Investigating Authorship in Classical Arabic Poetry Using Large Language Models

Lama Alqurashi

A thesis presented for the degree of
Doctor of Philosophy

School of Languages, Cultures and Societies
University of Leeds
United Kingdom
August, 2024

Abstract

This study investigates authorship attribution in Arabic poetry using the entire Classic Arabic Poetry corpus for the first time. Authorship attribution in Arabic poetry dates back to the 6th century during the pre-Islamic period when oral recitation was the primary method of preserving and disseminating poems. Limited written documentation, mainly for treaties, resulted in the loss of much pre-Islamic poetry and the misattribution of post-Islamic poems to pre-Islamic poets. While previous studies have qualitatively explored this issue, this research quantitatively addresses it for the first time.

The study collected and augmented data with metadata to ensure accurate temporal separation. To address potential confusion between style and topic, topic modeling experiments identified five prominent topics, revealing patterns in topic distribution across centuries and poetic meters. Random poems from each century were qualitatively analyzed to validate the topic modeling process.

A classification model was applied to delve deeper into authorship attribution. An ensemble model was developed and tested on applicable data, excluding the pre-Islamic era. The model's performance was evaluated based on topic, number of poets, and number of examples. Topic segregation slightly improved performance, with optimal results observed when one poet was included in the opposite class. The best performance occurred with 60 examples on average.

After selecting the most effective parameters, the model achieved accuracies of 0.97 to 1.0 and corresponding F1 scores. Misclassifications mostly occurred at probabilities below 90%, while correct classifications approached 100%. These findings demonstrate the model's robustness and its potential for addressing real cases of misattribution in Arabic poetry.

Acknowledgment

First and foremost, I thank God for the countless blessings in my life, with the completion of my PhD being just one of them. I am grateful for His continuous care and guidance throughout every stage of my life.

I extend my deepest gratitude to my parents, whose unwavering support and encouragement have been the foundation of all my achievements. I am also immensely thankful to my husband, who has stood by my side from the very beginning of this journey, making it easier and constantly motivating me.

A special thank you goes to my mentor, Professor **Mohammed Abdeldaiem**, who guided me toward this field of study and encouraged me to pursue my academic path. I am also grateful to my sister Alia, who served as an invaluable academic advisor to me.

I am deeply appreciative of my supervisor, **Serge**, who provided the best possible support and was always there to assist me. My second supervisor, **Jacob**, who unfortunately left us before the completion of this stage, offered enriching insights and important discussions that I will always cherish. I would also like to thank my supervisor, **Janet**, for her significant contributions to the linguistic aspects of my work, patiently correcting every letter in my thesis.

I would like to extend my heartfelt gratitude to **Sarah Byford**, Associate Dean for Doctoral Studies at KCL, for her unwavering support during the darkest moments of my journey. Her encouragement and understanding provided me with strength and resilience when I needed it most, and her kindness will always be remembered.

I acknowledge the use of data from **Adab** and **AlDiwan** websites, accessed in compliance with their copyright policies. Their extensive collections of Arabic poetry were invaluable to this research.

Lastly, I extend my thanks to everyone who offered their support and assistance, particularly the faculty members at the University of Leeds and Taif University.

Table of Contents

Abbreviations

ABBREVIATION	DEFINITION
AA	Authorship Attribution
BERT	Bidirectional Encoder Representation from Transformers
HTML	HyperText Markup Language
HTTP	HyperText Transfer Protocol
LDA	Latent Dirichlet Allocation
BOW	Bag of Words
NLP	Natural Language Processing
LR	Logistic Regression
SVM	Support Vector Machines
NB	Naive Bayes
NNS	Neural Networks
CNG	Common N-Grams
POS	part-of-speech
TF-IDF	Term Frequency Inverse Document Frequency
RNN	Recurrent Neural Networks
LSTM	Long Short-Term Memory
SVM	Support Vector Machine
KNN	k Nearest Neighbours
NB	Naive Bayes

List of Figures

List of tables

Chapter 1: Introduction

1. Research Problem and Motivation

This research seeks to address questions related to text attribution in classical Arabic poetry, with a particular focus on the complex issues surrounding pre-Islamic texts. The application of deep learning techniques to address questions in Arabic literary studies represents a novel approach, unprecedented in the field. Historically, analyses of literary issues have relied on qualitative methods. Although qualitative analysis can yield insights, it is inherently subjective, leading to a wide range of conflicting opinions. Additionally, qualitative approaches often rely on historical evidence, which can be incomplete or challenging to trace, especially when exploring the origins of specific texts. Conversely, quantitative methods focus exclusively on the text, mitigating the influence of subjective opinions. They derive their outcomes from linguistic and literary analysis, providing results that are intrinsically connected to the language and style of the author. Furthermore, quantitative methods can address multiple issues within a single study, whereas qualitative approaches often require significant effort to gather external information.

This study marks the first instance of an Authorship Attribution (AA) analysis addressing actual cases of attribution problems in Arabic literature, with a comprehensive focus on linguistic, literary, and technical aspects. The outcomes are expected to shed light on longstanding literary questions that have remained unanswered.

Moreover, this study is unique due to its extensive dataset, encompassing the complete corpus of Arabic poetry, and its use of advanced computational methods. Previous studies in Arabic poetry have utilized conventional machine learning techniques like Support Vector Machines (SVM), which are not considered state-of-the-art (Al-Falahi et al., 2019). This research employs neural networks, specifically Large Language Models, and trains a pre-trained model on Arabic poetry. This approach facilitates a deeper understanding of the linguistic characteristics and stylistic variations in classical Arabic poetry, providing a more robust foundation for analysis and attribution, as the model already has a strong grasp of the language irrespective of the attribution task.

2. Objectives of the Study

The primary objective of this study is to construct deep learning models capable of accurately assigning texts to their corresponding authors. To achieve this goal, the following key milestones will be pursued:

- Conduct a thorough review and critical analysis of the relevant literature.
- Design a comprehensive dataset of Arabic poetry.
- Create deep learning models capable of producing reliable results across various AA scenarios.
- Apply the developed techniques to resolve the attribution of disputed poems, utilizing the knowledge gained from earlier phases.
- Evaluate the model's efficiency and effectiveness using a range of statistical metrics.

3. Research questions

1. What are the existing computational methods for addressing the problem of answering AA questions generally and particularly in Arabic poetry?
2. What are the optimal data sources for this research, and what pre-processing steps are necessary to prepare the data for analysis?
3. How does the topic of the poetry influence the performance and accuracy of the AA model?
4. Which BERT model variant is most suitable for addressing AA questions in Arabic poetry, and what are its specific advantages?
5. What methodologies can be employed to evaluate and validate the performance of these computational models to ensure high accuracy and reliability?
6. How can an AA model be effectively trained given the scarcity of texts per author, particularly in the context of pre-Islamic poetry?
7. What is the impact of the quantity of training examples on the performance of the AA model?
8. Does a one-vs-one classification approach yield better performance compared to a one-vs-all approach in the context of this research?
9. How does the performance of the proposed model compare to traditional AA methods, such as Burrow's Delta?

10. Does the computational model effectively answer AA questions specifically in the context of pre-Islamic Arabic poetry?

Structure of the Thesis

The structure of this thesis is as follows

- Chapter 1: Introduction - outlines the research, highlighting its driving factors, research question and objectives.
- Chapter 2: Background - discusses the foundational context of various methods, techniques, and algorithms employed in AA, with a special emphasis on their application in Arabic and offers a comprehensive perspective on Arabic poetry and the authorship issues within it.
- Chapter 3: Literature review - summarizes the body of related work, concentrating on machine learning techniques employed in AA and verification.
- Chapter 4: Methodology - provides an overview of the research methodology and design.
- Chapter 5: Data extraction and preparation - outlines the procedures used for data extraction and cleaning, offering a detailed description of the corpus.
- Chapter 6: Topic modelling - elaborates on the use of a topic modelling technique to create more accurate topical labels for the poems, allowing for better topic control in authorship classification tasks.
- Chapter 7: Ensemble model for authorship classification - explains the development of a model designed to achieve accurate classification based on authorship, contrasting this approach with traditional methods in AA. It concludes by applying the model to address authorship questions in pre-Islamic poetry.
- Chapter 8: Conclusion and future work – the thesis concludes by summarizing the main findings from the research, reflecting on the key outcomes of the work, and highlighting the novel contributions made. It also explores possible directions for future studies.

Chapter 2: Background

1. Introduction

The issue of AA has captivated scholarly attention over extended periods for diverse reasons, as expounded upon in the introductory chapter. The quest to unravel the intricacies of authorial identification has prompted the utilization of a spectrum of strategies and techniques, ranging from qualitative investigations to cutting-edge quantitative methodologies. The nuances and evolution of these approaches will be meticulously examined and elucidated in the forthcoming Literature Review Chapter.

Embarking on our research odyssey, we commenced with an in-depth exploration of Arabic poetry, delving into its rich history and artistic evolution. Numerous scholarly works dedicated to classical poetry have meticulously examined this art form from multifaceted perspectives, offering a comprehensive historical context. In the linguistic segment of our study, we underscore the profound societal significance of poetry in Arab culture, reflecting the considerable attention devoted to it within literary studies. Subsequently, we delve into the diverse genres that characterize Arabic poetry, with a particular focus on the principal genres that have shaped its expressive landscape.

An additional pivotal facet of our exploration delves into the prosodic system of Arabic poetry, a fundamental element that governs its rhythmic structure. In elucidating this intricate system, we provide a nuanced background on the sixteen metres employed in Arabic poetry, each distinguished by unique syllabic distributions.

Moreover, our research incorporates an examination of the computational aspect, an integral facet of our research initiative. Commencing with the utilization of web scraping, we expound upon its pivotal role in the data collection process. Subsequently, our exploration extends to the domain of topic modelling, recognized as a critical component within AA studies. Notably, scholarly investigations have consistently emphasized the importance of mitigating topic confounding, particularly when the primary focus revolves around style or authorship. This segues into a comprehensive dialogue on the AA techniques implemented in this study, featuring contemporary methodologies such as Bidirectional Encoder Representation from Transformers (BERT) models. This examination serves to provide a basic understanding of the computational methodologies applied in data acquisition, topic modelling, and AA within the specific context of our study.

2. Linguistic Section

Poetry is of crucial significance in Arab culture, serving as both a record and a science. Regarded as art and an integral part of life, poetry wielded significant influence, to the extent that it could instigate or conclude conflicts. The Arab people placed high value on maintaining good conduct, particularly in interactions with poets, as they understood that the impact of their actions would resonate throughout the Arabian Peninsula. In a tribal society where reputation is paramount, the birth of a new poet was celebrated akin to gaining a new platform to contribute to society and elevate one's status. The persuasive power of poetry was comparable to today's media influence, captivating the masses who often accepted poetic narratives as truth.

In a tale illustrating the profound impact of poetry, the king of Al-Ḥīrah,Al- Nuʿmān, declared to envoys from various Arab tribes that he would gift his Abaya to the noblest man among them. On the appointed day, all envoys, except a man named Aws Ibn-Suʿdā, readied themselves to meet the king. Aws's absence caught the king's attention, and he was summoned. When questioned, Aws replied, "If the prize is meant for me, I'll receive it whether I come or not. If it's for someone else, it's not fitting for me to embarrass myself by attending." Remarkably, Aws won the prize.

Envy fueled a conspiracy among the others to tarnish his reputation. They urged Al-Ḥuṭay'ah to compose a satirical poem, known as Hījaa', but Al-Ḥuṭay'ah, grateful to Aws, refused. Offered three hundred camels, Bishr Ibn Abī Khāzim accepted the task. When Aws learned of the poem, he apprehended Bishr, and the tribes refused to shield him from Aws's retribution. The poem's egregious offence included disparaging remarks about Aws's mother. Confronted by Aws, Bishr's fate rested on his mother's chosen punishment. Astonishingly, she opted to return the camels and more, asserting that nothing could erase satire better than panegyric. Aws's magnanimity left Bishr so impressed that he spent the rest of his life praising him.

This narrative underscores the potency of poetry in the Arab world, providing insight into the milieu where poetry thrived and how it endured without the aid of writing implements. Poetry permeated everyday life, capturing the interest of a broad audience rather than being confined to a select group of intellectuals or specialists.

Poetry embraced a diverse array of themes, known as *aghrāḍ*, that evolved with the changing times, incorporating new subjects reflective of contemporary lifestyles. A concise exploration of poetry themes is gleaned from recorded insights in the Britannica Encyclopaedia (1993).

> "first, panegyric *madḥ*, the praise of the tribe and its elders, a genre of poetry that was to become the primary mode of poetic expression during the Islamic period; second, praise's opposite—lampoon *hijāʾ*—whereby the poet would be expected to take verbal aim at the community's enemies and impugn their honour (most often at the expense of women); and third, praise of the dead, or elegy *rithāʾ*"

These main purposes of composing poems served as the primary themes, akin to main courses. However, poems often encompassed additional themes akin to side dishes. Notable themes included descriptions of nature and animals, hunting journeys, references to wine, expressions of love, and reflections on wisdom. More infrequently, poets delved into themes like apologia. While poets might compose poems across various themes, they were often recognized for their specific focus and proficiency in expressing themselves within particular thematic domains. This is evident in the analysis of four of the most famous poets of the pre-Islamic era, who were distinguished by their adeptness in specific areas of poetic expression.

وحكى الأصمعي عن ابن أبي طرفة: كفاك من الشعراء أربعة: زهير إذا رغب، والنابغة إذا رهب، والأعشى إذا طرب، وعنترة إذا كلب، وزاد قوم: وجرير إذا غضب

The esteemed position of poetry in Arab culture allowed this oral art form to thrive, establishing a set of rules akin to a constitution adhered to by all poets. These rules govern various aspects, from prosody to the structure of poems. In terms of prosody, Al-Khalīl delineated 16 patterns that Arabs followed in their poetry composition. Understanding these rules requires knowledge of the syllabic nature of Arabic. Most Arabic words have a triconsonantal root, such as the word "kataba" (to write), with the root KTB. To form different words within the same root, syllables are constructed by combining root consonants with vowels, of which there are six (three short and three long). This morphological simplicity and regularity enable the existence of these constrained prosodic rules. In all 16 patterns, there is a consistent flow of consonants and vowels that must be maintained throughout the entire poem. In other words, if the poem is rewritten in an abstract form of consonants and vowels, all lines will appear identical. The abstraction reveals that each line consists of two identical parts *šaṭr*, with each part having three components, known as *Tafʕīlah* (table). The first two forms are identical (\\o\o\o), and the last one is (\\o\o). The first two forms, in their original form, should be \\o\\\o, but the poet chose to make a change, which appears only once in the original form (in red). This metre is called *Wāfir*, and it is entirely acceptable to make this slight change to the metre.

إذا غامرت في شرف مروم	فلا تقنع بما دون النجوم
\\o\\\o \\o\o\o \\o\o	\\o\o\o \\o\o\o \\o\o
فطعم الموت في أمر حقير	كطعم الموت في أمر عظيم
\\o\o\o \\o\o\o \\o\o	\\o\o\o \\o\o\o \\o\o

To explain how the actual words are transformed into this abstract form, we can use the traditional method which has two simple rules based on what follows consonants.

Cluster of consonant and vowels	Abstraction	Example
Consonant + short vowel	\	Bi (the proposition بـ)
1. Consonant + long vowel 2. Consonant + short vowel + consonant (not followed by a vowel)	\o	1. Bā (the first syllable of the word باب) *2.* Bar (the first syllable of the word بَر)

Table 1: Explaining syllabic Arabic system.

These are the potential types of syllables in Arabic, and we can represent each of them as either (\) or (\o). In creating this abstract form, we disregard word boundaries, allowing a word to be split between two forms or two words to be combined into one form, as we focus solely on the flow of consonants and vowels. Al-Khalīl examined Arabic poetry and identified 16 possible combinations of consonants and vowels, as illustrated in the table.

The metre name	Abstract form
Ṭawīl	\\o\o \\o\o\o \\o\o \\o\o\o
Madīd	\o\\o\o \o\\o \o\\o\o
Basīṭ	\o\o\\o \o\\o \o\o\\o \o\\o
Wāfir	\\o\\\o \\o\\\o \o\\o
Kāmil	\\\o\\o \\\o\\o \\\o\\o
Hazaj	\\o\o\o \\o\o\o
Rajaz	\o\o\\o \o\o\\o \o\o\\o
Ramal	\o\\o\o \o\\o\o \o\\o\o
Sarī ʿ	\o\o\\o \o\o\\o \o\\o
Munsariḥ	\o\o\\o \o\o\o\ \o\\\o
Khafīf	\o\\o\o \o\o\\o \o\\o\o
Muḍāri ʿ	\\o\o\ \o\\o\o
Muqtaḍab	\o\\o\ \o\\\o
Mujtath	\o\o\\o \o\\o\o
Mutaqārib	\\o\o \\o\o \\o\o \\o\o
Mutadārak	\\\o \\\o \\\o \\\o

Table 2: Arabic metres as syllabic forms.

Certain rules in Arabic poetry were regarded as traditions that enhanced the quality of a poem if adhered to. The typical structure began with reminiscing about loved ones and pausing at the remains of a camp, a scene often encountered in Arab life due to the nomadic nature of people who frequently sought water sources. This was typically followed by depicting the journey through the desert and describing the poet's path, highlighting desirable attributes like strength and beauty. Following this extensive introduction, the primary aim of the poem emerged, usually centred around expressions of pride or praise.

Nevertheless, these rules started to be disregarded one by one. Initially, during the Abbasi era, poets began to feel constrained by the conventional themes of poetry and sought to explore new subjects influenced by changes in Arab life. They challenged the established structure that commenced with ruins, dismissing it as outdated, and opted to omit this introductory section. However, prosodic rules endured for a longer period and were generally adhered to until the 19th century when the era of Ḥur poetry, characterized by a more liberated form, commenced.

The field of Arabic literature faces significant challenges related to attribution and authentication, particularly evident in the context of pre-Islamic poetry. The absence of a written form during that era has led to extensive investigations aimed at establishing accurate attributions. Various outcomes have emerged from these studies, ranging from the outright rejection of the entire corpus of pre-Islamic poetry as genuinely from that era to doubts surrounding the attribution of specific texts or even portions of texts. Additionally, the existence of multiple variants of a single text, where a particular composition is reported differently by various sources, further complicates the matter.

This attribution problem has been a focal point of scholarly inquiry from the early days of Islam in the 8th century until the present. Notably, a period of heightened interest and research occurred after the contributions of orientalists who approached the matter from a different standpoint, expressing scepticism towards much of the Jāhilī (pre-Islamic) poetry. In response, Arab scholars passionately defended the authenticity of Jāhilī poetry against the claims of orientalists, leading to a polarized debate.

The conflicting perspectives of these two groups have been described as overly emotional, with accusations of bias either against Arabs and the Arabic language or in favour of them. Historical evidence exists to support both positions, with each group selectively employing such evidence to bolster their arguments. Consequently, arriving at a definitive conclusion becomes challenging when relying solely on the conflicting assertions of these two factions.

Adding to this complexity is the oral-formulaic theory of composition, proposed by Milman Parry and Albert Lord (Parry, 1928; Lord, 1960), which challenges the notion of singular authorship in oral traditions. This theory, originally developed through studies of Homeric epics and Serbian guslari traditions, highlights how oral poets relied on fixed phrases and formulas to compose and perform narratives improvisationally. Applied to pre-Islamic Arabic poetry by scholars like James Monroe (1972) and Michael Zwettler (1978), this theory suggests that such poetry represents a composite tradition rather than individual authorship.

3. Computational Section

Web Scraping

Web scraping refers to retrieving information or data from websites by automatically fetching and analysing the HTML code of a web page to extract the required information. It includes several steps:

- Sending HTTP Request: An HTTP or HTTPS request is initiated to the web server to fetch the HTML content of a particular webpage. In our case, we had to retrieve the poetry corpus from AlDiwan website. The data in AlDiwan follows a hierarchical structure, with the main page containing pages for different eras. Each era page, in turn, has pages for various poets, and each poet's page contains links to their respective poems. In our process, we gather the primary texts from the final layer, preserving the structure of all the layers we traverse. This forms the foundation of our dataset.

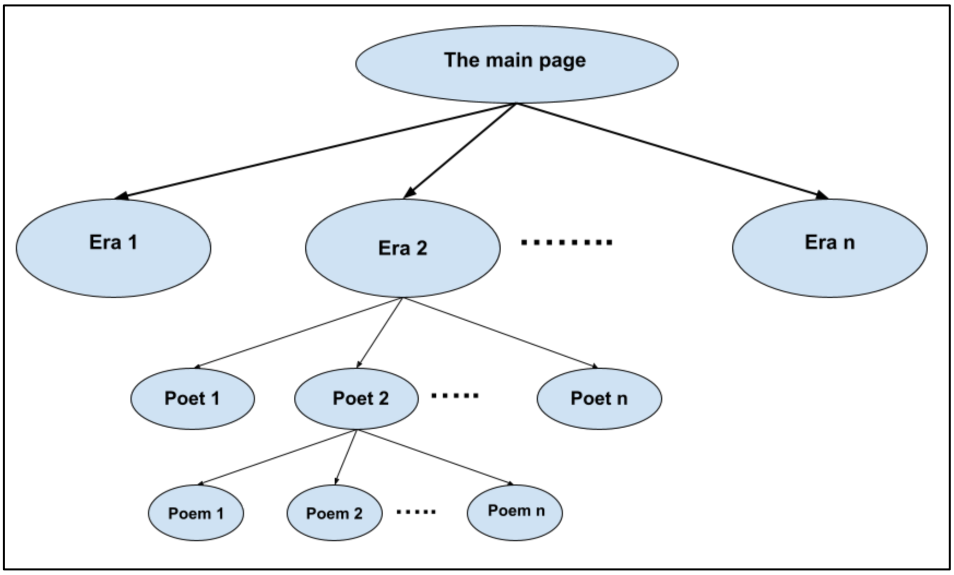

Figure 1: Structure of the website.

- Downloading HTML Content: The server replies to the request by returning the HTML content of the webpage.
- Parsing HTML: The HTML content is analysed to retrieve the necessary information. We utilized the Python library Beautiful Soup for cleaning and extracting the desired text.
- Data Storage: Once all the required data was gathered, we organized the poems based on eras and poets, structuring the information in a CSV file for further analysis.

Topic Modelling

The primary traditional approach to Topic Modelling is Latent Dirichlet Allocation (LDA) (Blei et al., 2003). In LDA, each document is assigned probabilities for different topics, and it also identifies the distribution of vocabulary over topics. This distribution indicates the likelihood of encountering a specific word within each topic. These probabilities are essential for gaining insights into the thematic organization of a set of documents and understanding how words contribute to defining different topics. However, in this model, the word representation is based on the Bag of Words (BoW).

Embedded Topic Modelling (ETM) (Dieng. et al., 2020) is an advanced approach that leverages the advantages of space embedding in topic modelling. In ETM, topics are represented as vectors in the semantic space of words. Unlike traditional models that use topic distribution and vocabulary distribution, ETM employs a topic embedding matrix and a word embedding matrix. The model calculates the probability of a word in a topic by assessing the agreement between the embeddings of the word and the topic. This innovative approach allows for a more nuanced understanding of the relationships between topics and words in a corpus Figure 2.

