	None	9	33.83	16.91	3.38	0.61	4	4	3	2
Baked Ziti	Main dish	None	7	8.23	4.28	4.61	0.30	4	3	3	1
Best Jambalaya	Main dish	None	7	6.53	1.96	0.65	0.53	4	3	3	2
Chicken Parmesan	Main dish	None	7	11.48	4.13	0.92	0.39	4	2	3	1
Pan-Fried Asparagus	Main dish	Vegetarian	8	13.74	6.11	1.53	0.40	4	3	3	3
Baked Kale Chips	Dessert/Snack	Vegan	8	85.67	0.00	0.00	5.28	4	5	3	3
Jamie's Sweet and Easy Corn on the Cob	Main dish	Vegan	7	11.76	0.00	82.34	0.16	4	3	3	3
Sarah's Homemade Applesauce	Dessert/Snack	Vegan	6	0.00	0.00	82.12	0.01	3	3	3	2
Simple Turkey Chili	Main dish	None	4	2.27	0.38	0.38	0.17	4	3	3	3
Roasted Garlic Lemon Broccoli	Main dish	Vegan	10	52.15	0.00	52.15	8.53	4	5	3	3
Ginger Veggie Stir-Fry	Main dish	Vegan	7	13.09	2.91	4.36	1.31	4	4	3	2
Spicy Vegan Potato Curry	Main dish	Vegan	7	6.86	4.46	2.06	0.40	4	3	3	2
Traditional Style Vegan Shepherd's Pie	Main dish	Vegan	7	10.43	1.74	3.48	0.57	4	3	3	2
Moroccan Lentil Soup	Soup	Vegan	4	1.09	0.27	1.36	0.09	4	5	3	4
Roasted Sweet Potato Quinoa Salad	Salad	Vegan	7	13.33	1.67	5.56	0.05	4	4	3	3
Dairy-Free Chocolate Pudding	Dessert/Snack	Vegan	6	3.07	0.61	20.23	0.06	4	3	3	2
Roast Beef and Yorkshire Pudding	Main dish	None	6	11.09	4.62	0.92	0.10	4	3	3	2
Scotch Eggs	Main dish	None	7	11.88	3.38	1.08	0.58	3	3	1	2
Bodacious Broccoli Salad	Salad	None	10	40.02	13.34	5.00	0.91	4	4	3	2
Shepherd's Pie	Main dish	None	6	9.24	3.80	3.26	0.16	4	4	3	2
Absolutely Ultimate Potato Soup	Soup	None	8	16.15	6.92	1.15	0.34	4	3	3	1
Best Cream Of Broccoli Soup	Soup	None	6	3.90	2.28	2.28	0.17	4	4	3	2
Healing Cabbage Soup	Soup	None	4	1.98	0.40	2.38	0.17	4	3	3	3
Catherine's Spicy Chicken Soup	Soup	None	5	2.46	0.82	2.46	0.40	3	2	2	3
Jamie's Cranberry Spinach Salad	Salad	Vegetarian	8	20.62	2.58	19.76	0.05	4	4	3	2
Authentic German Potato Salad	Salad	None	6	2.50	0.62	6.87	0.50	4	1	3	2
Zesty Quinoa Salad	Salad	Vegan	6	6.70	0.56	1.12	0.38	4	3	3	3
Spinach and Strawberry Salad	Salad	Vegetarian	8	18.17	2.27	19.30	0.08	4	4	3	2
Super-Delicious Zuppa Toscana	Soup	None	8	14.49	6.59	2.20	1.05	3	2	2	1

Appendix G

Feature Ablation Study for One-Stage Models: Full Evaluation Results

Table G.1: *Feature ablation study for one-stage models: full evaluation results*

	XGBoost			Ridge Regression			MLP			SVR		
	RMSE	**MAE**	**R2**	**RMSE**	**MAE**	**R2**	**RMSE**	**MAE**	**R2**	**RMSE**	**MAE**	**R2**
Baseline group												
UI	1.299	1.109	0.148	1.408	1.203	-0.001	1.4	1.207	0.01	1.407	1.206	0.001
UI+Contextual features												
UI+cs	1.276	1.078	0.178	1.402	1.204	0.008	1.394	1.201	0.018	1.403	1.201	0.005
UI+Contextual features and user feature set												
UI+cs+udi	1.256	1.056	0.203	1.374	1.183	0.047	1.315	1.111	0.127	1.352	1.154	0.077
UI+Contextual features and recipe feature sets												
UI+cs+nibri	1.262	1.054	0.195	1.375	1.18	0.045	1.362	1.163	0.063	1.366	1.171	0.057
UI+cs+nibri+tfidfrn	1.287	1.076	0.163	1.363	1.168	0.061	1.35	1.142	0.08	1.368	1.166	0.055
UI+cs+nibri+tfidfingr	1.281	1.07	0.172	1.364	1.169	0.061	1.349	1.149	0.08	1.37	1.169	0.052
UI+cs+nibri+BERTembcd	1.269	1.06	0.187	1.366	1.172	0.058	1.357	1.15	0.07	1.367	1.17	0.056
UI+cs+nibri+if	1.288	1.085	0.162	1.363	1.169	0.061	1.36	1.158	0.066	1.383	1.179	0.034
UI+cs+nibri+Gloveembcd	1.273	1.062	0.182	1.366	1.172	0.058	1.352	1.146	0.076	1.366	1.166	0.057
UI+cs+nibri+tfidfcmm	1.262	1.054	0.196	1.37	1.175	0.052	1.353	1.149	0.076	1.368	1.17	0.055
UI+cs+nibri+tfidfrn+tfidfingr	1.292	1.091	0.157	1.363	1.169	0.061	1.352	1.143	0.077	1.373	1.169	0.048
UI+cs+nibri+tfidfrn+BERTembcd	1.275	1.068	0.179	1.363	1.169	0.061	1.356	1.149	0.072	1.37	1.168	0.052
UI+cs+nibri+tfidfrn+if	1.283	1.071	0.168	1.363	1.168	0.061	1.357	1.147	0.07	1.368	1.166	0.055
UI+cs+nibri+tfidfingr+BERTembcd	1.275	1.061	0.179	1.363	1.169	0.061	1.354	1.146	0.075	1.372	1.169	0.05
UI+cs+nibri+tfidfingr+if	1.289	1.088	0.161	1.364	1.169	0.061	1.356	1.146	0.071	1.371	1.169	0.051
UI+cs+nibri+BERTembcd+if	1.274	1.064	0.181	1.367	1.173	0.056	1.356	1.149	0.072	1.367	1.17	0.056
UI+cs+nibri+tfidfrn+tfidfingr+BERTembcd	1.289	1.087	0.161	1.363	1.169	0.061	1.347	1.143	0.083	1.378	1.169	0.042
UI+cs+nibri+tfidfrn+tfidfingr+if	1.291	1.091	0.158	1.363	1.169	0.061	1.37	1.149	0.052	1.373	1.169	0.047
UI+cs+nibri+tfidfingr+BERTembcd+if	1.277	1.068	0.176	1.364	1.169	0.061	1.354	1.151	0.074	1.372	1.169	0.05
UI+cs+nibri+BERTembcd+if+tfidfrn	1.277	1.069	0.177	1.363	1.169	0.061	1.357	1.148	0.071	1.37	1.168	0.052