In the ETM, having a larger dataset tends to improve coherence, which is not the case in LDA. In LDA, there's a limit to performance improvement based on vocabulary size.

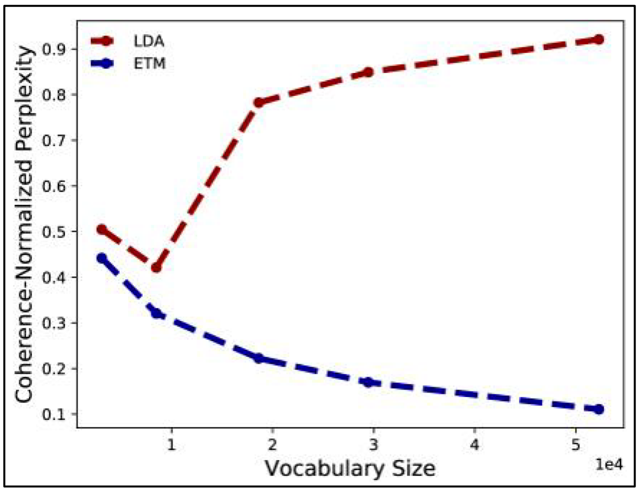

Figure 2: Comparing embedding topic modelling with topic modelling with no embedding (the source: Dieng et al., 2020)

Burrow's Delta

The Delta method (Burrows 1987) utilizes the most frequently occurring words in the training corpus as features to assess authorship. It computes the mean of the absolute differences between the z-scores for a set of word variables within a particular group of texts and the z-scores for the same set of word variables in a target text.

The Delta for the test document is calculated concerning each training document, and the author whose training document exhibits the minimal Delta with the test document is selected for attribution. Burrows's Delta can be seen as a method to rank authorship candidates based on their likelihood. Let X and Y be n-dimensional vectors representing the word frequencies in two documents. The z-score is derived by subtracting the mean and dividing by the standard deviation. The Delta measure between these documents can then be expressed as follows:

$$\sum_{i=1}^{n} |z(x_i) - z(y_i)| = \sum_{i=1}^{n} \left|\frac{x_i - \mu_i}{\sigma_i} - \frac{y_i - \mu_i}{\sigma_i}\right|$$

Machine Learning for classification

In this section, I will present a broad overview of Natural Language Processing (NLP) and the primary methods employed in computational linguistics. I will begin by discussing the pre-processing steps typically undertaken before any analysis, elucidating key terms. Prior to applying machine learning algorithms, raw text undergoes various cleaning and preparation steps. The objective of this phase is to enhance the data quality, directly impacting the performance of the algorithms. In the pre-processing phase, several key steps are implemented to enhance the quality of the data:

- Removing Noise: This involves eliminating irrelevant elements, such as URL links, that do not contribute to the analysis.

- Normalization: This step includes converting all words to lowercase to prevent duplication (e.g., treating "Word" and "word" as the same word without normalization).

- Removing Stopwords: Stopwords, including prepositions, pronouns, and modal verbs, are excluded as they do not carry significant meaning. The specific list of stopwords may vary based on the study's objectives.

- Tokenization: Text is segmented into chunks, either words or sentences. Punctuation marks and whitespaces assist in identifying the boundaries of words and sentences.

- Lemmatization: This advanced step aims to further normalize the data. For instance, it involves converting plural forms to singular forms. Lemmatization typically requires sophisticated tools and is often language-independent.

Once the data undergoes pre-processing, it's prepared for the algorithmic phase. In the realm of machine learning, we encounter primarily two problem types: regression and classification. Disregarding regression, which deals with continuous values like house prices, my focus is on classification problems. In classification, the algorithm works to distinguish between two or more classes. Consider a scenario where we have labelled texts by topics, such as Science and Linguistics, with the goal of classifying them. To simplify, let's say we'll use the word frequencies of "nature" and "language" to create a representation space. Each text is then plotted as a point in this space based on the frequency of these two words. The ML algorithm's task is to discover the optimal boundary (depicted by the red dotted line) that effectively separates the classes with minimal loss. In practical applications, dimensions can extend to thousands, and the representation of these dimensions varies, with raw frequency being the most fundamental.

Conversely, the methodologies in machine learning diverge in their approaches. Traditional machine learning algorithms typically commence by establishing an initial decision boundary. Subsequently, they assess misclassified documents and adjust the boundary line equation accordingly. This iterative updating process repeats until the equation converges to a point where the loss is minimized, ideally approaching zero. Logistic Regression (LR), Support Vector Machines (SVM), and Naive Bayes (NB) are among the most renowned traditional ML algorithms, each presenting unique characteristics while adhering to the fundamental concept of machine learning Figure 3.

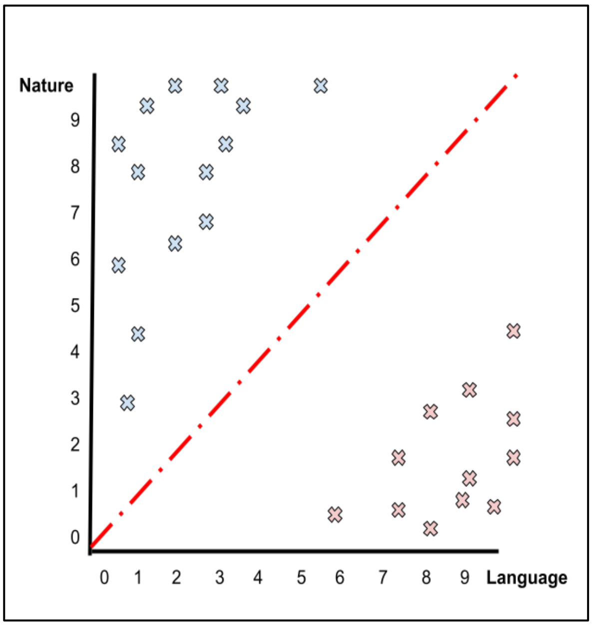

Figure 3: Illustration of a simplified linear classification example.

Deep learning stands out as a revolutionary advancement in the realm of machine learning. Its design mirrors the functioning of neurons in the human brain, with each perceptron serving as a cell endeavouring to comprehend the relationship between input and output. To illustrate this concept simply, let's consider a scenario where only two inputs are processed, and texts are classified into two distinct outputs.

Deep learning constitutes a network with multiple hidden layers, enhancing its capacity to comprehend the intricate relationship between input and output Figure 3. A key advantage of neural networks (NNs) lies in their proficiency in learning non-linear connections. For instance, envision a scenario where data points are plotted, and traditional machine learning fails to accurately classify them due to its limitation in finding linear boundaries Figure 5. This highlights one of the advantages of employing NNs. However, a notable drawback is the requirement for extensive data, which led to the adoption of transfer learning to mitigate this limitation.

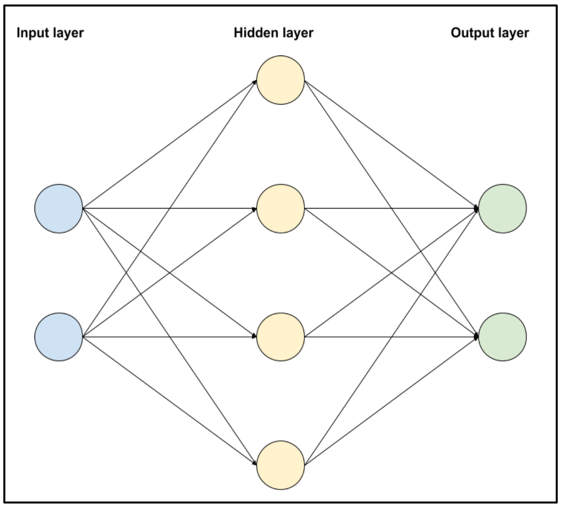

Figure 4: Simple neural network.

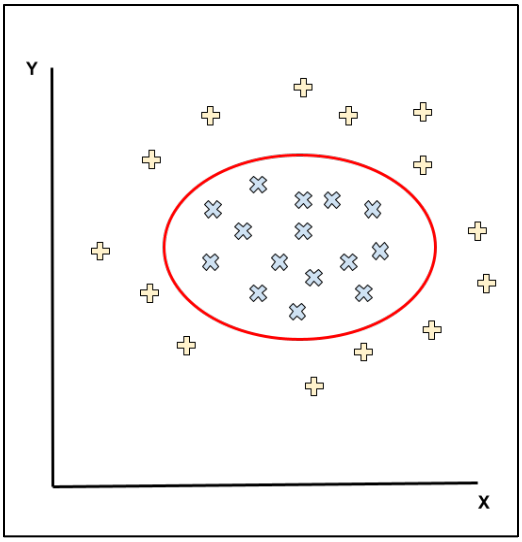

Figure 5: Non-linear classification.

BERT

BERT is a research paper (Devlin, 2018) released by scientists at Google AI Language. Ever since its introduction, BERT has demonstrated significant capabilities in addressing various natural language processing (NLP) challenges, such as classification and question answering.

The fundamental idea behind BERT and transformer-based models is to construct multiple layers of neural networks and train them on an extensive corpus. This process imbues the model with weights or a comprehensive understanding of language without specific task or genre targeting. The acquired knowledge can subsequently be utilized in additional training for diverse tasks.

BERT is a specific implementation of transformers which was introduced first in 2017 by Vaswani et al. in the paper “Attention is All You Need”. The attention mechanism in transformers enables them to effectively capture contextual information, making them suitable for tasks involving comprehension and generation of word sequences, such as machine translation. The strength of transformers lies in their capability to understand long dependencies in language. An example of a linguistic challenge they address is anaphora, where a word or phrase refers back to a preceding one. Traditional neural network models struggle with such cases, but transformers, with their attention mechanisms, can effectively handle situations like the repetition of mentioning a person in the sentences "John went to the store. He bought some groceries."

The transformer architecture adopts an encoder-decoder structure, but it doesn't depend on recurrence or convolutions to produce an output. In essence, the encoder transforms input sequences into continuous representations that are both concise and rich in information. The decoder takes the encoder's output and the decoder's output from the preceding time step to produce a sequence of outputs. The attention mechanism is useful in emphasizing relevant components, regardless of their order or the distance between them. An example of translating a sentence from English to Arabic demonstrates how the model pays attention to differences in word order and word count between the two languages Figure 6.

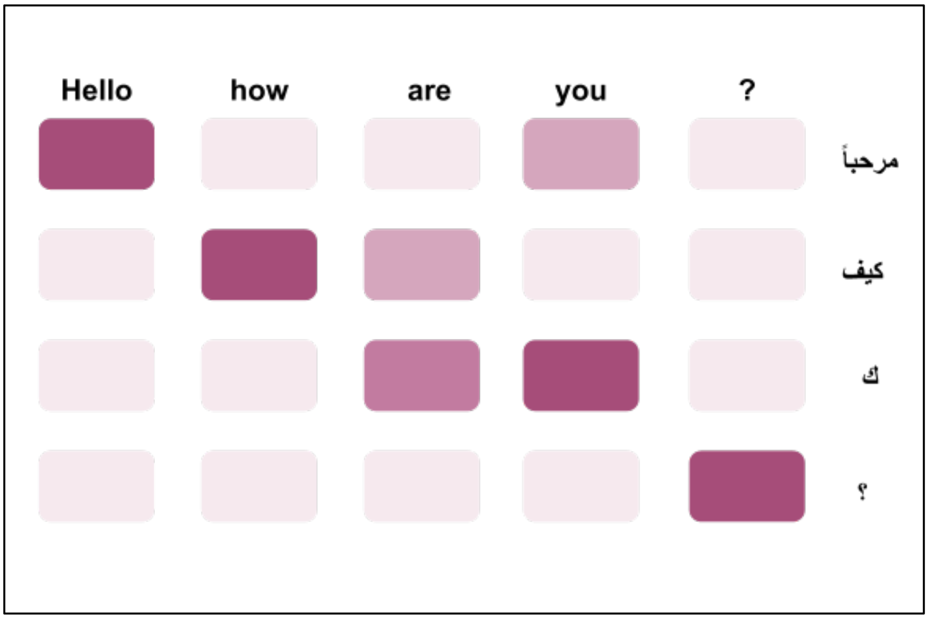

Figure 6: An example of how attention works.

In BERT, we only use the encoder part of the transformer structure, As BERT aims to create a language model. Unlike sequential models that directionally process text input, the Transformer encoder reads the entire sequence of words simultaneously. Although it is often referred to as bidirectional, a more precise description would be non-directional. This feature enables the model to grasp the context of a word by considering all its surroundings, both to the left and right of the word (Devlin, 2018).

Two generic tasks are used to train BERT: Masked words and next-sentence prediction. For masked words, before inputting word sequences into BERT, approximately 15% of the words in each sequence are substituted with a [MASK] token. Subsequently, the model endeavours to forecast the initial values of the masked words by considering the context supplied by the other, non-masked, words in the sequence.

For next-sentence prediction, in the training procedure of BERT, the model is presented with pairs of sentences and is trained to determine whether the second sentence in the pair follows the first sentence in the original document. Throughout the training, 50% of the inputs consist of pairs where the second sentence is the subsequent sentence in the original document, while in the remaining 50%, a randomly selected sentence from the corpus is used as the second sentence. The rationale is that the randomly chosen sentence will lack a cohesive connection with the first sentence.

BERT is versatile in addressing numerous language tasks with the addition of a minimal layer to the main model. During the fine-tuning training, the majority of hyperparametres remain consistent with those used in the original BERT training.

Since the introduction of the BERT model, several versions have been developed to cater to different languages and language variations. Two notable BERT models trained specifically for Arabic are AraBERT (Ara stands for Arabic) and CAMeLBERT (CAMel stands for Computational Approaches to Modeling Language Lab[1]).

AraBERT (Wissam Antoun et al., 2020) comprises six variants, with the initial release of V0.1 trained on a smaller dataset. Subsequent versions include two trained on Twitter data, primarily consisting of colloquial Arabic. The other two variants are more relevant to our research, with corpora primarily focused on Modern Standard Arabic (MSA):

- OSCAR unshuffled and filtered.
- Arabic Wikipedia dump from 2020/09/01
- The 1.5B word Arabic Corpus
- The OSIAN Corpus
- Assafir news articles.

Model	Size (MB/Params)	DataSet (Sentences/Size/nWords)
AraBERTv0.2-base	543MB / 136M	200M / 77GB / 8.6B
AraBERTv0.2-large	1.38G / 371M	200M / 77GB / 8.6B

Table 3: AraBERT variations.

CAMeLBERT (Go Inoue et al., 2021) is a set of BERT models that have undergone pre-training on Arabic texts, encompassing various sizes and variants. They provide pre-

[1] https://nyuad.nyu.edu/en/research/faculty-labs-and-projects/computational-approaches-to-modeling-language-lab.html

trained language models for Modern Standard Arabic (MSA), dialectal Arabic (DA), and Classical Arabic (CA). Additionally, there is a model pre-trained on a combination of these three variants. We only used the CA model as it was trained on a Classical Arabic dataset which can be closer to our data.

Model	Variant	Size	Number of Words
bert-base-arabic-camelbert-ca	CA	6GB	847M

Table 4: CAMeLBERT training information.

4. Conclusion

In this chapter, we offered a broad overview of our research, covering both its aspects: the literary and linguistic, as well as the computational dimension. The linguistic segment highlighted the significance of poetic art in the ancient Arab world, explored various genres of Arabic poetry, and elucidated the prosodic system, providing a detailed explanation of the sixteen metres.

This computational part provided a broad overview of the statistical and machine learning methodologies relevant to addressing the issue of AA. Commencing our research journey, we initiated with web scraping, a technique employed to collect pertinent data for this study. The subsequent discussion delved into topic modelling, elucidating both the conventional approach of Latent Dirichlet Allocation (LDA) and the alternative method adopted in our study, the Embedding Topic Model (ETM). Continuing our exploration, the discussion extended to AA methods, with a focus on differentiating traditional statistical approaches from contemporary state-of-the-art methodologies. Notably, our emphasis turned towards sophisticated techniques within the realm of neural networks, encompassing linear classification models and advanced architectures. The culmination of this comprehensive exploration centres on the spotlighted methodology involving BERT models. This in-depth investigation serves as a foundational understanding of the methodologies employed in data acquisition, topic modelling, and AA within the specific context of our study.

Chapter 3: Literature Review

1. Introduction

This critical review seeks to offer a comprehensive understanding of the historical trajectory and contemporary landscape of AA methodologies. Beginning with an overview of Arabic studies delving into the intricacies of authorship and highlighting distinctive authorial styles, particular attention will be given to the enduring problem of pre-Islamic poetry authorship, known as *Al-Naḥl* and *Al-Intiḥāl*.

The discussion will then transition to computational methods employed to address challenges in AA and verification, with a specific focus on studies specializing in Arabic texts or incorporating Arabic elements into their analyses.

2. Arabic Literary Side

This phenomenon has undergone comprehensive examination from various perspectives, with its earliest mentions dating back to the 9th century in the book *Ṭabaqāt fuḥūl al-shuʿarāʾ* (Ibn Sallām, 2003)

وفي الشعر مصنوع مفتعل موضوع كثير لا خير فيه، ولا حجة في عربيته ولا أدب يستفاد، ولا معنى يستخرج، ولا مثل يضرب، ولا مديح رائع، ولا هجاء مقذع، ولا فخر معجب، ولا نسيب مستطرف، وقد تداوله قوم من كتاب إلى كتاب لم يأخذوه عن أهل البادية، ولم يعرضوه على العلماء.

The passage criticizes a considerable amount of poetry for being fabricated and lacking in value. It states that such works do not demonstrate authentic Arabic, provide no educational benefit, and lack meaningful content. These poems, lacking in various poetic qualities, have been circulated without proper verification from authentic sources or scholars.

Building upon these discussions of authenticity and fabrication in pre-Islamic poetry, the oral-formulaic theory of composition offers a compelling framework for understanding how such works might have evolved. Proposed by Milman Parry and Albert Lord (Parry, 1928; Lord, 1960), this theory challenges the notion of singular authorship in oral traditions. Originally developed through studies of Homeric epics and Serbian guslari traditions, it highlights how oral poets relied on fixed phrases and formulas to compose and perform narratives improvisationally. Applied to pre-Islamic Arabic poetry by scholars like James Monroe (1972) and Michael Zwettler (1978), this theory suggests that such poetry represents a composite tradition rather than individual authorship.

The theory has also faced criticism for its oversimplification of the relationship between oral and written compositions. Critics argue that oral and written forms frequently overlap, as demonstrated in traditions like Nabati poetry, where oral poets often compose with significant deliberation, and written compositions may still rely on oral conventions (Sowayan, 1985). This critique highlights the challenges of applying the theory universally to diverse poetic traditions.

Researchers exploring this phenomenon have employed three distinct methodologies in their investigations:

1. **Historical evidence:**

Studies in this field predominantly leaned on historical evidence to unravel the authenticity of contested texts. Before delving into the intricacies of this approach, it is noteworthy that the majority of these investigations were centred on Jāhilī poetry, given that this era bore the brunt of challenges within the broader landscape of Arabic poetry. Several factors underlie the susceptibility of Jāhilī poetry to such scrutiny, with a key contributing factor being the absence of a written form. The earliest recorded instances of written poetry did not emerge until at least 50 years after the advent of Islam, according to the most optimistic sources. This delayed documentation gave rise to confusion, misattribution, and even deliberate interference. The following list outlines the factors that contributed to this phenomenon:

- **Memory dependence:** Memory mistakes play a pivotal role in the challenges faced by the transmission of poetry through generations. Regardless of the robustness of the memory among poetry transmitters, preserving every word becomes an impossible task. This method of passing down poems from one generation to another has influenced the written version in several ways. Issues include the omission of certain verses, alterations in the poem's sequence, substitution of words, and, significantly, confusion regarding attribution.
- **Rectifying lost records:** Ibn sallām (2003) proposed a connection between the emergence of counterfeit poems and the lost documentation of certain tribes' history, which included the poems chronicling their glory days. In an attempt to remedy this

gap, individuals from these tribes reportedly fabricated poems and attributed them to their ancient poets. This perspective prompts us to view the issue with a measure of scepticism. Additionally, this type of falsification is relatively easier to identify, given that it involves mimicking traditional poems rather than originating from genuine poetic creations.

- **Status pursue:** The growing significance of poetry collectors in the pursuit of ancient poetry collections elevated their social status. This newfound status sparked a desire in some to compose poems or verses whenever queried about a particular topic. A notable anecdote exemplifies this practice: Khalifat once inquired of the trusted scholar Al-Aṣma‘ī about a Jāhilī poem, specifically questioning if any verses were missing from its beginning, as the poem initiated with a sentence alluding to prior discourse. Al-Aṣma‘ī claimed ignorance of any missing portions. Subsequently, Khalifat sought the opinion of the poetry reporter Ḥammād Al-Rāwiyah, who asserted that there were three previously unreported verses. He proceeded to compose these verses spontaneously. When accused of falsehood by Khalifat, Ḥammād Al-Rāwiyah confessed to fabricating them. This incident underscores the influence of the social status associated with poetry reporting, leading to instances of contrived poetic creations.
- The occurrence of copying mistakes is pertinent to the period when written records started, but it took some time before clear and accurate written versions became available.

2. Clues in the text:

In this analytical approach, textual clues play a crucial role in drawing meaningful conclusions. For instance, the inclusion of events the author never witnessed or references to places they never visited raises suspicions about the attribution of the entire text. Any discrepancies between such textual details and the known facts about the author's life can trigger scepticism regarding the authenticity of the text.

3. Stylometric evidence:

Despite its significance, this approach is the least utilized method in qualitative studies. Its limited application stems from the questionable reliability of interpreting language usage, given the considerable variability in opinions about how a specific poet employs language. For instance, a poem ascribed to a brigand poet from the pre-Islamic era faced criticism for one line deemed too intricate for a seemingly straightforward individual. However, conflicting views arose, with some dismissing this criticism and asserting that pre-Islamic poets could indeed achieve such complexity. The inherent subjectivity in evaluating language usage makes this approach less dependable and susceptible to subjective interpretations in qualitative studies.

Apart from historical approaches, AA studies have increasingly turned to statistical and computational methodologies, showcasing their proven reliability. Several studies have employed these innovative approaches in analysing Arabic texts, yielding promising results. It is important to note that these studies typically focused on well-attributed texts for testing, initially avoiding cases with uncertain authorship. The rationale behind this was the assumption that studying texts with doubtful attribution could not be empirically assessed. While this assertion holds some truth, the present study aims to address this gap by developing an evaluation process tailored to disputed texts, forming a substantial part of the research endeavour.

As observed in the preceding section, qualitative studies predominantly rely on extrinsic clues to explore authorship, yet the tools employed often bypass the central question of authorship, which revolves around style. Style, despite its elusive nature, is construed as the distinctive manner in which an author expresses themselves, akin to a fingerprint. It is defined as "the patterning of choices made by a particular author within the resources and limitations of the language and of the literary genre in which he is working." Style imparts uniqueness to a text, establishing a personal connection between the text and its author. It is crucial to recognize that style is context-dependent, as articulated by Leech and Short (2007), existing "in a given context, by a given person, for a given purpose, and so on." Despite variations arising from different topics, descriptions, and purposes, an author's recognizable stylistic touch can permeate a range of texts penned by the same writer.

The examination of style in Arabic has been a focus since the inception of rigorous linguistic studies. Arabic language scholars initially embarked on studying style with the primary aim of illustrating the distinctive features that set the Holy

Quran apart from other writings. Despite its religious underpinnings, their work laid the foundation for an extensive body of research on style, contributing to linguistic analyses and descriptions of what is commonly referred to as style or *Uslūb* in Arabic.

Arabic studies encompass two definitions of style. The first, a broad interpretation, revolves around the manner in which an author approaches writing within a specific topic or genre. Al-Jurjānī (1960), a prominent figure in this field, articulated this perspective:

"واعلم أن الاحتذاء عند الشعراء ، وأهل العلم بالشعر وتقديره وتميزه، أن يبتدئ الشاعر في معنى له وغرض وأسلوب، والأسلوب الضرب من النظم، والطريقة فيه، فيعمد شاعر آخر إلى ذلك الأسلوب فيجيء به في شعره" .
و يقول : " ولما كانت الأغراض الشعرية يوقع في واحد منها الجملة الكبيرة من المعاني والمقاصد وكانت لتلك المعاني جهات فيها توجد ومسائل منها تقتني كجهة وصف المحبوب ، وجهة وصف الطلول وكانت تحصل للنفس بالاستمرار على تلك الجهات، والنقلة من بعضها إلى بعض وبكيفية الإطراء في المعاني صورة وهيئة تسمى الأسلوب".

As illustrated by Al-Jurjānī, it becomes evident that Arabic poetry, particularly in its earlier manifestations, adhered to specific style rules dictated by topics, with authors generally conforming to these guidelines.

The second definition of style pertains to its common usage, representing the individual choices a poet makes in expressing ideas. This aspect is distinct from grammar or prescribed poetry rules, highlighting the author's freedom to make personal stylistic selections. Ibn Khaldūn (2001) offered insights into style, stating:

" فاعلم أنه عبارة عندهم من المنوال الذي ينسج فيه التراكيب أو القالب الذي يفرغ فيه ولا يرجع إلى الكلام باعتبار إفادته أصل المعنى الذي هو وظيفة الإعراب ولاعتبارها إفادته كمال المعنى من خواص التراكيب الذي هو وظيفة البلاغة والبيان ، ولاعتبار الوزن كما استعمله العرب فيه الذي هو وظيفة العروض، فهذه العلوم الثلاثة خارجة عن هذه الصناعات الشعرية، وإنما ترجع إلى صورة ذهنية التراكيب المنتظمة كلية باعتبار انطباقها على تركيب خاص وتلك الصورة ينتزعها الذهن من أعيان التراكيب ويصيرها في الخيال كالقالب أو المنوال"

Al-Bāqillānī (1951) observed that achieving identicality between any two authors is an impossible endeavour:

"فإنه لا يخفى على أحد أن يميز سبك أبي فراس من سبك ابن الرومي، أو سبكه من سبك البحتري، ولا يخفى على أحد أن يميز بين شعر جرير والأخطل، فلكل منه معروف، وطريق"

"It is obvious that anyone can differentiate between the cohesion of Abī-Firās and the cohesion of Ibn-Al-Rūmī, or Abī-Firās's and the cohesion of Al-Buḥturī, and anyone can differentiate between Jarīr's poems and Al-Akhṭal's poems. Every author has his own way".

Nevertheless, there are limited discussions on how one author's style may differ from another. Al-Marzūqī (1991) provided concise insights into the style of Abū-Tammām, stating:

".. إن أبا تمام معروف المذهب فيما يقرضه، مألوف المسلك فيما ينظمه نازع في الإبداع إلى كل غاية، حامل في الاستعارات كل مشقة، متوصل إلى الظفر بمطلوبه من الصنعة أين اعتسف وبماذا عثر، متغلغل في توعير اللفظ وتغميض المعنى أنى تأتى له، وقدر؛ وهو عادل فيما انتخبه في هذا المجموع عن سلوك معاطف ميدانه، ومرتض ما لم يكن فيما يصوغه من أمره وشأنه فقد فليته فلم أجد فيه ما يوافق ذلك الأسلوب إلا اليسير ومعلوم أن طبع كل امرئ - إذا ملك زمام الاختيار- يجذبه إلى ما يستلذه ويهواه ويصرفه عما ينفر منه ولا يرضاه"

Abū-Tammām's method of poetry composition is renowned for its extreme metaphoric richness, characterized by the utilization of rarely used words and intricate nuances of meaning. However, when compiling a collection of poems attributed to other poets, he conscientiously adhered to his own criteria in the selection process. Interestingly, he demonstrated a willingness to accept and include poems that deviated significantly from his distinctive style, showcasing a certain openness to diverse poetic expressions.

While these insights are insightful, they remain general and applicable to a wide array of authors, highlighting a limitation of the qualitative approach. Despite considerable efforts in studying style and extracting linguistic features for specific authors, there has been a scarcity of endeavours to compare and discriminate between two or more authors based on their styles. This study represents a pioneering effort, aiming to address attribution issues by relying on style and textual information, filling a notable gap in the existing literature.

In this context, Abū-Faraj al-Aṣfahānī (1927) recounted a story from the Umayyad era involving the renowned poets Dhū Al-Rummah and Al-Farazdaq. This anecdote vividly illustrates how a poet's style can be unmistakably recognized.

في سياق حديث أبو الفرج الأصفهاني عن قصة هجاء ذي الرمة للمرئي أن جرير لقي ذا الرمة فقال: فأنشدني ما هجوت به المرئي فأنشده قوله:

نبت عيناك عن طللٍ بحزوى عفته الريح وامتضح القطارا

فأطال جداً فقال له جرير: ما صنعت شيئاً أفأرفدك قال: نعم قال: قل:

يعد الناسبون إلى تميم بيوت المجد أربـعةً كـبـارا

يعدون الرباب وآل سعدٍ وعمراً ثم حنظلة الـخـيارا

ويهلك بينها المرئي لغواً كما ألغيت في الدية الحوارا

فغلبه ذو الرمة بها.

قال: حدثني محمد بن عمر الجرجاني قال: حدثني جماعة من أهل العلم أن ذا الرمة مر بالفرزدق فقال له: أنشدني أحدث ما قلت في المرئي فأنشده هذه الأبيات فأطرق الفرزدق ساعة ثم قال: أعد فأعاد فقال: كذبت وايم الله ما هذا لك ولقد قاله أشد لحيين منك وما هذا إلا شعر ابن الأتان.

The tale unfolds as Dhū Al-Rummah, desiring to jest at the expense of the poet Almara'i, crafted a poem for this specific purpose. Upon encountering Jarīr, he was urged to recite his composition. After attentively listening, Jarīr offered insightful feedback, deeming the poem somewhat lacking in achieving its intended effect. To enhance its impact, Jarīr proposed the addition of three lines of his own composition.

In a subsequent encounter with Al-Farazdaq, Dhū al-Rummah was once again prompted to recite his poem. As he reached the three lines suggested by Jarīr, al-Farazdaq requested a repetition. Astonishingly, he pointed out that the verses belonged to someone more seasoned, expressing certainty that they were the creation of Ibn Al-Atān, as Jarīr was colloquially known.

The inclusion of these statements and stories serves to emphasize that specialists are indeed capable of identifying whether a poem or a set of lines was authored by a poet they are familiar with or not. Ibn Sallām (2003) asserted that not all counterfeit poems share the same degree of falseness. Some fraudulent poems, he noted, were crafted by individuals from the same cultural milieu, enabling them to adeptly mimic the style of the original author.

"وليس يشكل على أهل العلم زيادة الرواة ولا ما وضعوا، ولا ما وضع المولدون، وإنما عضل بهم أن يقول الرجل من أهل البادية من ولد الشعراء، أو الرجل ليس من ولدهم، فيشكل ذلك بعض الإشكال"

During the 19th century, certain orientalists delved into the matter of pre-Islamic poetry, introducing fresh perspectives. In the year 1864, the Orientalist Noldeke (Al-Asad, 1988) addressed the subject, pointing out the doubts raised by the appearance of pre-Islamic poetry. Eight years later, the Orientalist Ahlwardt (Al-Asad, 1988) revisited the issue without introducing any substantial innovation. However, he presented it with precision that surpassed his predecessor's treatment. He stated that the transmitted documents lack reliability in terms of authorship, poetic circumstances, or the arrangement of verses. Consequently, it is imperative to subject every work from the sixth and early seventh centuries to thorough scrutiny before acceptance.

In 1916, the orientalist Margoliouth (Al-Asad, 1988)published an inquiry into pre-Islamic poetry in the Royal Asiatic Society Journal. He had previously addressed the context of pre-Islamic poetry in the entry "Muhammad" in *the Encyclopaedia of Religion and Ethics* and in his book "Muhammad and the rise of Islam." Charles Lyall responded to Margoliouth's views in the introduction to his translation of "Al-Mufaḍḍaliyyāt."

However, Margoliouth revisited and published in the aforementioned journal's July 1925 issue, an extensive research paper titled "The Origins of Arabic Poetry." In this thorough inquiry, he elaborated on the resemblances that led him to doubt the authenticity of pre-Islamic poetry. His assertion in this study was that the poetry collected and attributed to the pre-Islamic poets was fabricated and manipulated, originating in the Islamic eras. Those who ascribed it falsely attached it to the poets of the pre-Islamic period, engaging in deceit and slander.

In summary, his article posits that there existed poetry and poets in the Arabian Peninsula before Islam, supported by references in the Quran. He proceeds to analyse the Quran's depiction of poets, noting the substantial number of pre-Islamic poets whose verses are documented in pre-Islamic literature. He expresses scepticism regarding the preservation of pre-Islamic poetry through oral tradition, arguing that Islam did not advocate for such preservation. Simultaneously, he dismisses the notion that pre-Islamic poetry was written down, asserting that if it were, the Arabs would have possessed a written record or records, which the Quran refutes by stating they did not have a book. He contends that the extant pre-Islamic poetry is a subsequent development to the Quran, suggesting that it emerged after Arabs were exposed to the Quran, influencing their poetic expression. Consequently, he posits that pre-Islamic poetry emerged post-Islam, evolving from the Quranic style to a more structured poetic form as Arabs, having heard the Quran, were influenced and began expressing their poetry in their language, transitioning from unconventional to regularized artistic forms.

He stated that the sheer abundance of what we term pre-Islamic literature does not originate from the pre-Islamic era in any sense; rather, it is a post-Islamic manifestation, representing the Islamic way of life, inclinations, and desires more than it reflects the life of the pre-Islamic era. I can scarcely doubt that what remains of authentic pre-Islamic literature is exceedingly scarce. It signifies nothing, reveals nothing, and should not be relied upon for extracting an accurate portrayal of the pre-Islamic period.

In conclusion, the scepticism surrounding pre-Islamic poetry has evolved from questioning the authorship of certain poems to doubting the existence of the entire corpus of pre-Islamic poetry. Although this radical perspective is largely dismissed by the scholarly community, its development highlights the ongoing debates in this field.

4. Computational Side

Methods

Feature engineering:

During the initial stages of machine learning (ML), feature engineering played a crucial role in constructing effective models. The careful selection of features held significant sway over the outcomes, particularly in the field of NLP and, more specifically, in AA studies. In this section, we will explore various methods for selecting features, highlighting their impact on model performance.

Character n-grams, consecutive characters, have been extensively employed in AA studies for various reasons, primarily due to their proven effectiveness (Stamatatos, 2006).

This effectiveness has been demonstrated specifically in Arabic AA as well (Ouamour & Sayoud, 2018). The simplicity of application adds to their appeal, requiring no pre-processing or language-specific information (Stamatatos, 2006). Moreover, character n-grams demonstrate resilience in handling noisy data, such as misspellings commonly found in social media texts. However, an inherent challenge arises from the increase in dimensionality as one word can be represented by multiple n-grams, resulting in features outnumbering the vocabulary size. For instance, the bigram of the word 'book' would yield 'bo,' 'oo,' and 'ok,' tripling the features for a single word.

Given the dimensionality challenge, studies employing character n-grams often resort to feature selection. Some studies apply character n-grams to the entire text without alterations, while others selectively choose certain types of n-grams (Hosseinia & Mukherjee, 2018; Potha & Stamatatos, 2014). Common practices involve concatenating all texts into one body, extracting the top 100k n-grams, and using them as the universal set for all texts (Koppel & Winter, 2014). However, relying solely on a small number of n-grams, such as the most common 100, may yield suboptimal results (Sari et al., 2018).

The success of character n-grams lies in their ability to capture multiple levels of linguistic features, encompassing lexical, contextual, stylistic, and topical aspects (Koppel & Schler, 2004; Sapkota et al., 2015). Moreover, character n-grams exhibit the potential to capture morphological information (Fuchun Peng, 2003), leading to the idea of controlling n-grams to specifically detect morphological properties. Various rules have been established to capture prefixes, suffixes, and other morphological properties in the context of AA studies (Markov et al., 2018; Sapkota et al., 2015).

Several distance measures based on character n-grams, such as Common N-Grams CNG, have been introduced. Initially, CNG considered all n-grams in the two texts under investigation (Fuchun Peng, 2003). Subsequent studies modified this approach by considering only the n-grams existing in both texts (Stamatatos, 2006) or focusing solely on the n-grams in the questioned text under the assumption of its shorter length (Potha & Stamatatos, 2014). Koppel (Koppel et al., 2013) also utilized the original CNG with free-space n-grams. Across these studies, the simplicity of this measure, relying on characters, has consistently demonstrated its efficiency in the context of AA.

A critical consideration for optimizing results with n-grams lies in choosing the appropriate value for 'n'. The optimal 'n' is influenced by the language being analysed. For instance, smaller 'n' is suitable for Chinese, where words typically consist of two characters or fewer. In contrast, German, with its longer words, benefits from larger 'n' (Stamatatos, 2009).

Lexical features, particularly words, have proven to be highly valuable in stylistic studies, showcasing their efficacy since early investigations (Burrows, 1987). However, dealing with tens or hundreds of thousands of features, such as an extensive vocabulary, can lead to inaccuracies as many algorithms struggle with high dimensionality. Critics argue that lexical features, while informative for style, often carry content-related information, making them insufficient for pure style classification (Gamon, 2004). A practical solution is to focus on a select set of influential features, such as the most frequent words.

In the realm of lexical features, the concept of "stopwords" has gained popularity in NLP. Stopwords constitute a list of words deemed meaningless on their own, and it is common practice to exclude them during pre-processing, as their inclusion tends to introduce noise without aiding in text classification. Despite their high frequency, stopwords have paradoxically proven useful for detecting style. Various studies adopt different methodologies for selecting top words, either based on predetermined thresholds or by creating a universe of top words from the corpus under investigation or the language itself (Kumar et al., 2017; Stamatatos, 2009; Badirli et al., 2021; Stamatatos, 2017).

In the context of Arabic, the utility of these features has been demonstrated (Ahmed, 2018). While some studies exclusively focus on using these words, others adopt more sophisticated approaches. For instance, less common words may be abstracted, reflecting only their length, or represented using part-of-speech tags, as seen in some topic classification tasks (Stamatatos, 2017; Sharoff, 2021). Normalizing words related to a specific topic, like personal pronouns, is another practice to enhance the use of these features, particularly as they can serve as indicators of certain themes. In general, removing topic words in AA tasks usually gives better results (Sundararajan & Woodard, 2018).

Notably, the decision on how to leverage these features depends on the specific aims of the study and the characteristics of the corpus. Researchers should carefully consider the most effective approach tailored to the unique objectives of their research.

Mosteller and Wallace (Mosteller, 1964) laid the foundation for AA research by utilizing a group of common words, prompting subsequent attention to the employment of such words in style investigations. The success of these common words lies not only in their frequency but also in their nature as function words, which carry no content. This alignment with style, divorced from content, stems from the idea that a writer's expression remains constant across diverse topics. Consequently, numerous lists of function words have been crafted to capture more stylistic elements and less topical content, yielding satisfactory results in several studies (Abbasi, 2005; Argamon et al., 2003; Argamon et al., 2005; Shlomo Argamon, 2007; Zhao, 2005; Pennebaker, 2010; Koppel et al., 2013), often surpassing character and word n-grams (Sari et al., 2018). Another study (Ramnial, 2016) incorporates words ending in short forms like *'ll* and *n't*.

However, conflicting views exist, with some studies suggesting that character n-grams can outperform approaches based on function words (Kestemont, 2014). Function words can be employed as a unified set or categorized based on their functionalities (Halvani et al., 2020). A comprehensive study (Shaker, 2012) delves into function words in Arabic, presenting an effective list, although various function word variations. Despite the utility of this feature, some studies found it to be ineffective, both in English and Arabic (Abbas et al., 2019; Rocha et al., 2020).

Moreover, it is essential to note that all word-level features necessitate pre-processing, ranging from basic steps like lowercase conversion, noise removal, and text normalization to more intricate procedures involving specific tools such as stemming and lemmatization. The requirements for pre-processing often vary based on the language; for instance, Arabic does not require lowercase conversion but does necessitate word boundary segmentation, a step frequently overlooked. A study comparing original words and stems in Arabic demonstrated that stems yielded superior results (Ahmed, 2018).

Several features can be considered partially syntactic, including character and word n-grams, as they encapsulate certain syntactic properties. Lexical features, like character n-grams emphasizing prefixes and suffixes, are employed in some studies to uncover syntax. More complex syntactic properties are targeted by other studies, with Baayen (Baayen, 1996) pioneering the incorporation of syntax into AA. Baayen's work presented rewrite rules derived from syntactically annotated data, revealing that the frequency of these rules could identify authors as accurately as word frequency. Another study (Gamon, 2004) utilized rewrite rules, offering specific measures such as ratios of adverbial clauses to adjective clauses or subordinate clauses to the rest of the document.

Additional ratio measures assessing ambiguity and syntactic complexity were introduced by Stamatatos (Stamatatos, 2001). Some studies utilize part-of-speech (POS) taggers to discern style (Koppel et al., 2013). Hitschler (Hitschler et al., 2018) employed POS tags to create one-hot encoding vectors for neural networks, while Goncalves (Goncalves & Vimieiro, 2021) used POS frequencies to generate dissimilarity matrices and other matrices based on various features.

Despite the efficiency of POS in capturing non-topical properties, it is seldom used in isolation and is typically combined with other lexical features. One challenge with syntactic features is the required pre-processing, including tagging and annotating, with some unavoidable errors. However, it is possible to extract syntactic features without specialized tools by relying on indicators. For instance, studies (Halvani et al., 2020; Ramezani, 2021) have used sentence starters and endings as indicators of author style.

Using prosodic features for Arabic poetry attribution detection might seem intuitive, but surprisingly, it is not highly indicative (Ahmed, 2017). The stringent prosodic rules in Arabic and the limited choices available to authors—restricted to only 16 metres—contribute to this. This constraint hinders significant variation among authors, not only in poetry but also in prose, where authors tend to unconsciously adhere to distinctive prosodic patterns (Omer, 2017). Recognizing the metre to which a poem belongs is not sufficient; there are also variations within each metre, as elucidated in the work by Berkani et al. (2020). It's plausible that authors express their individualities prosodically through specific variations within the established prosodic norms.

Feature representation

All the discussed features must undergo a transformation into a manageable format for algorithmic calculations. Before applying any algorithms, these features need to be translated into numerical representations, and there are various methods for converting language to numbers. The simplest approach involves converting a document into a vector with word frequencies, where each word is assigned a unique index in a dictionary containing the entire corpus vocabulary. Mosteller (1964) utilized the raw frequencies of a closed set of words, and raw frequencies remain a potent technique, as evidenced by recent studies (Bakly et al., 2020; Goncalves & Vimieiro, 2021). This method is known as Bag of Words (BoW), as it collects all word frequencies in one "bag" regardless of their order.

Another representation of features is achieved by indicating their absence or presence in what is termed one-hot encoding. This approach aids in obtaining normalized values, crucial because raw frequencies can vary depending on text length. In NLP studies, a vital consideration is to focus on influential features, necessitating the removal of repeatedly occurring words, known as "stopwords," in all documents. Alternatively, each word can be assigned a weight that diminishes the importance of stopwords, despite their high frequencies, using the Term Frequency Inverse Document Frequency (TF-IDF) algorithm. This method assigns high weights to words that appear frequently in one class while occurring less in other classes. Notably, stopwords typically receive very low weights as they appear almost equally in all classes. TF-IDF proves most useful in topic classification, where distinguishing topical words between different classes is crucial, and it proved to be effective in AA tasks (Gamon & Michael, 2004; Koppel et al., 2006; Kumar et al., 2017; Ramezani, 2021; Abuhammad, 2021). Other researchers explored Latent Dirichlet Allocation (LDA) and They discovered that this topic modelling approach can compete with the imposter's method while demanding reduced computational resources (Seroussi et al., 2011).

The previously discussed techniques effectively capture word distribution but fall short in considering context. Treating text as a collection of random words overlooks the inherent structure of language (Kilgarriff, 2005). Word embedding emerges as a solution to the drawbacks of BOW approaches by reversing word order. This technique represents text in a co-occurrence matrix, positioning adjacent words in the text closely within the embedding space.

However, word embedding demands a substantial amount of data to yield optimal results. Some NNs methods, such as Recurrent Neural Networks (RNN) and Long Short-Term Memory (LSTM), address the context issue by considering a sequence of words and transmitting information from previous cells. This becomes particularly advantageous when a longer context is crucial.

Bidirectional models further enhance context retention by memorizing words in both directions.Transfer learning stands out as a cutting-edge method that leverages the linguistic knowledge embedded in extensive language models without the need for vast amounts of data for the task under investigation. Attention-based mechanisms in these models facilitate language comprehension. Numerous studies have delved into the specific linguistic knowledge embedded in these models, exemplified by BERT (Rogers, 2020). While analysing such models, and NN models in general poses a challenge, less complex models can still yield reasonable results and are more amenable to analysis. This discussion prompts a more detailed exploration of these techniques.

Algorithms

There is no universal formula applicable to all AA challenges and data types. Consequently, each researcher asserts the superiority of a particular technique, resulting in a multitude of explored methods over time. Evaluating these techniques can be challenging due to their application across diverse corpora. Nevertheless, certain generalizations can be made about what tends to be effective in AA tasks.

In the initial endeavours, researchers employed straightforward algorithms that provided probabilities for potential authors (Mosteller, 1964; Zhao, 2005; Clement, 2003; Madigan, 2005). These models relied on the frequencies of various features in known-author texts, attributing authorship to the one with the most similar distribution of these features. Many researchers adhered to this approach. Distance measures were also prevalent, aiming to find the nearest neighbour among authors and assign the unseen document accordingly. However, the methods for measuring distance varied, with some calculating the actual distance between two vectors, such as Euclidean and Manhattan measures (Ahmed, 2018; Halvani et al., 2020). In contrast, cosine similarity measured the angle between vectors, proving advantageous when raw frequencies were employed (Bakly et al., 2020; Koppel & Winter, 2014; Ramezani, 2021).

A specialized measure for AA, known as CNG, was introduced to assess dissimilarity between two texts (Selj, 2003). Variations of CNG (Potha & Stamatatos, 2014; Stamatatos, 2017) were proposed to make it applicable to unbalanced documents. Additional distance measures, such as min-max (Koppel & Winter, 2014), Hel (Ramezani, 2021), and L1-norm (Kocher, 2017), were utilized. These methods prove particularly useful when there is an insufficient number of examples to train machine learning models. When employing distance measures, it is crucial to establish a threshold to determine when the distance between two documents is short enough to attribute them to the same author.