Continued on next page

	XGBoost			Ridge Regression			MLP			SVR		
	RMSE	MAE	R2	RMSE	MAE	R2	RMSE	MAE	R2	RMSE	MAE	R2
UI+cs+nibri+tfidfrn+tfidfingr+BERTembcd+if	1.29	1.089	0.16	1.363	1.169	0.061	1.357	1.147	0.07	1.377	1.169	0.042
UI+Contextual feature with user and recipe feature sets												
UI+cs+udi+nibri	1.228	1.02	0.239	1.347	1.155	0.084	1.293	1.074	0.155	1.291	1.089	0.159
UI+cs+udi+nibri+tfidfrn	1.227	1.02	0.24	1.336	1.142	0.099	1.286	1.086	0.165	1.31	1.101	0.134
UI+cs+udi+nibri+tfidfingr	1.247	1.046	0.215	1.336	1.142	0.099	1.283	1.072	0.169	1.311	1.103	0.133
UI+cs+udi+nibri+BERTembcd	1.229	1.023	0.237	1.338	1.146	0.096	1.286	1.083	0.165	1.301	1.096	0.145
UI+cs+udi+nibri+if	1.223	1.02	0.245	1.349	1.157	0.081	1.295	1.088	0.153	1.294	1.092	0.154
UI+cs+udi+nibri+tfidfrn+tfidfingr	1.234	1.03	0.231	1.336	1.142	0.099	1.361	1.126	0.065	1.319	1.11	0.121
UI+cs+udi+nibri+tfidfrn+BERTembcd	1.234	1.023	0.231	1.336	1.142	0.099	1.289	1.078	0.161	1.314	1.105	0.129
UI+cs+udi+nibri+tfidfrn+if	1.238	1.037	0.226	1.336	1.142	0.099	1.289	1.077	0.161	1.311	1.103	0.132
UI+cs+udi+nibri+tfidfingr+BERTembcd	1.247	1.049	0.214	1.336	1.142	0.099	1.298	1.083	0.149	1.315	1.107	0.126
UI+cs+udi+nibri+tfidfingr+if	1.247	1.036	0.214	1.336	1.142	0.099	1.292	1.08	0.157	1.312	1.105	0.13
UI+cs+udi+nibri+BERTembcd+if	**1.223**	**1.02**	**0.244**	**1.339**	**1.146**	**0.094**	**1.288**	**1.075**	**0.162**	**1.305**	**1.099**	**0.14**
UI+cs+udi+nibri+tfidfrn+tfidfingr+BERTembcd	1.239	1.032	0.224	1.336	1.142	0.099	1.29	1.089	0.159	1.322	1.114	0.118
UI+cs+udi+nibri+tfidfrn+tfidfingr+if	1.241	1.035	0.223	1.336	1.142	0.099	1.283	1.074	0.169	1.32	1.112	0.119
UI+cs+udi+nibri+tfidfingr+BERTembcd+if	1.242	1.042	0.22	1.336	1.142	0.099	1.317	1.108	0.125	1.317	1.108	0.125
UI+cs+udi+nibri+BERTembcd+if+tfidfrn	1.231	1.021	0.235	1.336	1.142	0.099	1.29	1.086	0.159	1.315	1.107	0.127
UI+All feature set												
UI+cs+udi+nibri+tfidfingr+BERTembcd+if	1.243	1.039	0.219	1.336	1.142	0.099	1.284	1.075	0.168	1.323	1.115	0.116
UI+ the best combination feature set without contextual factors												
UI+udi+nibri+BERTembcd+if	1.259	1.057	0.2	1.343	1.149	0.089	1.318	1.121	0.122	1.343	1.136	0.089

Note: Abbreviations of feature groups shows as follows, UI: user_id and recipe_id, cs: contextual scenarios, udi: user demographic information, nibri: nutritional information and basic recipe information, tfidfrn:tfidf recipe name, tfidfingr: tfidf ingredients, BERTembcd: BERT embedding cooking direction, Gloveembck: Glove embedding cooking direction, tfidfcmm: tfidf cooking methods matching, if: image features. UIC: UI+cs.