Recent studies predominantly leverage ML techniques for AA. Some ML techniques share similarities with distance measures, such as k Nearest Neighbours (kNN) (Abbas et al., 2019; Maurya et al., 2016; Ramezani, 2021), while others, like Naive Bayes (NB) and its variations, use probability to make decisions (Koppel et al., 2013; Ramezani, 2021). SVM stands out as one of the most popular models in NLP and AA tasks specifically (Abbas et al., 2019; Al-Harbi et al., 2008; Badirli et al., 2021; Gamon,2004; Markov et al., 2018; Posadas-Durán et al., 2017; Rocha et al., 2020; Stamatatos, 2017). Its capability to handle high-dimensional data proves particularly advantageous when dealing with texts, where the number of dimensions can be as large as the vocabulary size or even more.

Decision Tree (Koppel et al., 2013) and Random Forest (Kumar et al., 2017) are among the techniques used. While Decision Trees may not be as powerful as SVM, they offer the advantage of extracting a graph of the decisions made, aiding in result interpretation. Additional ML algorithms have also been employed, such as MBN (Abbas et al., 2019; Kumar et al., 2017; Markov et al., 2018), Logistic Regression (LR) (Kumar et al., 2017; Posadas-Durán et al., 2017; Rocha et al., 2020; Sari et al., 2018), SMO (Koppel et al., 2013; Maurya et al., 2016), C5.0 (Al-Harbi et al., 2008), RMW (Koppel et al., 2013), and BMR (Koppel et al., 2013). However, these are generally less popular than the previously mentioned techniques.

In recent times, researchers have increasingly turned to NNs to address AA challenges. The versatility of NNs extends beyond AA studies, gaining prominence in the broader field of artificial intelligence. RNNs and their variants, such as Long-Short Term Memories, are widely favoured in this regard (Hosseinia & Mukherjee, 2018). These structures excel at capturing relationships within sequential data, making them particularly suited for language-related tasks. Simpler NN versions like Feedforward Neural Networks (FNN) have also proven effective, surpassing traditional machine learning techniques due to their ability to learn intricate relationships between input and output (Al-Sarem et al., 2020; Benzebouchi et al., 2019; Sari et al., 2018).

Convolutional Neural Networks (CNNs), originally designed for image processing, have found applicability in textual data, especially in handling short texts like tweets (Hitschler et al., 2018; Shrestha et al., 2017). A significant advantage of NNs lies in their ability to eliminate the need for extensive feature engineering. However, harnessing their full potential demands substantial data. Pre-trained models, leveraging knowledge gained from vast datasets, have emerged as valuable tools. BERT is a prominent pre-trained model, achieving state-of-the-art results in various language tasks, including AA (Barlas & Stamatatos, 2020; Tyo et al., 2021; Wang & Iwaihara, 2021). It was found that pre-trained models such as BERT and ELMo deliver optimal performance in scenarios involving cross-topic and cross-genre contexts (Barlas & Stamatatos, 2020). The paper titled "BertAA: BERT Fine-Tuning for AA" (Fabien, M. et al. 2020) presents a method for identifying the author of a given text, which is useful in areas like historical literature analysis, plagiarism detection, and police investigations. The authors introduce BertAA, a model that fine-tunes a pre-trained BERT language model by adding a dense layer and softmax activation to perform authorship classification. BertAA achieves competitive performance, outperforming current state-of-the-art methods by up to 5.3% on datasets such as Enron Email, Blog Authorship, and IMDb. Additionally, the study explores the benefits of incorporating additional features, such as stylometric and hybrid features, into an ensemble approach, which further improves performance by 2.7%.

It is a cutting-edge language model and proved its efficiency in AA studies, therefore the Arabic version of it AraBERT (Antoun, 2020) will be used in this project for the first time in Arabic AA studies. There is one drawback when working with NNs, particularly the more complicated ones like BERT, is that we work with a black box and it is challenging to understand how the model makes the classification decision. Conducting experiments with the shallower layers of BERT to discern the layer that captures the most valuable information for an AA task reveals that, although there might be notable differences in results in specific cases, it remains unclear whether one layer is consistently superior to the others in general (Barlas & Stamatatos, 2020).

Studies on Arabic texts

In an effort to focus solely on poetic-related attributes, Alanazi (2015) examined 100 poems from four authors. The study involved inputting features such as rhythm, metre, the three most frequent letters, the number of verses, and the average number of words and word length per verse into a random forest model. The outcomes revealed varying levels of precision and recall, suggesting that relying exclusively on poetic features may not be sufficient for effective AA.

In a study by (Al-Falahi et al., 2019), the determination of authorship in Arabic poetry is undertaken, assigning authorship to a specified text based on documents with identified authors. The study evaluates the performance impact of Naïve Bayes, Support Vector Machine, and Linear Discriminant Analysis in the context of Arabic poetry AA through text mining classification. Various features, including lexical, character, structural, poetry-related, syntactic, semantic, and specific word features, serve as input data for text mining. The classification algorithms employed are Linear Discriminant Analysis, Support Vector Machine, and Naïve Bayes, implemented by the Arabic Poetry AA Model (APAAM). In the experiment, a set of 114 poets from distinct eras is randomly chosen. The highest accuracy performance value is 99.12%, with attribute-level performance at 98.246% and technique-level performance at 92.836%. They found that poetry features, with an average performance of 81.58%, fall short in author identification due to limitations imposed by common metre, rhyme, and sentence length characteristics shared among poets.

In his doctoral thesis, Shaker (2010) employed a collection of Arabic function words for classifying fourteen classical Arabic books composed by six authors. This approach yielded accuracies ranging from 82.67% to 93.82%, validating the early assumption about the effectiveness of Function Words.

AbdulRazzaq & Mustafa (2014) were the first to assess the applicability of Arabic in AA studies, specifically employing Burrow's-Delta algorithm. They utilized lexical features and discovered that tri-gram words could serve as more effective indicators. Notably, their approach did not involve using specific metrics; instead, they determined authorship based on a predefined threshold to identify the most similar author for a given text.

In their survey on Authorship Identification in Arabic texts, AlQahtani & Dohler (2023) examined 27 studies utilizing AI techniques. The studies differed in terms of data type and size, with eight focusing on Classical Arabic, while others applied models to Modern Standard Arabic alone or in combination with Colloquial Arabic. The linguistic features explored included lexical aspects like n-grams and individual words (2, 4, 5, 7, 8, 9, 10, 11, 12, 14, 15, 16, 18, 20, 21, 22, 24, 25, 26), function words (1, 19), characters (2, 3, 4, 23), syntax (4, 5, 7, 8, 11, 12) , structure (5, 6, 7, 8, 11), poetic features(6), HTML features(8), morphological features(10, 13, 14, 17, 22, 25, 26), content-specific features (5, 11), stylometric features (7, 13, 16, 19 24), and diacritics (22, 25, 26).

1	Shaker, 2010	8	Benjamin et al., 2013	15	Khalil et al., 2020	22	Ahmed, 2018
2	Ouamour and Sayoud, 2012	9	Alwajeeh et al., 2014	16	Al-Sarem et al., 2020	23	Kumar, 2012
3	Ouamour and Sayoud, 2013	10	Baraka et al., 2014	17	Omar and Hamouda, 2020	24	Al-sarem et al., 2018
4	Howedi and Mohd, 2014	11	Otoom et al., 2014	18	Albadarneh et al., 2015	25	Ahmed, 2019a
5	Altheneyan and Menia, 2014	12	Abuhaiba and Eltibi, 2016	19	Altakrori et al., 2018	26	Ahmed, 2019b
6	Alanazi, 2015	13	Al-Ayyoub et al., 2017b	20	Alsager, 2021	27	
7	Abbasi and Chen, 2005	14	Al-Sarem and Emara, 2019	21	Ahmed, 2017		

Table 5: Arabic studies.

This study represents a notable milestone as the inaugural exploration of the Classic Arabic poetry corpus. In the context of AA Arabic studies, our corpus distinguishes itself by its considerable data volume and diverse authorship. Our study's distinction is not solely predicated upon the voluminous nature of the dataset; rather, it is equally underpinned by the distinctive character of the data itself. This corpus, representing Classic Arabic poetry, holds particular significance due to its composition by renowned artists and authors spanning a historical trajectory encompassing 16 centuries. The inclusion of works from such diverse and eminent literary figures adds a nuanced dimension to our dataset, elevating the scholarly value of our exploration. This nuanced amalgamation of substantial data volume and the historical breadth encapsulated by known artists across centuries substantiates the scholarly uniqueness of our study within the domain of AA Arabic studies.

It is noteworthy that this research incorporates the use of a pre-trained model, showcasing a departure from traditional approaches. Additionally, the application of deep learning techniques, while a limited practice within the field, reflects our commitment to embracing innovative methodologies. It is important to acknowledge that, as of now, this study stands as the sole endeavour within Arabic AA research utilizing pre-trained models, underlining its distinctive contribution to the scholarly landscape.

Moreover, our study holds the distinction of being the first among Arabic studies and one of a select few on an international scale that transcends the conventional paradigm of merely constructing a model. Instead, we extend our investigation to practical applications by testing the model on real cases, particularly instances where the authorship is uncertain. In contrast to the predominant qualitative approaches in the existing body of literature, which often address such inquiries through subjective means, our research adopts a computational framework to provide a quantitative perspective on AA.

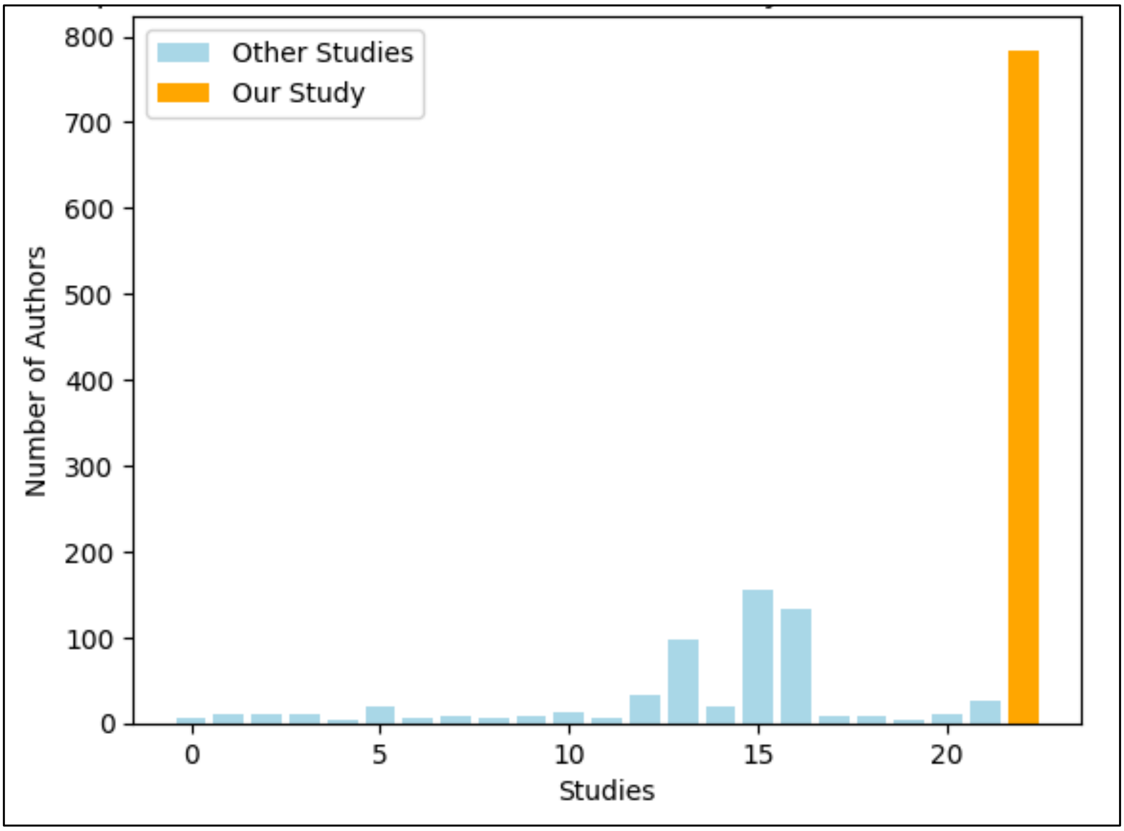

Figure 7: Comparing the authors' number in this study with other Arabic studies.

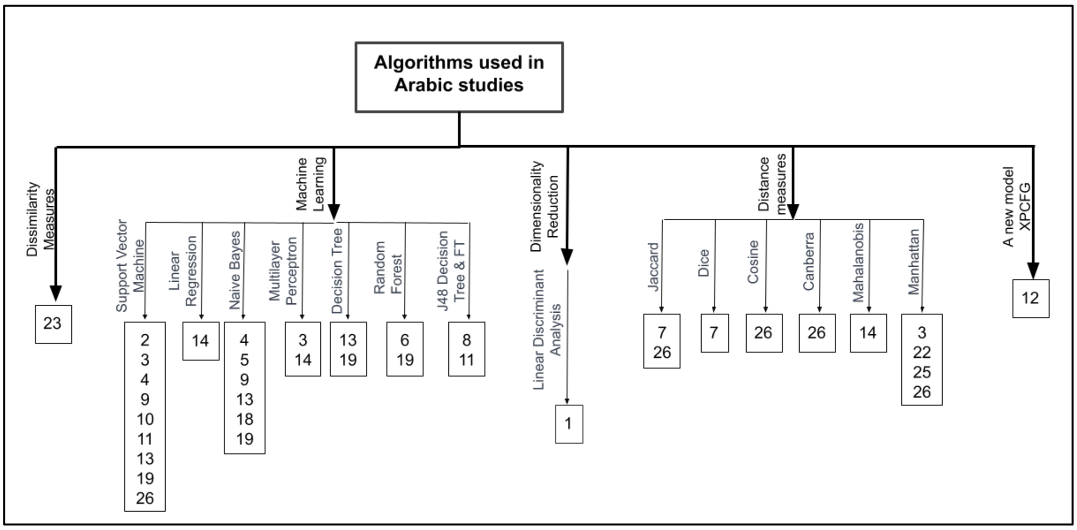

Figure 8: Algorithms that has been used in other Arabic studies. Numbers refer to studies in Table 5

5. Conclusion

In conclusion, this review aimed to explore the historical and modern aspects of AA methodologies. Starting with a survey of Arabic studies, the review delved into the nuances of authorship, emphasizing unique authorial styles. Special consideration was given to the longstanding challenge of pre-Islamic poetry authorship, represented by *Al-Naḥl and Al-Intiḥāl*.

As the narrative shifted, the focus turned to computational methods utilized for addressing AA challenges, particularly in studies dedicated to Arabic texts or those integrating Arabic elements. This comprehensive examination aimed to contribute to the broader understanding of the intricate landscape of AA methodologies and the evolving scholarly efforts in this enduring field.

Chapter 4: Methodology

1. Introduction

The primary objective of this chapter is to elucidate the conceptual framework adopted for tackling the authorship classification problem. Section 4.2 delineates the overarching methodology framework, providing a detailed exposition of each constituent. This encompasses the illustration of data preparation, topic modelling, authorship classification, and the subsequent evaluation process.

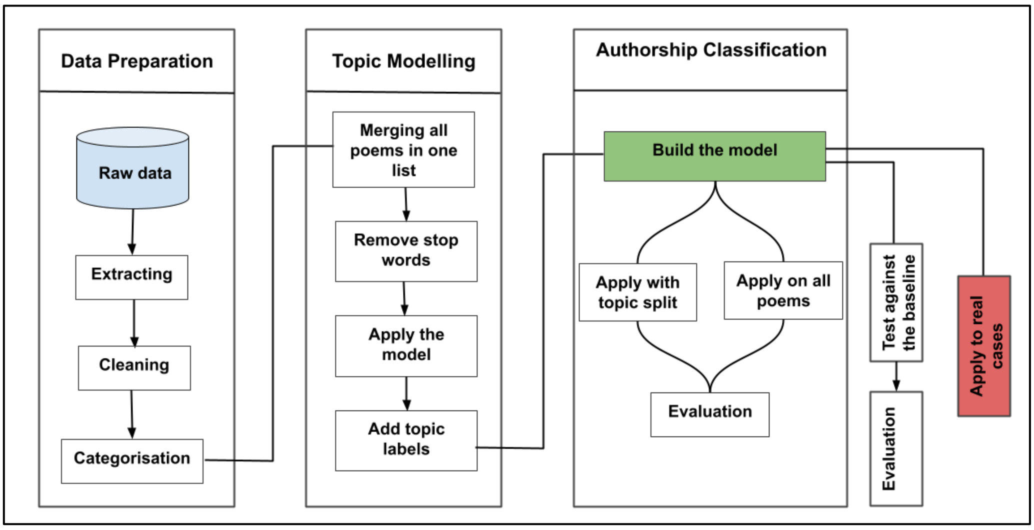

Figure 9: Flowchart of the study methodology.

2. Research methodology framework

In the literature review, we observed a diverse range of methods employed to address AA and Authorship Verification (AV) challenges. These methods spanned from basic statistical approaches to more advanced techniques such as Machine Learning and Deep Learning. Researchers explored an extensive array of linguistic features. However, there is no unanimous agreement on a universally applicable method for all types of AA problems. The trend indicates a shift towards more sophisticated approaches, particularly involving Deep Learning algorithms and, more recently, transfer-learning methods.

In our specific case, despite having ample data, the challenge lies in the distribution of this data across a considerable number of authors. In AA tasks, having more data within each class is advantageous. However, the presence of numerous authors makes the task more challenging.

Furthermore, the primary objective of this research is to assess the suitability of ML methods for addressing a practical AA problem, particularly in the context of pre-Islamic poetry. Therefore, we require a method that considers the unique characteristics of pre-Islamic poetry, including its scarcity.

To tackle this challenge, we suggest employing transfer-learning algorithms, specifically CAMeLBERT. This approach leverages the linguistic knowledge embedded in the pre-trained model, eliminating the need for a large number of examples in each class during training. With the model already possessing language understanding, only minimal fine-tuning is required for a specific task. The provided Figure 9 illustrates our proposed research framework for Authorship Classification, comprising three key components: Data Preparation, Topic Modelling, and Authorship Classification. The subsequent sections will provide a detailed explanation of the objectives and functions of each component.

Data preparation

This element consists of three primary stages: data extraction, data cleaning and categorization.

Data Extraction

In this study, two data sources were utilized: Adab corpus and AlDiwan corpus, both accessible online on their respective websites. The data from Adab is available in Word documents on their website[2], and the extraction process focused on the format and structure of these documents.

Regarding the AlDiwan corpus, the data is exclusively available online, necessitating web scraping to extract information from entire pages. This process involves locating the

[2] https://www.adab.com/

relevant section of the page and eliminating HTML comments. More details about this process in Chapter 5.

Data Cleaning

In this study, the models used are sensitive to diacritics, punctuation, and other noise factors such as double spacing and elongation. To ensure accurate processing in subsequent steps, we prioritize cleaning the texts before constructing our data frame. The specifics of this step will be elaborated in Chapter 5.

Data Categorizing

It is customary to categorise Arabic poetry based on the historical periods corresponding to Islamic states. However, grouping poets with several centuries between them may not be chronologically logical. To address this, we reorganized the data by centuries for a more coherent approach. Further details are provided in Chapter 5.

Topic Modelling

In AA tasks, there is a tendency to make the data or the process topic-agnostic. This is because the algorithm is trained to differentiate between classes representing various authors. If authors write in unrelated genres, the model might focus on classifying topics rather than learning about the unique writing style of the authors.

Some datasets already have topic sections, enabling the identification of genres and sub-genres within the dataset. In the case of Arabic poetry, dividing data based on reasonable time windows partially addresses this issue. However, within the same era, there are several topics referred to in Arabic as *Aghrāḍ*. Poems collected from online resources come with topic labels, but these labels may not be accurate. To enhance accuracy, topic modelling is applied in this research. For more detailed information on this aspect, please refer to Chapter 6.

Authorship Classification

Building the Model

In this study, pre-trained models are identified as the optimal choice for the task at hand, aligning with their effectiveness in various applications. Among the Arabic language BERT models available, including multilingual BERT, AraBERT, and CAMeLBERT, multilingual BERT is excluded due to its training on multiple languages, with Arabic data not being as predominant. To make an informed selection between AraBERT and CAMeLBERT, an experimental comparison is conducted on the dataset, ultimately leading to the choice of CAMeLBERT. Further insights into this decision-making process are provided in Chapter 7.

The model is specifically tailored to meet the needs of its subsequent application on pre-Islamic poetry, which constitutes the primary focus of this research. To address the challenge posed by the limited number of examples, ranging from 40 to 80, we construct an ensemble of models utilizing a voting technique. Further details on this ensemble approach are discussed in Chapter 7.

Test against the Baseline

The literature review highlights that methods for solving AA problems vary in their concepts and complexities, and their effectiveness varies widely. Burrow's Delta, established by John F. Burrows in the late 20th century, is a key method in AA. It has been employed since its creation and was among the first methods applied to Arabic texts (AbdulRazzaq and Mustafa, 2014). Our study found that the BERT model outperformed Burrow's Delta, and the specifics are discussed in Chapter 7.

Applying to Real Cases

In this section, our objective is to tackle enduring literary questions that have traditionally been addressed qualitatively, utilizing our computational approach. While ancient literary inquiries have explored the authorship of numerous poems and various authors, our research will concentrate on specific key authorship issues. We will utilise our model to furnish answers and conduct an in-depth analysis, elaborated further in Chapter 7.

Evaluation

Data Partition Strategy for Training and Testing

In our model, we divided the data into two sets: training and testing, with the testing set comprising 25% of the data provided to the model. Additionally, we created a validation dataset of the same size as the training and testing data. After training the model, we utilize the validation data to assess the effectiveness of the model on unseen data.

Evaluation of Effectiveness

1. Binary Class Confusion Matrices

   A binary confusion matrix is a 2x2 table that displays the actual and predicted values from the proposed classifier.

<table>
<tr><td></td><td></td><td colspan="2">Target
(Actual Value)</td></tr>
<tr><td></td><td></td><td>Positive</td><td>Negative</td></tr>
<tr><td rowspan="2">Model
(Predicte Value)</td><td>Positive</td><td>TP</td><td>FP</td></tr>
<tr><td>Negative</td><td>FN</td><td>TN</td></tr>
</table>

- True Positives (TP): The number of correctly predicted positive cases.
- False Positives (FP): The number of incorrectly predicted positive cases.
- False Negatives (FN): The number of incorrectly predicted negative cases.
- True Negatives (TN): The number of correctly predicted negative cases.

2. Average Accuracy

   The model's prediction accuracy can be determined using the following equation derived from its confusion matrix.

$$Accuracy = \frac{True\ Positives + True\ Negatives}{Total\ Instances}$$

3. Precision

   The precision of a prediction is the proportion of accurately predicted positive observations to the overall count of correctly predicted positive observations.

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives}$$

4. Recall

   The recall is calculated as the fraction of correctly predicted positive observations relative to the total number of positive observations in the actual class.

$$Recall = \frac{True\ Positives}{True\ Positives + False\ Negatives}$$

5. F1

   The **F1 score** is a measure used in machine learning and statistical analysis to evaluate the performance of a classification model. It represents the harmonic mean of precision and recall, combining both metrics to provide a single score that balances false positives and false negatives.

   The "1" in F1 comes from the fact that it gives equal weight to precision and recall. The general formula for the F-score is:

$$F\beta = (1 + \beta^2) \,.\, \frac{precision \,.\, recall}{(\beta^2 \,.\, precision + recall)}$$

   Where **β** is a parameter that controls the weight of recall relative to precision.

   For F1, β is set to 1, meaning precision and recall are equally important. If β were set to 0, the formula would ignore recall (making it just a measure of precision), and if β were set to 2, recall would be given more importance than precision.

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Significance Tests

Conducting statistical analysis is essential for evaluating the proposed method's performance and determining its significant superiority over other methods. The Wilcoxon test and the Friedman test will be employed to assess the ensemble's performance compared to Burrows' Delta.

Wilcoxon-Test

Wilcoxon signed-rank tests can be employed for pairwise comparisons. This test computes the disparities in the performance of two classifiers under consideration, assigns ranks based on the computed differences, and compares their positive and negative ranks. To establish a baseline for comparisons, a control algorithm is utilized, and the statistical significance threshold is set at 0.05 (alpha = 0.05).

3.Conclusion

In this chapter, we have presented the overall framework that will guide our research. Each component of the framework, such as data preparation, modelling, and evaluation, has been discussed individually. The following chapter will delve into the details of the data source and the preparation methods employed for this research.

Chapter 5: Data Extraction and Preparation

1. Introduction

In this chapter, we present a comprehensive overview of the data utilized in the experiments conducted for this study. This encompasses an exploration of the data sources and the methodologies employed for acquisition, involving both file extraction and the process of web scraping. We then delve into the steps of pre-processing and cleaning, outlining the steps taken to prepare the data for the subsequent stages of experimentation.

Furthermore, we elucidate the methodology applied to recategorize the data based on centuries, facilitating a structured organization for analysis. The chapter culminates with a detailed description of both corpora, considering factors such as the number of poems, poets, and the overall size in terms of words. Additionally, we provide a comprehensive examination of the distribution of data across eras, shedding light on the chronological representation of poems across different centuries.

2. The data sources

As previously noted, our dataset comprises two types of data. In this section, I will expound upon the sources of these two corpora.

Adab Website

The primary source of our data is the Adab website, from which the entire content was downloaded in Doc files. The Adab corpus encompasses Arabic poetic texts spanning from the pre-Islamic age to contemporary works and is organized in Word documents, totalling 47 files.

To process this data, we employed a script designed to extract both poems and author names from the Word documents. Additionally, an extensive research endeavour has been undertaken to gather the birthdates of all poets. Consequently, poets are being systematically categorised based on their respective centuries, contributing to a structured and chronological organization of the dataset.

AlDiwan Website

Furthermore, another source of our data consists of Arabic poetry spanning all periods, sourced from the AlDiwan website, so its format is HTML. The significance of this data lies in its richness, characterized by a notably larger number of poets and poems. In the pre-Islamic period alone, it encompasses over 200 poets. Notably, this dataset is labelled with information pertaining to topics, metres, genre, and rhyme.

This extensive dataset comprises over 100,000 poems spanning various historical periods. The classification is structured according to the classical approach, delineating the following chronological classes: Pre-Islamic (~350 - ~600 AD), Veteran (~550 - ~650 AD), Islamic (~610 - 662 AD), Umayyad (662 - 750 AD), Abbasid (750 - 1258 AD), Andalusian (711 - 1492 AD), Ayyubi (1171 - ~1250 AD), Mamluk (1250 - 1517 AD), Ottoman (1299 - 1922 AD). The current segmentation lacks precision, featuring varied time windows that do not sufficiently contribute to meaningful analysis.

Processing and cleaning

To extract this dataset, a web-scraping script is crafted to retrieve all the data and restructure it into a dataframe table. The same data cleaning and reorganization process was applied to both the Adab and AlDiwan sources.

The main webpage links to ten additional pages, each representing a specific historical era like pre-Islamic. These pages list poets from that period, each with a biography and all their poems on sub-pages, which were used for web scraping.

On the poem pages, the top shows the directory path (main page → age → poet → poem), while the bottom has details about the poem's topic, genre, metre, and rhyme. To facilitate further analysis, I created a dataframe table with rows for each poet. The table includes the poet's name, biography, and a list of their poems. This approach consolidates all website data into one table for easier access and processing Figure 10.

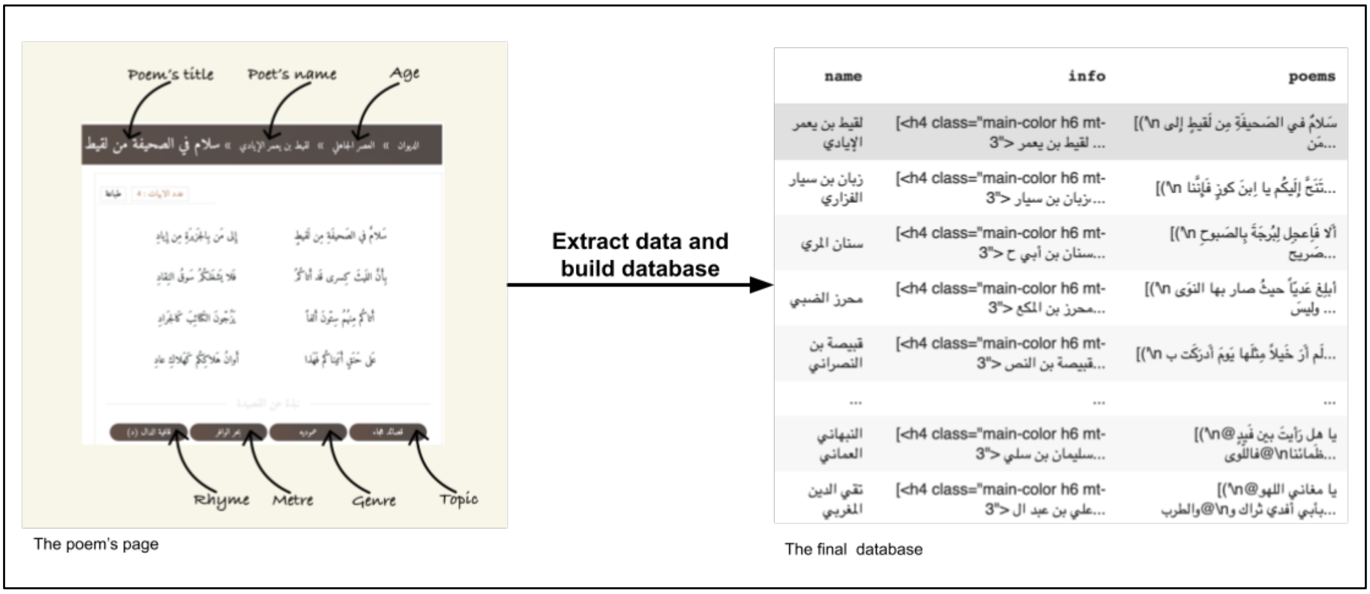

Figure 10: Exporting information from the website to dataframe.

Once all the data is consolidated into a unified table, the next crucial step involves cleaning the texts to eliminate any potential noise that could impact our models. The text may contain HTTP markup segments, line breaks, unnecessary punctuations, and diacritics. The accompanying visual representation illustrates the appearance of the text both before and after the cleaning process Figure 11. This cleaning procedure aims to enhance the quality of the text data, ensuring its suitability for subsequent modelling and analysis.

Figure 11: Data cleaning steps.

As previously highlighted, the website is organized according to the classic method, which categorizes eras based on historical periods. However, this method lacks precision as these eras vary in length and some overlap. In an effort to illustrate the limitations of this approach, I have initially described the data according to these sections. The accompanying pie chart visually depicts the distribution of poets across different eras, providing insight into the drawbacks associated with the classic categorization method.

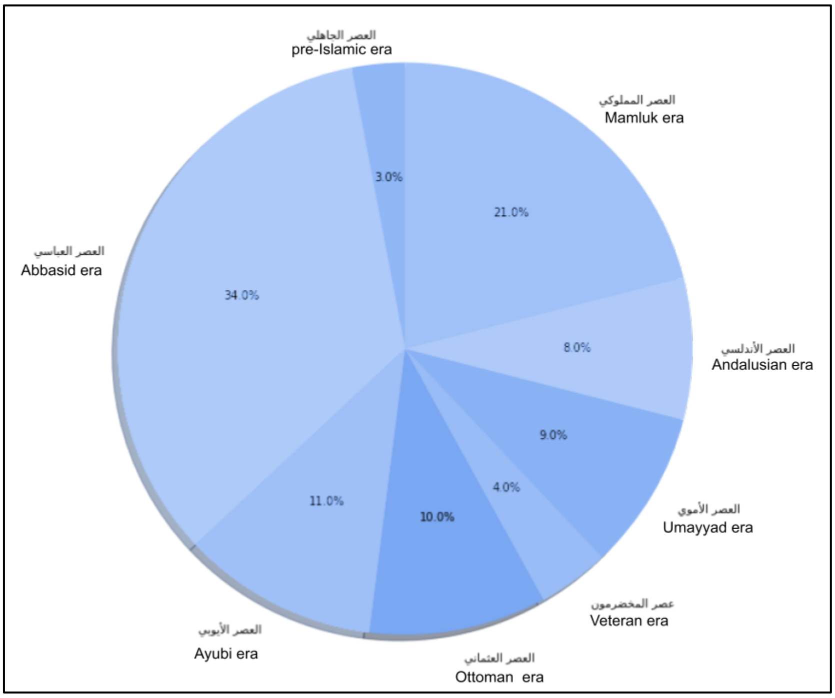

Figure 12: Chronological division of traditional poetry eras.

The data analysis reveals that approximately one-third of poets belong to the Abbasid era, spanning the period from 750 to 1517 AD. Notably, during this timeframe, the Andalusian state was established in 756 in Spain, and some of its poets were contemporaneous with Abbasid poets.

Re-sectioning data based on century division

As demonstrated, the classic division proves to be imprecise and has the potential to introduce noise to our model. To address this limitation, I have opted for a more refined approach by assigning each poet to the century in which they lived. This process entails manual annotation of all poets' dates of birth and death. A comprehensive web search was conducted for each poet, with a focus on obtaining the date of death. While not all poets had information on their date of birth, a significant majority had details regarding their date of death.

The accompanying diagram visually outlines the process of re-sectioning the data based on the poets' birth and death dates, illustrating the meticulous steps taken to enhance the accuracy and granularity of the dataset.

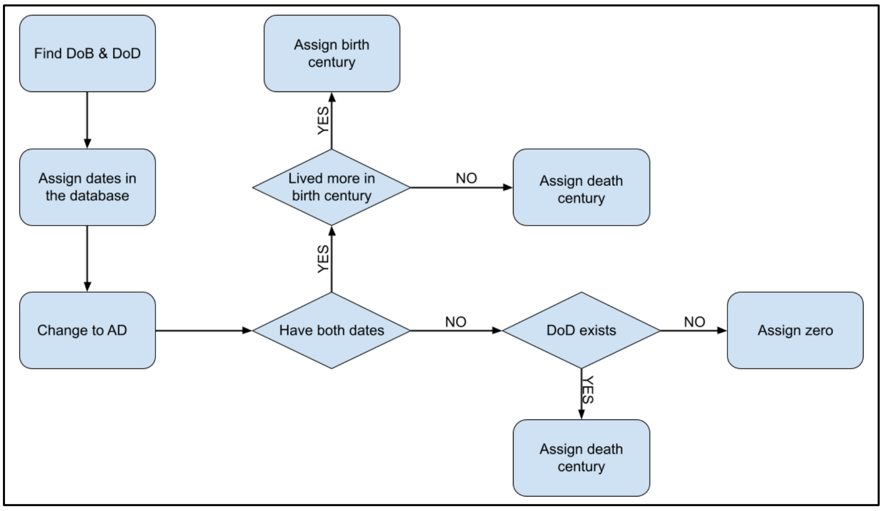

Figure 13: Flowchart illustrating the decision-making process for assigning poets to centuries.

3. Data description

At this juncture, the data stands prepared for a comprehensive description. The first table provides a general overview of the dataset's size, offering insights into the overall composition. The second table delves into the distribution of data per century, presenting a detailed breakdown based on the total number of poets and poems within each century. These tables serve as valuable references for understanding the quantitative aspects of the dataset and its distribution across different historical periods.

Category	Total number
Poets	784
Poems	77,850
Words	6,609,495

Table 6: Total numbers of our corpus.

Centuries	**4th**	**5th**	**6th**	**7th**	**8th**	**9th**	**10th**	**11th**	**12th**	**13th**	**14th**	**16th**	**17th**	**18th**	**19th**	**20th**
poets	2	4	40	118	75	59	40	42	59	71	18	7	14	19	9	2
poems	4	82	1,281	6,958	7,928	10,174	6,492	9,847	8,518	7,610	7,609	824	2,406	3,788	1,223	382

Table 7: Detailed numbers of our corpus.

An alternative representation of the distribution of poets across centuries is depicted in the form of a pie chart, highlighting the proportional distribution of poets in each century. This visual presentation offers a clear and concise view of the percentage distribution, contributing to a more intuitive understanding of the dataset's historical composition.

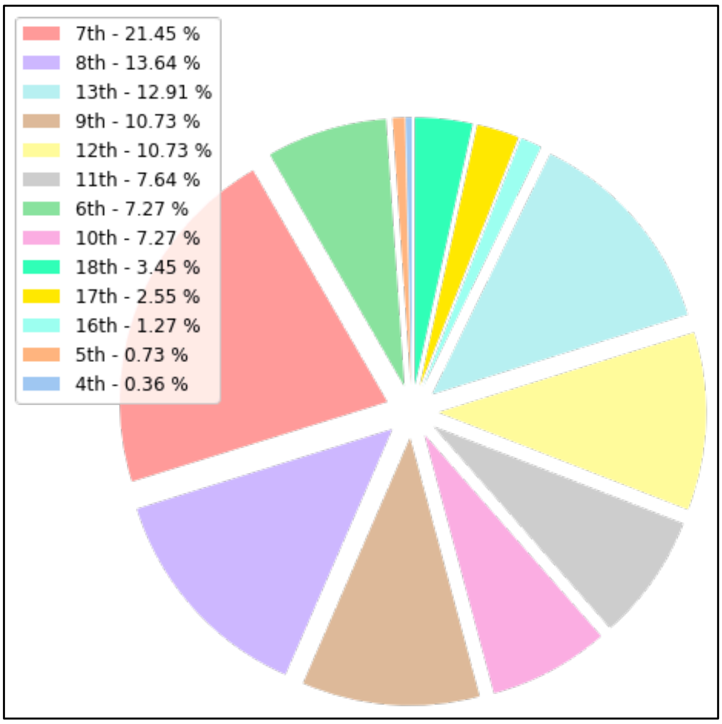

Figure 14: Percentage of poems per century.

Topic labels

As depicted in Figure 10, each poem's page contains various pieces of information, including details about the poem such as its topic. In AA studies, it is imperative to ensure that the distinguishing factor is solely the author's style. Failing to control for topics may introduce the risk of attributing differences to thematic content rather than stylistic nuances. To address this concern, I will initiate an exploration of the topic labelling on AlDiwan website. This exploration aims to provide a comprehensive understanding of the topics assigned to poems, enabling a more nuanced and controlled approach in AA experiments.

The analysis of topic labelling on AlDiwan website reveals notable insights. The pie chart highlights that the two most prominent proportions are 38% for the label (short) and approximately 16% for the label (general). Remarkably, more than 50% of the labels fall into these less informative categories. Furthermore, the remaining labels lack a distinct demarcation, evident in cases such as "romance" and "love." This observation underscores the challenge of relying on topics in AA studies, as a significant portion of the labels lacks specificity, potentially complicating the isolation of stylistic features from thematic content.

In the absence of precise topic labelling, it becomes necessary to establish one. A more accurate topic labelling can offer several significant benefits: it can enhance our understanding of the data, facilitate the examination of correlations between topics and other factors like time periods and poetic metres, and enable an evaluation of the impact of topics on the performance of the authorship model.

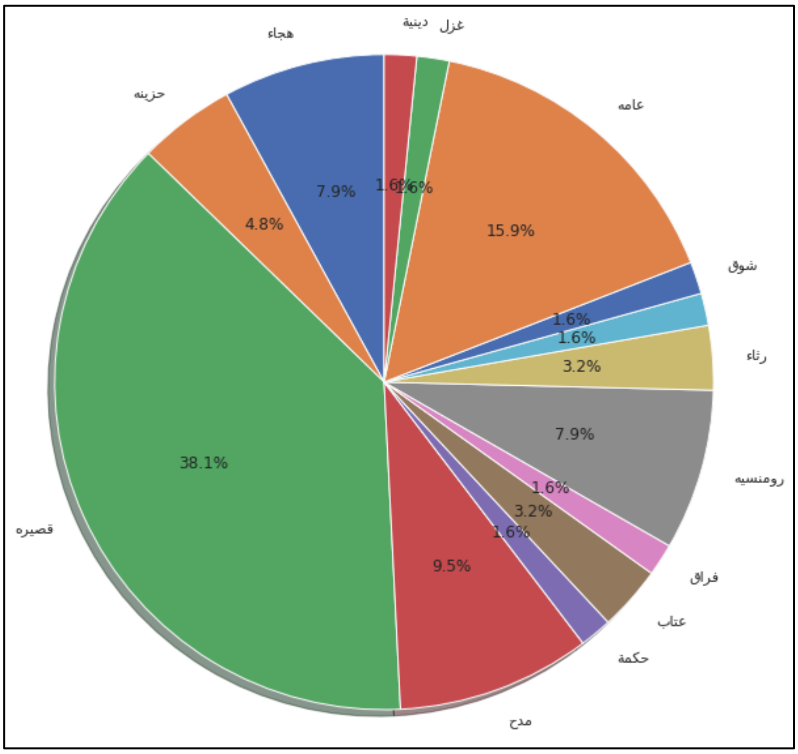

Figure 15: Topic sectioning in the main source.

4. Conclusion

In conclusion, this chapter serves as a foundational exploration of the data landscape crucial to the experiments undertaken in this study. The intricate journey begins with an examination of data sources and the diverse methodologies employed for its acquisition. The subsequent discussion delves into the meticulous steps of pre-processing and cleaning, underscoring the significance of meticulous data preparation for ensuing experimentation phases.

The methodical recategorization of data based on centuries emerges as a key element, offering a structured framework for in-depth analysis. As the chapter unfolds, a detailed portrayal of both corpora is presented, capturing essential metrics such as the number of poems, poets, and overall size in terms of words. Furthermore, the examination of data distribution across eras contributes valuable insights into the chronological representation of poems across different centuries, setting the stage for the analytical journey that follows.

Chapter 6: Topic Modelling[3]

1. Introduction

In our research, the primary objective is AA, which involves identifying the true authors of texts based on their unique writing styles. However, one common challenge is that authorship models often mistakenly focus on topics rather than individual writing styles. This can lead to models learning an author's topic preferences, which may hinder their ability to correctly attribute authorship when the author writes about different topics.

To address this challenge, we are implementing a strategy to control for topics in our authorship experiments. In the data chapter of our research, we obtained data from AlDiwan website, which includes topic labels. However, these labels lack precision and informative details. Therefore, in this chapter, our initial step is to train a topic model using the ETM technique. ETM combines traditional topic modelling with word embeddings to generate more accurate topic representations.

Using the ETM approach, we will explore and identify the various topics present in our data. Subsequently, we will assign approximate topic labels to each text. This will enable us to effectively control for topics when applying authorship models in our experiments.

To implement the ETM technique, we will transform our single-line corpus into a BoW format. This will allow us to utilize a pre-existing topic model and apply it to our dataset of poems. By leveraging this trained model, we will be able to generate appropriate topic labels for each text, enhancing our ability to control for topics in the AA process. Additionally, we examined the correlation between the topics derived from the ETM model and the meters, which was visualized using a heat map.

[3] This chapter was presented in BRISMES Conference.

2. Method

In this study, we selected the Embedded Topic Model (ETM) (Dieng et al., 2020) to identify and categorize the latent topics within our corpus of Arabic poetry. This choice was based on ETM's distinctive capacity to utilize word embeddings in topic modelling. Unlike traditional topic modelling approaches, which are typically constrained by word co-occurrence within documents, the ETM can capture semantic relationships among words, leading to more coherent and interpretable topics. Given the intricate and semantically rich nature of Arabic poetry, we considered the ETM an optimal solution for our research objectives.

To prepare our corpus for ETM application, we first transformed the dataset into a one-line BoW format[4]. This transformation is crucial for topic modelling, as it simplifies text data into a representation based on word frequencies, ignoring word order. This approach aligns with the underlying assumptions of topic modelling, allowing each document to be represented as a vector of word frequencies, thus facilitating compatibility with ETM's input requirements.

Applying the ETM involves tuning several hyperparameters, including the number of topics, learning rate, and the number of epochs. To determine the optimal settings, we conducted a grid search across various parameter ranges. During this process, we assessed the model's performance using the evaluation metrics Diversity and Coherence, aiming to achieve the most accurate and robust topic distribution for the Arabic poetry corpus.

3. Results

In the context of Arabic Poetry topic modelling, the application of the ETM model aims to examine how the diversity and coherence of generated topics are influenced by different numbers of epochs and topics. By experimenting with various configurations, researchers can gain insights into the underlying structure of Arabic poetry, including its themes and motifs.

[4] https://github.com/ssharoff/ETM

In my study, I conducted experiments using three different sets of epoch values: 15, 50, and 70. I also trained the model on different numbers of topics, specifically 5, 10, and 15, to thoroughly analyse the topic generation process. Through these experiments, I aimed to identify the optimal combination of epoch and topic settings that would yield the most coherent and diverse topics for our data.

After conducting the experiments and evaluating the results, it was found that the combination of 50 epochs and 5 topics achieved the best balance of coherence and diversity. This configuration provided the most meaningful and distinct topics for our dataset.

Furthermore, I also explored the performance of the model using four and six topics with 50 epochs. However, the experiment with five topics demonstrated superior results in terms of topic quality, coherence, and diversity compared to these alternative configurations.

By conducting these experiments and analysing the outcomes, I gained valuable insights into the impact of epochs and topic settings on the generated topics. This knowledge contributes to a deeper understanding of the thematic structure and richness of Arabic poetry.

N	**Epochs**	**Topics**	**Diversity**	**Coherence**
1	15	5	0.768	0.231
2		10	0.44	0.219
3		15	0.381	0.223
4	**50**	**5**	**0.872**	**0.236**
5		10	0.712	0.230
6		15	0.554	0.231
7	70	5	0.864	0.230

8		10	0.752	0.234
9		15	0.562	0.235

Figure 16: Topic modelling parameters.

Based on the word sets generated by the model, we can identify the following motifs associated with each topic:

Topic 0: Nature, Beauty, and Aesthetics

This topic revolves around elements of nature, beauty, and aesthetics. It includes words such as "night," "morning," "sun," "beauty," "water," and "roses." The imagery and appreciation of nature's beauty are likely prominent themes within this topic.

Topic 1: Love, Desire, Heartache, and Longing

This topic is centred around emotions of love, desire, heartache, and longing. It incorporates words like "love," "heart," "desire," "tears," and "yearning." It tends to explore themes of romantic love, emotional experiences, and the yearning for a beloved.

Topic 2: Faith, God, Prophet Muhammad, and Morality

This topic delves into matters of faith, featuring words such as "faith," "God," "Prophet Muhammad," and "morality." It tends to explore themes of religious devotion, moral values, and the teachings of Prophet Muhammad.

Topic 3: Life, Death, and Human Existence

This topic addresses the concepts of life, death, and human existence. It includes words like "people," "life," "death," "earth," and "youth." Themes related to the human experience, mortality, and reflections on life's meaning tend to be explored within this topic.

Topic 4: Royalty, Power, Wealth, and Honour

This topic focuses on concepts of royalty, power, wealth, and honour. It incorporates words such as "royalty," "power," "wealth," "glory," and "honour." Themes related to rulership, social status, wealth, and honour tend to be explored within this topic.

By analysing the recurring vocabulary sets and their frequency within each topic, we can identify these motifs that frequently appear in the Arabic poetry corpus. These motifs provide insights into the thematic elements and subjects that poets commonly address in their works.

Topic number	Words
Topic 4	الدهر', 'الزمان', 'المجد', 'ملك', 'الندى', 'الملك', 'الدين', 'العلى', 'الورى', 'الأيام', 'الملوك', 'الجود', 'العلا', 'الدنيا', 'الأرض', 'يد', 'المعالي', 'الوغى', 'فتى
Topic 2	خير', 'قال', 'الناس', 'أهل', 'الحق', 'رب', 'شيء', 'الأمر', 'قلت', 'محمد', 'الإله', 'علم', 'والله', 'النبي', 'الورى', 'عبد', 'الخلق', 'أتى', 'الدنيا
Topic 3	الناس', 'الدهر', 'بن', 'ترى', 'القوم', 'يوما', 'الموت', 'أرى', 'الأرض', 'قوم', 'الليل', 'رأيت', 'فتى', 'فلما', 'الحي', 'الفتى', 'فقلت', 'آل', 'المرء
Topic 1	الهوى', 'قلبي', 'القلب', 'الحب', 'عيني', 'نفسي', 'قلت', 'النوى', 'أرى', 'الدهر', 'قلب', 'فؤادي', 'فقلت', 'هوى', 'الشوق', 'الغرام', 'الدمع', 'البين', 'يوما
Topic 0	الليل', 'الصبا', 'الشمس', 'الحسن', 'ماء', 'بدا', 'بدر', 'الدجى', 'ليل', 'الصبح', 'البدر', 'تحت', 'حسن', 'ترى', 'شمس', 'وجه', 'الصباح', 'الورد', 'اذا

Figure 17: Topics with their most frequent Arabic words.

Qualitative model analysis

In this section, I will randomly select poems from each topic spanning across all eras. The goal is to analyse and evaluate the labelling process based on the model we have trained.

By examining a diverse set of poems from each topic, we can gain insights into the effectiveness of the model's labelling process. This evaluation will help us assess how accurately the model assigns topics to poems and identify any potential areas of improvement.

Through this analysis, we aim to validate the quality and reliability of the topic labels generated by the trained model not only statistically, but also qualitatively. By selecting random poems from different eras, we ensure a comprehensive representation of the entire corpus, allowing us to make meaningful observations and draw reliable conclusions about the performance of the model in assigning topics.

In the appendix, you will find a table that contains the selection of poems for each topic. Here, I will provide an analysis of these poems, focusing on the variations observed between centuries within the same topics, as well as the typical range of topics associated with each label.

By examining the selected poems, we can gain insights into the evolution and variation of topics across different centuries. This analysis allows us to understand how certain topics have been emphasized or transformed over time and provides a deeper understanding of the thematic development within Arabic poetry.

Furthermore, we can observe the range of topics typically associated with each label. This information helps us establish a clearer picture of the dominant themes and motifs within specific topic categories. It also highlights any overlaps or recurring patterns across different labels.

By conducting this analysis, we can uncover valuable insights into the temporal and thematic dynamics of Arabic poetry. This knowledge enhances our understanding of the cultural and historical contexts in which these poems were created and contributes to the broader field of literary studies.

First and foremost, the model has produced highly sensible results, successfully associating the majority of poems with their respective categories. This indicates that the topic modelling approach has effectively captured the underlying themes and motifs present in Arabic poetry.

Through the model's categorization, we can discern clear connections between the selected poems and their assigned topics. The poems exhibit recognizable characteristics and patterns that align with the expected themes within each category. This demonstrates the model's ability to identify and differentiate between different topics based on the textual content of the poems.

When analysing the selected poems for each topic, we can observe various shades or nuances within each topic category. These shades represent the different facets, sub-themes, or variations within a broader topic. By identifying and exploring these shades, we gain a more comprehensive understanding of the depth and complexity of the topics in Arabic poetry.

Certainly, within the topic of **nature, beauty, and aesthetics**, we can observe shades that involve the portrayal of the beloved woman in relation to natural phenomena. These shades highlight the strong association between descriptions of the beloved and natural elements such as the moon and flowers. In Arabic poetry, the beloved woman is often depicted using metaphors and imagery drawn from nature. The moon, with its luminosity and captivating beauty, is frequently compared to the radiance and allure of the beloved. Likewise, flowers symbolize elegance, delicacy, and the enchanting qualities attributed to the beloved woman.

Furthermore, it is noteworthy to mention the connection between the topic of "nature, beauty, and aesthetics" and the theme of wine in Arabic poetry. Wine poems often incorporate descriptions of the surrounding environment and the fruits from which the wine is produced. The imagery of nature is utilized to evoke a sense of sensory pleasure and to create a vivid backdrop for the wine-related experiences described in the poems.

Although the beauty of the beloved woman can be categorized under love, our model exhibits a distinct division between expressions of love and descriptions of women. This separation occurs because the act of describing is associated with nature, treating the beloved as a natural element or experience, akin to other elements found in the natural world.

Under **love**, we see the expression of feelings, especially longing, unrequited love, passionate desire, heartbreak, or the yearning for a beloved. Generally, the experiences depicted in these poems tend to be less joyful or positive in nature.

The exploration of **faith** reveals a progression throughout the centuries. Initially, the focus is primarily on praising Islam and the Prophet Mohammad, reflecting a simpler perspective. However, as time passes, the topic expands to encompass political aspects, including the mention of state leaders. Additionally, the influence of preaching remains prevalent throughout the different eras.

Around the 12th century, a shift occurred, introducing a philosophical element to the discussion of faith. Poets delve into deeper contemplations about the essence of life and humanity, adding a new layer of complexity to the exploration of faith.

This evolution in the portrayal of faith across centuries demonstrates the dynamic nature of the topic and the way poets engage with it in their works. It showcases the multidimensional aspects of faith, encompassing religious devotion, political dynamics, and philosophical reflections, all contributing to a comprehensive understanding of the topic in the context of Arabic poetry.

The **honour** category predominantly revolves around themes of wisdom and life experience. Within this category, poets often delve into topics related to the virtues of honour, dignity, and moral conduct. They draw upon their own life experiences and reflect upon the lessons learned from various situations. The honour poems convey a sense of wisdom, offering insights into the values and principles that shape a person's character. Through their verses, poets highlight the importance of integrity, resilience, and honourable conduct in navigating life's challenges.

The **praise** category is primarily characterized by poems that pay homage and tribute to leaders or famous figures, often highlighting their generosity and bravery. Poets in this category express admiration for individuals who have demonstrated exceptional qualities and accomplishments.

Poets may also express admiration and praise for deceased individuals. These poems serve as a way to honour the memory and legacy of those who have passed away, highlighting their contributions and impact during their lifetime.

It is worth noting that within the praise category, there is a nuanced aspect called lampoon, where poets criticize or fault individuals for lacking the personal traits they admire and celebrate.

Topic distribution

Once the topic modelling was completed, the trained model was utilized to assign likely topics to each poem in the entire corpus. This labelling process provided each poem with a list of probable topics and the corresponding percentage contribution of each topic. The next step involves examining the topical structure of the corpus.

To visualize the distribution of topics, a pie chart displays the relative contribution of each topic across the entire corpus. This chart offers an overview of the overall topic proportions within the collection of poems. Additionally, a line chart illustrates the distribution of topics across different centuries. This chart provides insights into how topics are distributed and potentially vary over time.

Topics across centuries:

By exploring the corpus's topical structure through these visualizations, a deeper understanding of the dominant and evolving topics can be gained. This analysis can be valuable for further exploration and interpretation of the poetic content within the corpus.

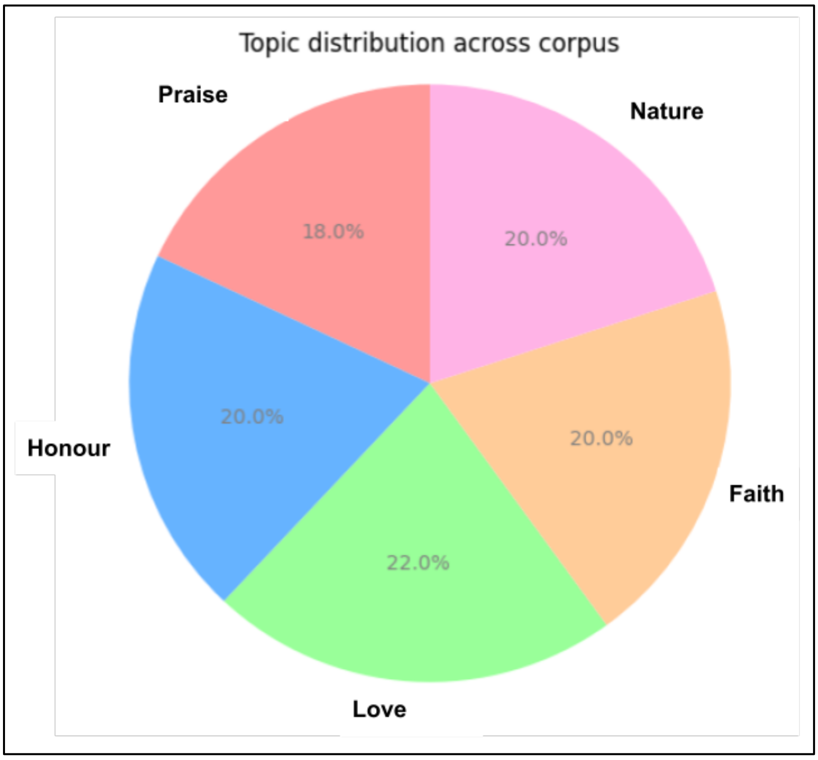

Figure 18: Topic distribution across the whole corpus.

Upon examining the topics distribution in the whole corpus, it appears that there are no significant differences. However, a closer analysis of each century reveals notable shifts in topic prevalence. The 19th and 20th centuries are characterized by a scarcity of poets' data. As a result, these centuries will be excluded from the subsequent graph analysis.

One striking example is the topic of "honour," which exhibits a substantial decline over time. It starts at a high frequency of 80 in the pre-Islamic era but gradually decreases by half over two centuries. This downward trend persists, with the topic rarely exceeding a frequency of 30 in the rest of centuries.

This decline can be understood within the context of social and political transitions from a tribal society to the establishment of Islamic states. Honour poems often reflect individual experiences and tribal conflicts, which became less prominent following the unification of Arab states under a single ruler. Consequently, societal changes and the consolidation of power could have contributed to the diminishing frequency of honour-themed poems in the corpus.

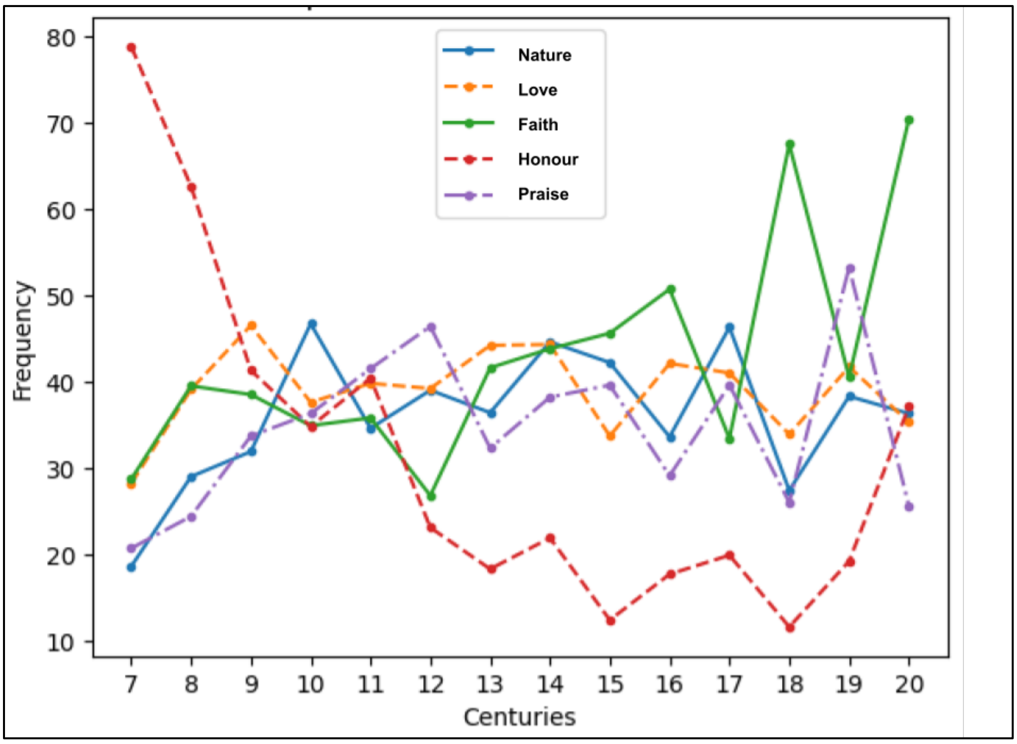

Figure 19: Topic distribution across centuries.

This observation provides insights into the interplay between poetry and the historical, social, and political dynamics of the time. By analysing topic shifts, we can uncover underlying factors that influence the thematic choices of poets throughout different eras.

With praise, we observe a consistent upward trend from the pre-Islamic era up to the 12th century. Following this period, the praise becomes more erratic and fails to reach the same pinnacle again. This pattern can be elucidated by considering the historical context: after the emergence of Islam, the Arab world witnessed the establishment of unified states for the first time, notably under the Umayyad and Abbasid caliphates. During these periods, poets competed to gain prestige within the royal courts and among prominent figures.

Nevertheless, a decline is evident in the 13th century. This decline aligns with a significant historical event—the Mongol invasion of Baghdad, the capital of the Islamic world. This invasion marked a pivotal moment, leading to upheaval in the Islamic realm. The subsequent disruption explains the observed drop in poetic production during that period, as the era of glory for great leaders and flourishing artistic circles came to an end.

Topics correlation with metres:

The heat map analysis reveals the correlation between the five topics and the Arabic poetry metres. Notably, praise poems predominantly utilize the *Kāmil* and *Mutadārak* metres. This suggests that these metres are commonly employed when composing poems of praise.

On the other hand, love and nature topics exhibit similar metre preferences, with *Muqtaḍab* and *Muḍāri* ʿ being frequently employed. This indicates that these metres are commonly associated with expressions of love and descriptions of nature in Arabic poetry.

For the topic of honour, the most commonly used metres are *Wāfir* and *Ṭawīl* . This implies that these metres are favoured when conveying themes related to honour in poetic compositions.

In the case of faith, the metres *Sarī*, *Ramal*, and *Hazaj* emerge as the most common choices. This suggests that these metres are often utilized when expressing themes of religious devotion or matters related to faith.

To read a heat map showing the correlation between topics and meters, look at the colour intensity or shading of the cells. Darker or more intense colours typically indicate weaker correlations between specific topics and meters or no correlations, while lighter colours suggest stronger. Each cell corresponds to a pair of topic and metre, and the scale bar will help you interpret the strength of these correlations.

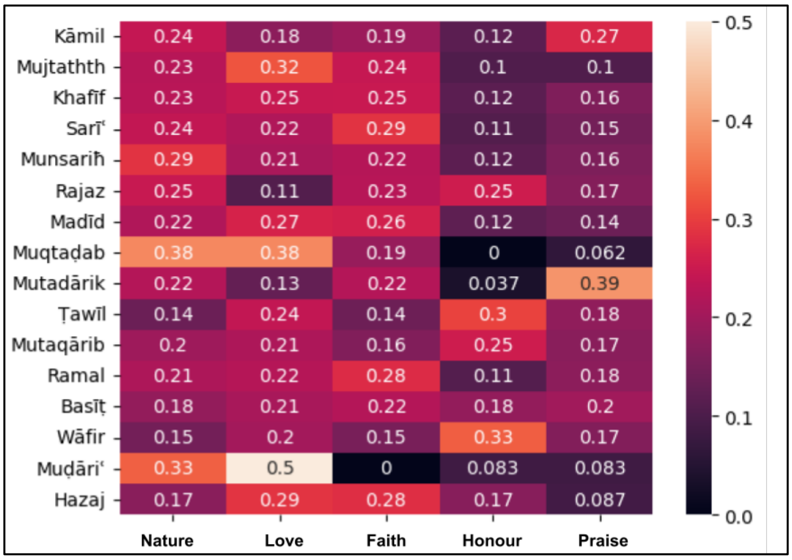

Figure 20: Topics correlations with metres.

The heat map analysis provides valuable insights into the correlation between specific topics and the metres employed in Arabic poetry. By understanding these associations, researchers and scholars can gain a deeper understanding of the poetic conventions and traditions associated with different thematic domains.

4. Conclusion

In conclusion, our research focused on AA, aiming to identify authors based on their unique writing styles. Recognizing the challenge of models emphasizing topics over styles, we implemented a strategy to control for topics. In the data chapter, we obtained data from the AlDiwan website, including topic labels with limited precision. To address this, we employed ETM to enhance topic representations. This involved transforming our corpus into a BoW format, allowing us to leverage a pre-existing topic model for accurate topic control in authorship experiments. The unveiling of this aspect will occur in the subsequent chapter upon the construction of the classification model.

Chapter 7: Ensemble Model for Authorship Classification[5]

1. Introduction

In this chapter, I will present my method for investigating doubted texts, particularly in the pre-Islamic era. The primary objective is to construct a scenario that closely resembles cases found in pre-Islamic data. An essential aspect to take into account during the model's design is the constrained data availability concerning pre-Islamic poetry. In this context, we might only have access to a few tens of examples, or potentially even fewer, to utilize for training. Consequently, determining the highest probability of assigning a text becomes challenging, as the model has not been previously tested on other cases where the truth is known.

Therefore, our objective is to develop a model and evaluate its performance on other eras where the correct answers are known. By subjecting the model to rigorous testing, we can identify the optimal configurations and parameters before applying it to the pre-Islamic cases. This approach will allow us to refine and validate the model's effectiveness, ensuring its accuracy and reliability in handling the scarcity of data typical in pre-Islamic poetry. Ultimately, our goal is to establish a robust and well-calibrated model that can confidently handle the challenges posed by the pre-Islamic era texts.

The problem

In the context of AA tasks, a widely-used method involves the creation of an opposite class, comprising writings from all authors except the one currently under investigation. The aim is to discern unique authorial patterns by contrasting the specific author's writing style against the collective pool of other authors' works.

[5] A paper derived from the work in this chapter was presented and recognized as an award-winning contribution at the 31st **COLING** conference.

However, in our specific case, we face a constraint where the available data for the author of interest, is limited to only a few examples. As a result, the opposite class, which amalgamates writings from all other authors, may still contain a substantial amount of data.

While having more data could potentially improve the model's performance, it is crucial to address concerns related to overfitting. Overfitting occurs when a model becomes overly specialized to the training data and loses its ability to generalize to unseen data. Given the data imbalance resulting from a larger opposite class, the model might excessively prioritize the more abundant class while inadequately representing the author of interest, compromising the model's effectiveness in practical AA tasks.

To mitigate the risk of overfitting, careful model training and regularization techniques should be employed. It is essential to strike a balance between utilizing the available data in the opposite class to improve model performance while avoiding excessive reliance on this data at the expense of the author of interest.

Conversely, if our intention is to achieve a balanced size between both classes and we opt to restrict the number of examples from other authors, this approach could lead to an unrepresentative class composition. This scenario might result in having merely one or two examples available for each author within that class.

2. Methodology

The model

To address the issues of data scarcity and potential overfitting in the context of AA, we propose a multi-model approach combined with an ensemble method. This entails developing multiple distinct models, each capturing different authors and authorial patterns. Each model is fine-tuned CAMeLBERT (Go Inoue et al., 2021) model.

Throughout the experimentation and evaluation, we will employ rigorous validation techniques to assess the ensemble's performance and its ability to generalize to unseen data.

Furthermore, we will fine-tune the ensemble's parameters to optimize its efficacy for the specific AA task.

In light of the challenges posed by data scarcity and potential overfitting in the AA task, an alternative approach is proposed. Rather than creating a single opposite class containing writings from all other authors, I will construct several distinct datasets, each excluding the author's writings under investigation. This approach aims to develop individual models for each dataset, capturing unique authorial patterns and enhancing model diversity.

Once the models are trained on their respective datasets, they will be evaluated on unseen data. To aggregate their predictions effectively, I will employ the voting technique, where the model outputs are combined to determine the highest probability of the predicted class. By leveraging the collective decisions of multiple models, the ensemble's final attribution will tend to yield improved accuracy and robustness.

To evaluate the models' performance, the accuracy metric will be employed, quantifying the number of correct predictions made by each model. A higher accuracy score signifies the model's ability to correctly identify the authors of the unseen data instances, underscoring its efficacy in AA.

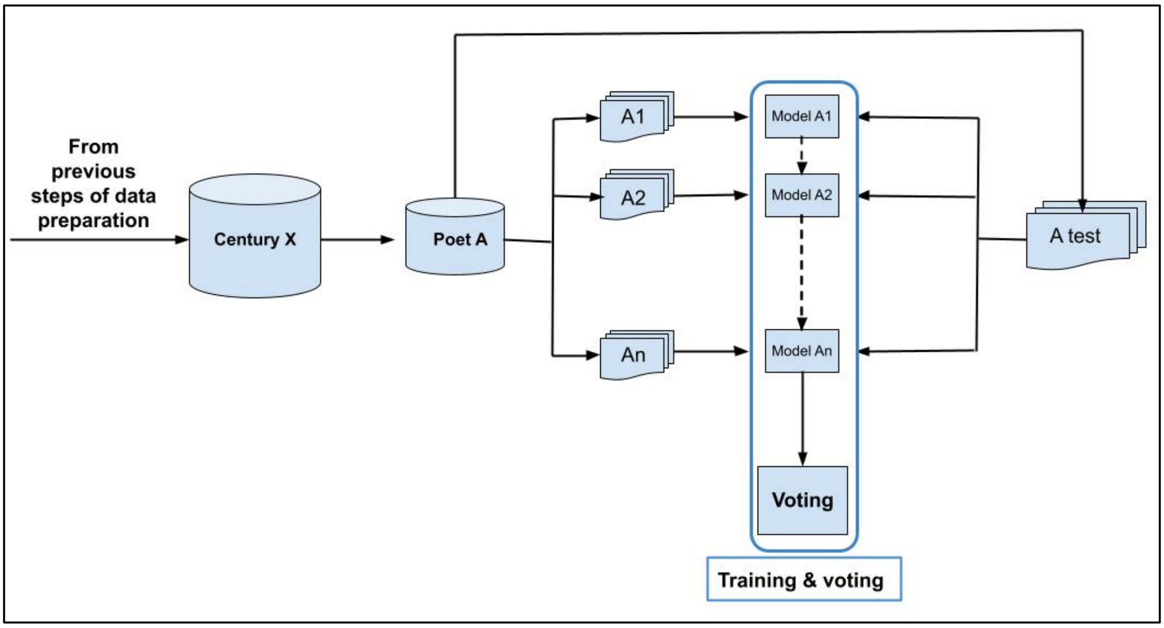

Figure 21: Flowchart of the model procedures.

Voting

In the domain of author classification, we are faced with the task of categorizing authors into specific classes. This scenario introduces a unique challenge in which we have one author of paramount importance (Class 1) and a group of authors with distinctive characteristics (Class 2). To address this challenge, we have developed a voting technique that tailors decision rules for each class, backed by empirical evidence from rigorous testing.

Class 1 Voting Requirement

For Class 1, representing our highly significant author, we have established a stringent requirement: at least 40% of the models in our ensemble should align in predicting this class. This requirement is not arbitrary but rather based on empirical findings from model testing.

Our choice to set the threshold at 40% is rooted in the fact that it has consistently yielded the best results during testing. Requiring a 40% consensus among models minimizes the likelihood of false negatives and ensures the highest reliability in predictions regarding this key author.

Class 2 Voting Requirement

Class 2 includes a diverse set of authors, each contributing unique attributes and writing styles. Acknowledging this diversity, we employ a more permissive voting requirement: a prediction by just one of the base models in our ensemble is sufficient to classify an instance into Class 2. This is because the author can only exist in one of the base models.

Our approach is underpinned by several key justifications:

1. Empirical Testing: Our voting technique is not theoretical but stems from empirical testing. It ensures the best results for our specific problem.

2. Handling Author Heterogeneity: The diversity within Class 2 demands flexibility in our decision rules to effectively capture the nuances of each author's work.

3. Prioritizing Key Author (Class 1): The 40% consensus requirement for Class 1 prioritizes accurate predictions for the highly important author, balancing between false positives and false negatives.

4. Balancing Sensitivity and Precision: Our approach achieves a balance between sensitivity and Precision based on the targeted author, enhancing overall model performance.

During the training phase, we employ binary model decisions; however, for testing on actual pre-Islamic poetry cases, we will utilize probabilistic outputs. Probabilities provide a more reliable measure of the model's confidence in its predictions. It is essential to consider the relative nature of these probability values. Typically, probabilities above 0.70 are regarded as high and indicate acceptable confidence in classification. Nonetheless, our model consistently achieved very high probabilities for correct answers, with values not falling below 0.95. Consequently, even misclassifications produced relatively high probabilities, often within the 0.80 range. Therefore, any probability in the 0.80s or lower will be interpreted as low confidence, suggesting that the text likely does not belong to the author under investigation.

Performance Validation

We have thoroughly verified the performance of our ensemble model through meticulous validation, which affirms that the selected voting percentages are the most effective for our specific problem. In conjunction with our model training and testing procedures, we devised a dedicated evaluation test. In this test, we provided each model with a dataset of the same size as the one initially fed to the model for training and testing, ensuring that it contained a completely unseen set of data for validation.

Factors to control

Constructing an optimal model involves the careful consideration of various factors, one of which pertains to the number of poets in the opposite class. First, we will focus on examining the influence of this specific parameter on model performance. To obtain a precise understanding of the optimal value, we conducted rigorous testing on poets from diverse age groups. The goal is to ensure the accuracy of our model by identifying the most suitable parameter value for the **number of poets** in the opposite class.

An additional factor to take into consideration is the varying **number of examples**. There is no a predetermined quantity that we can anticipate for the pre-Islamic context. Furthermore, we encounter multiple instances where each case exhibits a distinct amount of available data. Given this situation, it becomes imperative to conduct tests across a range of numerical values and examine how the model's performance evolves when the data size is altered.

Considering probabilities

Furthermore, I will investigate whether there existed any variance in the probability distribution between correctly classified cases and misclassified cases. This analysis will provide us with a more profound comprehension of the model's performance, allowing us to determine the threshold probability score at which we can confidently deem the output as highly reliable.

3. Results

To thoroughly evaluate the performance of our AA model, we tested it across three distinct scenarios: one author against one (one-vs-one), one author against two (one-vs-two), and one author against three (one-vs-three). The goal was to determine the optimal configuration for the model to achieve the highest accuracy and reliability in classification.

In these experiments, the one-vs-one configuration consistently yielded the best results. This finding is depicted in the box charts, which illustrate the distribution of classification accuracies across the different test cases. The one-vs-one scenario demonstrated the highest median accuracy and the smallest interquartile range, indicating not only superior performance but also greater consistency and reliability.

In contrast, the one-vs-two and one-vs-three scenarios exhibited a decline in accuracy. The box charts for these cases show lower median accuracies and wider interquartile ranges, reflecting increased variability and reduced performance. The degradation in accuracy as we added more authors to the opposite class suggests that the model struggles with distinguishing the target author's style when more distractor authors are present.

To address the observed decline in performance in one-vs-multiple author scenarios, we implemented a voting technique. This technique involved using multiple one-vs-one classifiers and aggregating their decisions through majority voting. By leveraging the strengths of the one-vs-one classifiers, the voting method improved the overall robustness of the model, resulting in more reliable AAs even in complex cases.

The results of these experiments underscore the efficacy of the one-vs-one approach in AA tasks. The box charts provide clear visual evidence of the superior performance and reliability of the one-vs-one configuration, supporting its adoption as the preferred parameter for our model.

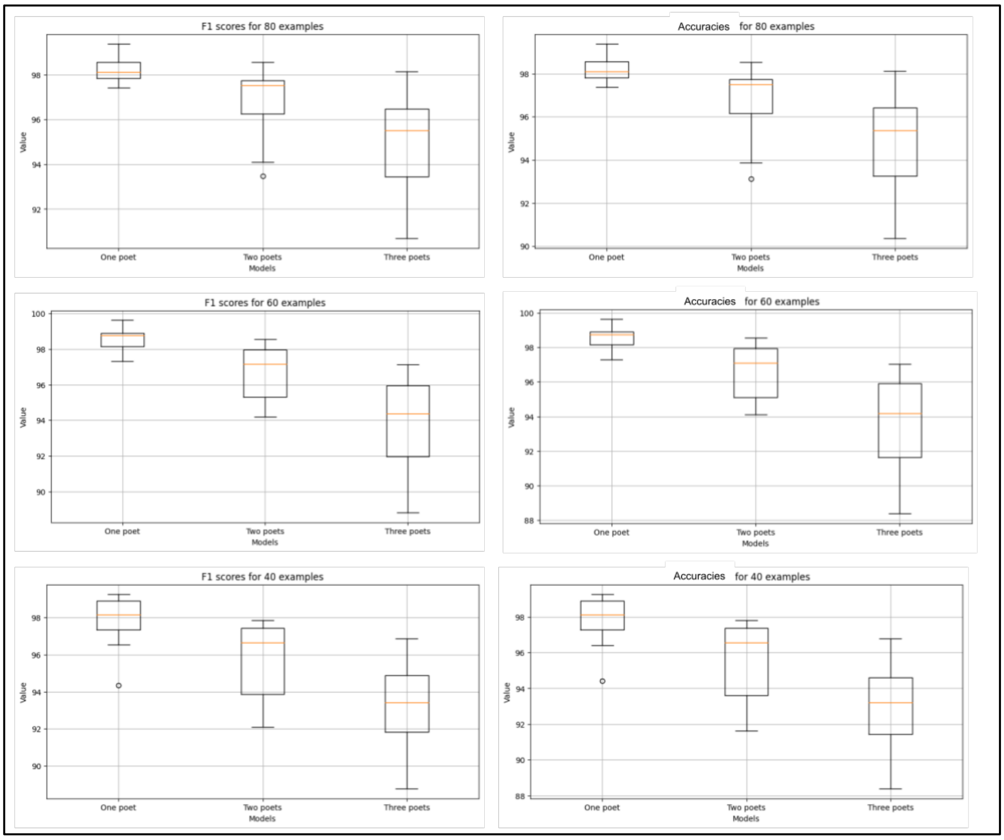

Figure 22: Accuracies and F1 scores for varying the number of poets in the opposite class.

The performance of models when varying the number of training examples indicates that utilizing 60 examples yields optimal results. To further analyse the outcomes, we compared the probabilities assigned to correct and incorrect classifications. This comparison revealed a significant difference in the confidence levels of the model, depending on whether it accurately predicted the outcome or not.

In our evaluation of the AA model, we observed a distinct pattern in the classification accuracies that underscores the model's reliability and confidence. Specifically, as depicted in the accompanying line chart, the accuracies for well-classified cases predominantly exceeded 0.95, indicating a very high level of confidence in these classifications. This suggests that when the model assigns a text to an author with such high probability, it is highly likely that the attribution is correct.

Conversely, for misclassified cases, the accuracies generally fell between 0.80 and 0.90. While these probabilities are relatively high, they are notably lower than the accuracies observed in well-classified instances. This difference in accuracy ranges highlights the model's ability to discern between correct and incorrect classifications. The fact that misclassified texts still receive moderately high probabilities suggests that the model is uncertain, yet it tends to err on the side of caution.

This distribution of accuracies, as illustrated in the line chart, reflects the model's overall robustness and effectiveness. The high confidence levels in well-classified cases (above 0.95) provide strong evidence of the model's precision and reliability in AA. The relatively lower confidence levels in misclassified cases (0.80-0.90) further reinforce the model's discriminative power, indicating that when the model is less certain, it signals a potential misclassification.

In practical terms, this means that for real-world applications, texts classified with probabilities above 0.95 can be considered highly reliable attributions. Texts with probabilities in the 0.80-0.90 range, while still reasonably confident, should be treated with more scrutiny and further verification may be warranted.

Overall, these findings, clearly depicted in the line chart, demonstrate that our model not only performs well in accurately attributing texts to authors but also provides meaningful confidence metrics that can guide the interpretation and validation of its classifications.

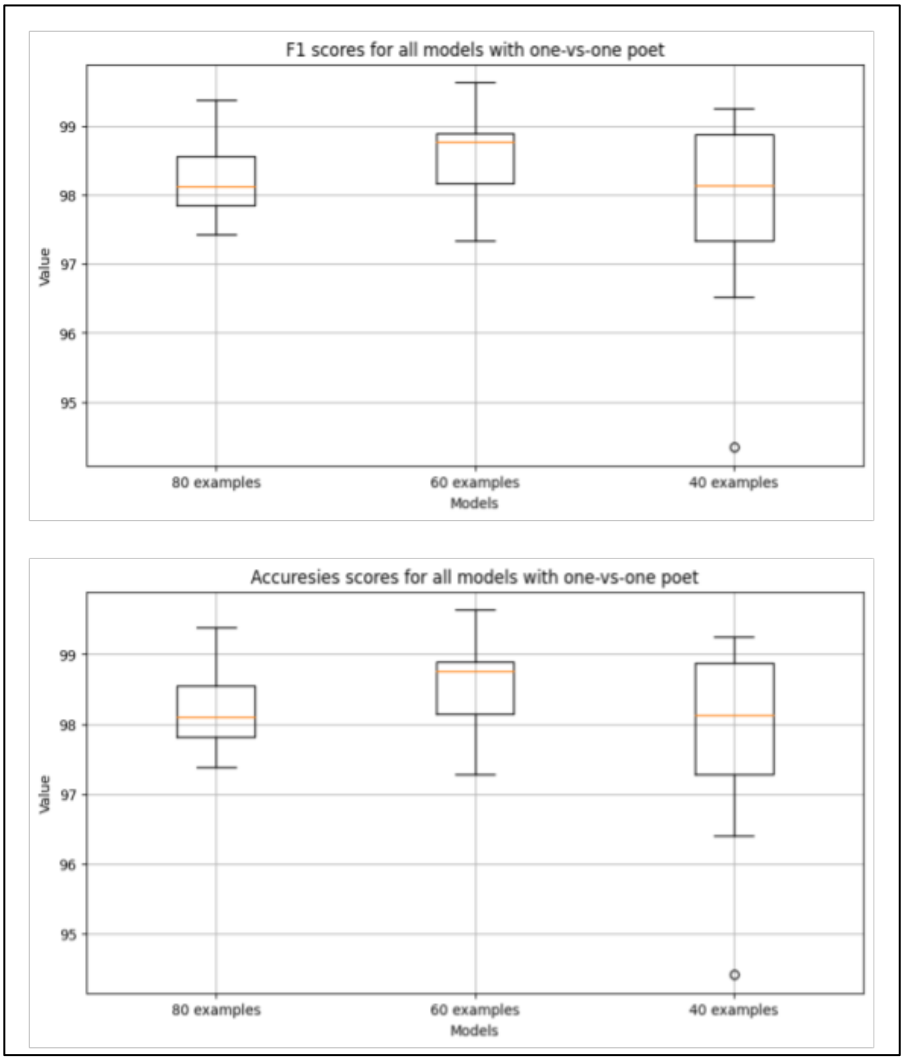

Figure 23: Box chart of accuracies and F1 scores for varying the number of examples.

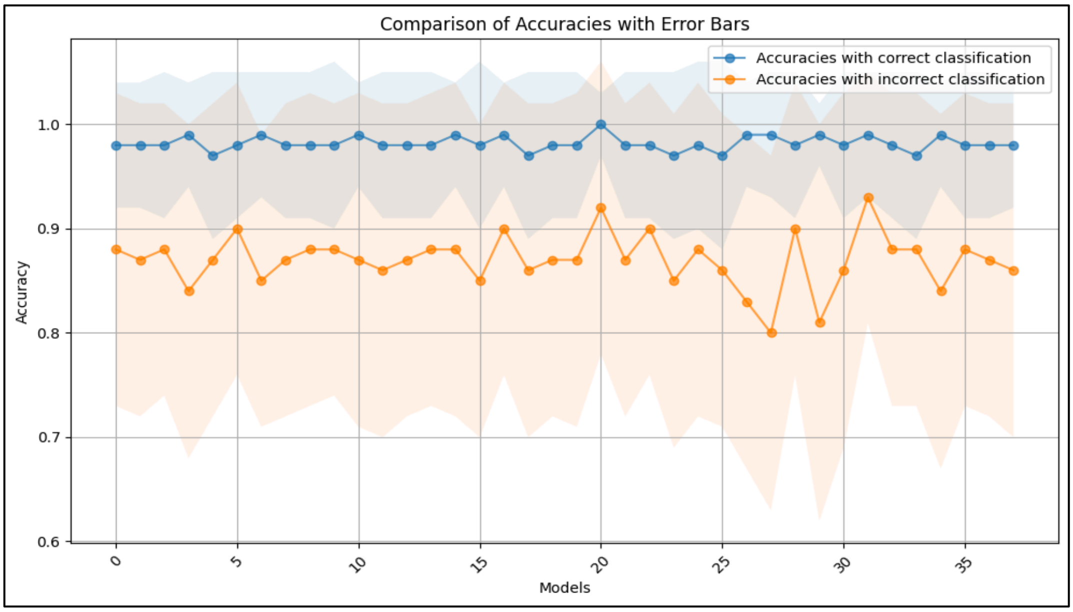

Figure 24: Comparison between accuracies with correct and in correct classification.

4. Applying topic control

Before constructing the authorship classifier, it is crucial to examine the intersection of topics and author styles by having dependable topic labels. This step is vital as machine learning models acquire knowledge of both topic-specific language and unique authorial traits, and understanding the overlap between these aspects is essential.

In order to accomplish this, I constructed one-versus-all binary classifiers using BERT. For each age group, I specifically chose poets who have a minimum of 100 poems, ensuring an adequate amount of training data for topic segmentation. Consequently, each classifier is trained on a particular topic for a given poet, distinguishing them from their peers in the same era.

With the intention of controlling the number of training examples and focusing on the variation between topics, I implemented a threshold for each topic. Specifically, I set a maximum limit of 100 poems and a minimum limit of 30 poems for each topic.

By imposing this threshold, I ensured that each topic had a sufficient number of training examples while preventing any topic from dominating the dataset. This approach allows for a balanced representation of poems across topics, enabling a fair comparison and analysis of the variations and distinctions between topics.

Broadly speaking when topic control was implemented, the accuracies of the classifiers were higher. This suggests that considering topics is important when applying authorship classifiers in subsequent experiments. By acknowledging the influence of topics on the performance of classifiers, it emphasizes the significance of incorporating topic-related factors into the design and analysis of authorship classification tasks.

The outliers depicted by the small circles in the box plot represent the cases where we have the fewest minimum examples available for analysis. These outliers have a noticeable impact on the results, as the limited number of examples can introduce higher variability and potentially affect the accuracy of the classification.

On the other hand, we do not observe outliers with low accuracy in the "mix" category. This is expected since the "mix" category generally consists of a combination of different topics, ensuring that we always have a sufficient number of examples available for analysis. The larger sample size in the "mix" category helps stabilize the results and reduces the likelihood of encountering outliers with low accuracy.

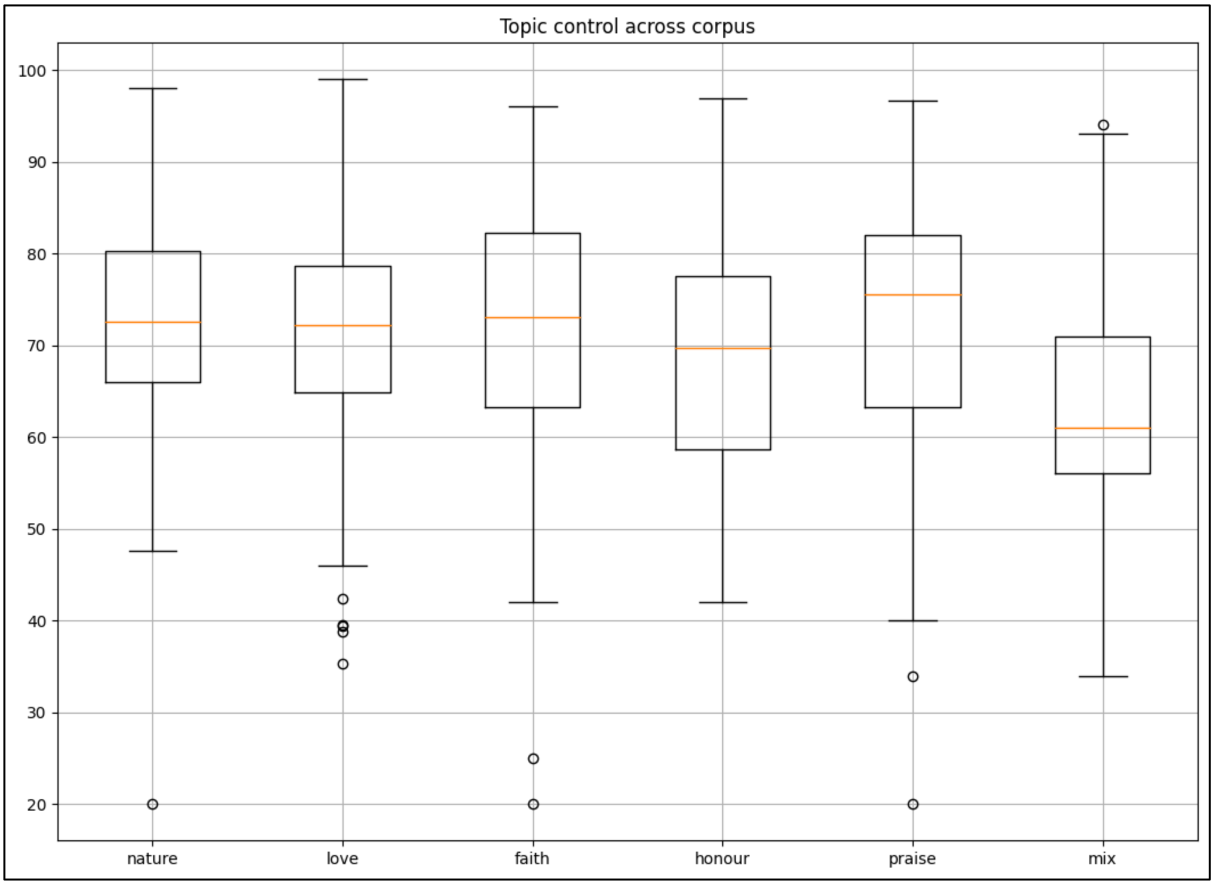

Figure 25: Comparing classification accuracies with certain topics control and no topic control.

To conduct a more comprehensive examination of the topic effect, I utilized the same binary classification task. In this case, I trained the model by considering one poet against all other poets from the same age group. To ensure an adequate amount of data variation, I specifically chose poets who had a minimum of 100 poems in their repertoire.

Unlike the previous approach, I included all the poems of each author without any sectioning. Once the model was trained, I examined the instances where the classification failed and assigned topics to those examples. By doing so, I determined the ratio of failed examples within each topic.

This analysis provided valuable insights into the weaknesses of each poet in terms of topics. Additionally, it offered a broader perspective on the areas where the models struggled the most. By identifying the topics that posed challenges to the model's classification accuracy, we can gain a better understanding of the limitations and areas for improvement in the model's performance.

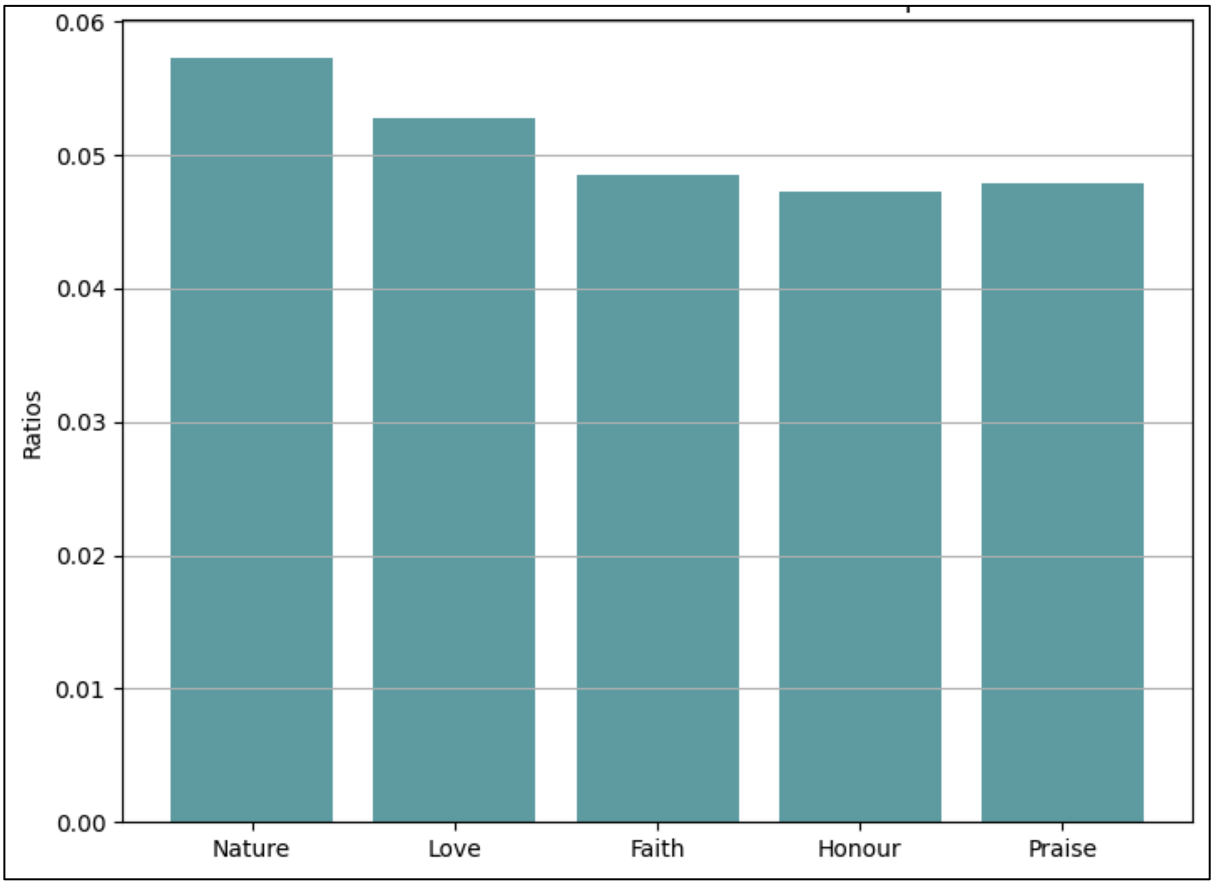

Figure 26: Ratio for classification fails for each topic.

The bar chart reveals that the topic "nature" exhibited the highest misclassification ratio, while the topic "honour" demonstrated the lowest. However, the differences between the misclassification ratios for the various topics were not significant, indicating that the models generally do not exhibit a preference for one topic over the others.

Nevertheless, this analysis provides valuable insights at a micro level. It highlights that in specific cases, there can be noticeable differences in the misclassification ratios. For instance, Abū al-ʻAtāhiyya had a misclassification ratio between the topic "nature" and the rest of the topics (except for "love") that was more than twice as high. This information, when examined for each individual poet, will prove useful in subsequent authorship experiments.

By identifying the specific patterns of misclassification and understanding how certain topics may be more challenging for the models to accurately classify, we can refine our approach and potentially enhance the AA accuracy in future experiments.

5. Comparing with Burrows-Delta

Given the prominence of Burrows-Delta as a widely recognized method for AA tasks, it became imperative to assess its efficacy within the context of our specific investigative task. This imperative arises from the inherent significance of benchmarking against established methodologies within the field. By subjecting Burrows-Delta to scrutiny in our study, we endeavour to discern its comparative performance and ascertain its utility vis-à-vis novel approaches under examination.

In this section, the Wilcoxon Test was employed to compare the outcomes of Burrows-Delta and our model. The results indicated a substantial disparity between the performance of our model and the Burrows-Delta method, with the LLM model exhibiting superior efficacy. The obtained p-value of **0.0048** further confirms this distinction.

6. Pre-Islamic Case Study

This section introduces the final objective of the AA model, following extensive testing across various data points within the corpus. Our goal was to create a model capable of assigning texts to their correct authors with high confidence, enabling its application to cases with disputed authorship.

Method

The methodology adopted in this section includes the following steps:

1. Selection of Core Poets in the Pre-Islamic Era

The primary focus is on the leading poets of the pre-Islamic period, specifically the ten poets who are renowned for composing the celebrated collection known as Al-Muʻallaqāt.

2. Literary Research and Categorization

A thorough investigation of literary sources was conducted to examine the attribution of these poets' works. The poems were then categorized into three groups: those with definite attribution, those with definite denial, and those where authorship is doubtful.

3. Model Training and Validation

The model was trained on poems with definite attribution and tested on those with definite denial to assess its performance and reliability in the pre-Islamic context. To enhance accuracy, this step relied on the probability assigned by the model to each class rather than just the final classification outcome.

4. Testing with Disputed Texts

The model was tested on texts where most literary studies have questioned the authorship, allowing for an analysis of the model's concordance with these doubts.

5. Testing with Ambiguously Attributed Texts

The model was also applied to texts in the third category, where attribution is ambiguous or less certain, to evaluate its capability in addressing such uncertainties.

By following this methodology, we aimed to build a robust AA model that could effectively resolve complex cases of disputed authorship in Arabic poetry, providing valuable insights into one of the most challenging aspects of literary studies.

Applying to pre-Islamic text

In this section we will show the results of our model verifying if certain texts belong to their assumed authors. It’s worth mentioning that all texts that will be tested are assigned to In this section, we present the outcomes of our model's assessment aimed at verifying the authorial attribution of specific texts. It is pertinent to note that all texts subjected to testing are initially ascribed to their respective assumed authors within both corpora. This assumption persists despite any lingering uncertainties surrounding their authorship.

Poem 1:

As per our outlined strategy, we will commence our evaluation process with texts that have garnered the highest degree of scepticism, aiming to ascertain the degree of alignment between our model's predictions and the prevailing doubts. The initial text under scrutiny pertains to Imru’ al-Qays, a figure widely regarded as one of the most eminent poets of the pre-Islamic era, if not the entirety of Arabic literature. This particular poem, however, has not been attributed to Imru’ al-Qays in any authoritative sources. Furthermore, its structural composition deviates notably from the established patterns observed in his other works. While it adheres to the conventional thematic trajectory characteristic of Imru’ al-Qays's compositions, wherein he reminisces about bygone days and romantic escapades, certain linguistic anomalies are discernible. These anomalies include the presence of uncommon and archaic vocabulary, as well as instances of repetitive phrases devoid of semantic coherence. To illustrate this point, a single line from the poem will be provided for further elucidation.

أَلا لا أَلَا إِلَّا لَآلاءِ لابِثٍ

ولا لَا أَلَا إِلا لِآلاءِ مَن رَحَل

The poem in question underwent classification against the collective body of works attributed to the ten authors under consideration. The poem failed to exhibit congruence with the established stylistic and thematic markers associated with any of the authors, including Imru’ al-Qays, thereby casting doubt upon its authenticity. Notably, the model evinced a degree of perplexity and ambiguity in its attempt to assign authorship to this particular poem. While the model consistently yielded high confidence scores upwards of 0.99 for the attribution of other poems to Imru’ al-Qays, its confidence rating plummeted to a mere **0.71** for this anomalous composition. This divergence underscores both the validity of our suspicions regarding the poem's attribution and the model's discernment in identifying instances of potential misattribution.

Poem 2:

The subsequent text under examination is a poem attributed to Al-Aʿshā, purportedly composed during the waning years of his life praising the Prophet Mohammad. While the initial verses exhibit semblances of Al-Aʿshā's characteristic style, a discernible departure in style becomes evident upon closer inspection of the latter stanzas. Scholars have observed a discernible weakness in compositional quality in the latter half of the poem, which diverges from the established literary repertoire of Al-Aʿshā. Furthermore, the inclusion of detailed doctrinal expositions on Islamic teachings, such as the prohibition of consuming carrion, raises scepticism regarding Al-Aʿshā's authorship. This scepticism is compounded by historical accounts suggesting Al-Aʿshā's resistance to Islamic conversion, evidenced by his predilection for alcohol consumption, as depicted in his prior works. Given these discrepancies, the attribution of such intricate Islamic doctrinal knowledge to Al-Aʿshā appears implausible. The computational model exhibited notable hesitancy in attributing this poem to Al-Aʿshā, despite the historical plausibility of its initial segment aligning with his stylistic tendencies. The model's assignment probability for the entire poem, incorporating both the undoubted and disputed segments, was recorded at **0.86**.

Poem 3:

The analysis of another poem attributed to Al-Aʿshā yielded a certainty score of **0.85,** indicating a comparatively lower degree of confidence when juxtaposed with other attribution cases. Historical scepticism surrounding this poem is evident, with ancient researchers deeming it to be fabricated due to its perceived mediocrity. Additional insight from the manuscript editor of Al-Aʿshā's anthology further supports these doubts, citing the poem's overly philosophical nature and its conspicuous departure from the characteristic Islamic spirit associated with Al-Aʿshā's compositions.

Based on the findings presented, it can be confidently asserted that we have developed a computational methodology capable of effectively scrutinizing the verification of contested poems with a satisfactory level of confidence. The model demonstrated higher certainty in authenticating poems suspected of being fabricated compared to those entangled in disputes involving multiple authors.

Poem 4:

The poem attributed to ʻUbayd Ibn Al-Abraṣ, consisting of 21 lines, has been a subject of scholarly debate regarding its authorship. The primary reason for the uncertainty is that the poem is found exclusively in ʻUbayd's anthology, with no other sources corroborating the attribution. Despite this, stylistic analysis has not raised significant doubts, except for the second line, which some scholars feel is more characteristic of the Islamic period. There is a scholarly consensus that while the poem may indeed belong to ʻUbayd, it is possible that parts of it could have been authored by Aws Ibn Ḥajar.

Linguistically and stylistically, the poem aligns with pre-Islamic traditions, and there is no indication that it was fabricated. Upon testing this poem with our AA model, we obtained an average probability of 0.99, strongly suggesting that the poem belongs to ‘Ubayd. It is important to note that our model is not designed to detect multi-author contributions, and since Aws Ibn Ḥajar was not included in the opposing group, any potential co-authorship could not be identified.

In summary, the existing doubts about the poem's authorship are not substantial, and our model's high probability score corroborates its attribution to ‘Ubayd Ibn Al-Abraṣ.

7. Conclusion

In conclusion, this chapter has outlined a methodical approach to investigating doubted texts from the pre-Islamic era, emphasizing the challenges posed by limited data availability. By developing and rigorously testing a model on eras with known authorship, we aimed to refine and validate its effectiveness. This process enabled us to identify the optimal configurations and parameters, ensuring the model's accuracy and reliability despite the scarcity of pre-Islamic poetry data. Ultimately, the goal was to create a robust and well-calibrated model capable of confidently addressing the complexities and uncertainties inherent in pre-Islamic texts.

We managed to test four poems and achieved good corroboration with existing scholarship. However, we faced several challenges. Authors varied significantly in data size, with some having no more than 16 examples, while others had over 100. Additionally, the length of poems varied greatly; some consisted of only one line, making them inadequate for model testing. These variations posed significant hurdles in developing a consistent and reliable model.

Furthermore, exploring the authorship of poems is an inherently complex task, often requiring extensive analysis and validation. It can demand detailed, book-length studies for single poems, as evidenced by the comprehensive work of Shaker (1994). This endeavour highlights the intricate nature of AA in ancient poetry, necessitating meticulous analysis and validation.

Chapter 8: Conclusions and Future Work

1. Introduction

This chapter provides a comprehensive summary of the key findings from my study exploring AA in Classical Arabic poetry, with an emphasis on attribution challenges in pre-Islamic texts. It also outlines the scope and limitations of the research while presenting notable contributions to the field of Arabic literary studies. Through a detailed examination of quantitative approaches, the chapter discusses how computational techniques can offer new insights into AA issues. It further suggests potential directions for future research to enhance the current understanding and address unresolved questions in Arabic poetry.

2. Summary of Key Findings

This research aimed to investigate AA in Classical Arabic poetry, with a particular emphasis on attribution challenges in pre-Islamic poetry. The key findings from this study are as follows:

1. First Topic Labelling for Classical Arabic Poetry
This study introduces a novel approach to topic labelling across the entire corpus of classical Arabic poetry. Unlike traditional methods that assign texts to concrete categories, this approach categorizes texts by the proportion of their contribution to major topics. We applied ETM to identify five main topics which they are nature, love, honour, faith and praise. When we tested the AA model with and without topic control, the results showed improved performance with topic control, but the minimal difference suggests limited impact of topic confounding with style, indicating minimal overlap between topics and authors.

2. Relation Between Time Periods and Topic Shifts
This study explores the connection between various historical periods and topic shifts in Arabic poetry. It provides insights that link these shifts to significant historical events, contributing to a better understanding of the broader context. For example, the exploration of faith has evolved over centuries. Initially centred on praising Islam and the Prophet

Mohammad, it later expanded to include political elements, such as references to state leaders. Preaching remained a consistent influence across eras. Around the 12th century, a philosophical dimension emerged, with poets exploring deeper questions about life and humanity, adding complexity to the discourse on faith.

3. Correlation Between Metres and Topics

The research uncovers a significant correlation between certain topics and specific poetic metres, delivering the first statistical proof for what had previously been observed only through qualitative analysis. This finding is noteworthy, as traditional studies on poetic metres have often lacked consensus and rarely connected these observations to a large corpus of texts. For example, praise poems primarily use the Kāmil and Mutadarik meters, whereas Wāfir and Ṭawīl meters are most commonly employed for the theme of honour.

4. Transformers-Based Model for AA

The study builds the first model that relies on transformers encoding to resolve AA issues in Arabic poetry, offering a sophisticated approach to an enduring problem. In this part of the study, the Wilcoxon Test was used to evaluate the performance differences between Burrows-Delta and our model. The findings revealed a significant gap in performance, with our model outperforming the Burrows-Delta approach. The p-value of 0.0048 reinforces this notable difference. An analysis of the average F1 scores for the tested models reveals that the ensemble model achieved a score of 0.98, outperforming the single model, which obtained 0.75, and Burrows' Delta, which scored 0.72.

5. Comprehensive Corpus Usage

This study examined the entire corpus of classical Arabic poetry, spanning sixteen centuries, making it one of the most comprehensive computational analyses in the field. This dataset contains 784 poets, 77,850 poems, and a total of 6,609,495 words.

6. Ensemble Model Achieving State-of-the-Art Results

The research achieves state-of-the-art results in AA classification using an ensemble model, demonstrating the effectiveness of advanced computational techniques in this context. Our ensemble model achieved an average accuracy and F1 score of 0.98, while the single model attained an accuracy of 0.73 and an F1 score of 0.75.

7. Quantitative Solutions for AA

This study is the first to offer quantitative solutions to AA problems in Arabic literary studies, providing concrete answers to previously unresolved questions as shown in the four poems analysis at the end of chapter 7.

As per our strategy, we will begin our evaluation by analysing texts that have generated the most scepticism, aiming to determine how closely our model's predictions align with existing doubts. The first text to be reviewed is attributed to Imru' al-Qays, a renowned poet of the pre-Islamic era, but this poem has not been recognized as his work in any authoritative sources. Additionally, its structure significantly differs from the typical patterns found in his other compositions. Although the poem follows the usual themes associated with Imru' al-Qays, such as reminiscing about past experiences and romantic encounters, there are notable linguistic irregularities, including rare and archaic words and repetitive phrases that lack coherence.

When this poem was compared to the works of ten other authors, it did not match the stylistic and thematic characteristics of any, including those of Imru' al-Qays, raising doubts about its authenticity. The model struggled to assign clear authorship, showing a sharp drop in confidence—from consistent high scores of over 0.99 for Imru' al-Qays's other works to just 0.71 for this poem. This difference supports our suspicion about the poem's questionable attribution and highlights the model's ability to detect potential misattributions.

3. Addressing Research Questions

The primary objective of this research was to determine whether computational methods can effectively address literary questions, particularly those related to AA in Arabic poetry. The concluding section of Chapter 7 demonstrated that our methodology was able to resolve key AA questions with a satisfactory level of confidence. For more detailed description, we will address each research question and how it was answered.

1. **What are the existing computational methods for addressing the problem of answering AA questions generally and particularly in Arabic poetry?**

   Through the literature review, we observe the evolution of AA methods over time. Initially, AA methodologies employed relatively simple statistical techniques, but they have progressively advanced to incorporate sophisticated deep learning approaches. Early methods used simple statistical algorithms to attribute authorship based on feature frequencies (Mosteller, 1964; Zhao, 2005). Distance measures, such as Euclidean and cosine similarity, were also popular (Ahmed, 2018; Bakly et al., 2020). Specialized measures like CNG were introduced for unbalanced documents (Selj, 2003).

Recent advancements favour machine learning (ML) techniques like k Nearest Neighbours (kNN) (Abbas et al., 2019), Naive Bayes (NB) (Koppel et al., 2013), and Support Vector Machines (SVM) (Stamatatos, 2017). Decision Trees and Random Forests are also used for their interpretability (Kumar et al., 2017). NNs including RNNs and CNNs, have gained prominence due to their ability to handle complex language tasks without extensive feature engineering (Hosseinia & Mukherjee, 2018; Hitschler et al., 2018). Pre-trained models like BERT achieve state-of-the-art results in AA (Barlas & Stamatatos, 2020). However, to date, no study has employed a BERT model for addressing the AA problem in Arabic texts. Consequently, this research represents the first attempt to utilize BERT for AA in Arabic poetry.

2. **What are the optimal data sources for this research, and what pre-processing steps are necessary to prepare the data for analysis?**

In this research, we aim to compile the largest possible dataset. Our search identified two comprehensive digital sources, Adab and AlDiwan, which were selected for data collection. These sources were merged, ensuring no duplication. Standard data cleaning procedures, such as noise removal, were applied. Additionally, labelling the dates of birth (DoB) and dates of death (DoD) was necessary to achieve accurate temporal segmentation. The final dataset comprises 784 poets, 77,850 poems, and a total word count of 6,609,495.

This dataset surpasses that of all known Arabic AA studies by a significant margin. Notably, this research employs the entire corpus of classical Arabic poetry, marking the first comprehensive study of its kind. The extensive volume of this dataset facilitates more robust experimentation using machine learning techniques, leading to more validated and accurate results.

3. **How does the topic of the poetry influence the performance and accuracy of the AA model?**

In the field of AA, it is observed that topical overlap can confound stylistic analysis, leading to misclassifications where the topic becomes the predominant factor rather than the author's unique style. To mitigate this issue, we applied topic modelling to our dataset and incorporated topic labelling. This approach allows us to gauge the extent to which the topic influences our AA classification.

Initially, we applied topic modelling using ETM, resulting in five primary topics. Upon building the AA model, we conducted the classification task with and without topic control. The results indicated an improvement in the model's performance when controlling for topic. However, the minimal difference between the topic-controlled and mixed cases suggests that topic confounding with style has a limited impact on our model. This implies minimal confounding between topics and authors in our dataset.

Notably, when applying our model to pre-Islamic poetry, we were unable to perform topic segregation due to the scarcity of data from this period. This limitation may introduce a small margin of misclassification in this subset, considering the earlier findings that controlling for topic generally yields higher accuracies.

4. **Which BERT model variant is most suitable for addressing AA questions in Arabic poetry, and what are its specific advantages?**

We opted to use BERT for this study due to its pretraining on the Arabic language, requiring only fine-tuning to optimize it for the AA task. Unlike other deep learning models, utilizing a pre-trained model such as BERT allows for effective handling of limited data while still achieving convincing results.

However, several BERT models have been trained on Arabic corpora. To determine the most suitable model for our task, we conducted a comparative analysis of three Arabic BERT models: AraBERT, CAMeLBERT, and multilingual BERT. The results demonstrated the superior performance of CAMeLBERT in handling our specific data and task requirements.

5. **What methodologies can be employed to evaluate and validate the performance of these computational models to ensure high accuracy and reliability,**

particularly in the context of pre-Islamic poetry given the scarcity of texts in this age?

Repeating the Experiment Multiple Times:

To ensure the robustness and reliability of our findings, we conducted the experiment multiple times. This iterative approach allowed us to mitigate the influence of random variations and stochastic processes inherent in machine learning models. By repeating the experiment, we could assess the consistency of our results and identify any anomalies or outliers. This practice enhances the validity of our conclusions and provides a more comprehensive understanding of the model's performance.

Repeating the Experiment Across Various Contexts and Historical Periods:

We extended the scope of our experimentation by applying the model to diverse cases and historical periods. This approach enabled us to evaluate the model's generalizability and adaptability to different contexts within Arabic poetry. By examining its performance across various temporal segments and poetic traditions, we aimed to ascertain the model's efficacy in accurately attributing authorship irrespective of the time period or stylistic variations present in the dataset. This comprehensive evaluation helps to establish the model's robustness across a wide range of conditions.

Comprehensive Testing Beyond the Initial Test Set:

In addition to evaluating the model on the test set derived from the initial data split, we further tested the model against additional examples. These examples were of comparable size to the training and testing data used during the model's development. This supplementary testing phase allowed us to assess the model's performance in a broader context and ensured that our findings were not confined to a specific subset of data. By validating the model against a diverse array of examples, we aimed to confirm its reliability and accuracy in real-world applications, thereby strengthening the credibility of our results.

6. **What is the impact of the quantity of training examples on the performance of the AA model?**

The quantity of training examples plays a critical role in the performance of AA models. In our study, we systematically varied the number of training examples to evaluate its impact on the model's effectiveness. Specifically, we experimented with datasets containing 40, 60, and 80 examples per author to identify the optimal quantity for training. This approach was particularly important given the scarcity of data from the pre-Islamic age, necessitating the evaluation of the model's performance with a lower number of examples.

Experimental Findings

Our findings revealed that using 60 training examples per author yielded the best results in terms of model performance. This was determined by assessing the accuracy and F1 score of the model, which were both maximized at this level of training data.

Analysis of Results

40 Training Examples: When the model was trained with 40 examples per author, the performance metrics indicated that the model struggled with underfitting. The limited quantity of data was insufficient to capture the complexities and stylistic nuances of each author's writing, leading to slightly lower F1 scores with average 0.98 and a wider stander deviation starts at 0.96. This scenario is particularly relevant for pre-Islamic poetry, where data scarcity is a significant issue.

60 Training Examples: Increasing the number of training examples to 60 significantly improved the model's performance. This amount of data provided a balanced trade-off, enabling the model to better learn the distinctive features of each author’s style without overwhelming it with excessive information. The model's accuracy and F1 score were highest at this level with average at 0.99, indicating an optimal fit. This suggests that even with limited data, such as in the pre-Islamic period, sufficient training examples can still yield reliable results.

80 Training Examples: While increasing the number of training examples to 80 initially seemed promising, it resulted in marginal declines compared to the 60-example scenario with F1 score average at 0.98. In some cases, it led to diminishing returns, where the additional data did not proportionally enhance the model’s performance. This suggests that beyond a certain point, adding more examples does not necessarily translate to better results and may even introduce noise, complicating the model's learning process.

Conclusion

The experiments underscore the importance of an optimal quantity of training data in the development of AA models. While too few examples can lead to underfitting, too many can introduce unnecessary complexity and noise. Our results indicate that 60 training examples per author strike the right balance, providing sufficient data for the model to accurately learn and attribute authorship. This finding is especially significant for periods like the pre-Islamic age, where data is scarce. It highlights the nuanced relationship between training data quantity and model performance, emphasizing the need for careful calibration to achieve optimal results even with limited examples.

7. **Does a one-vs-one classification approach yield better performance compared to a one-vs-all approach in the context of this research?**

In our study, we conducted experiments to determine the most effective approach for training the AA model. Specifically, we aimed to assess whether a one-vs-one or a one-vs-many classification strategy would yield better results. This experimentation involved varying the number of poets in the opposite class during training.

Experimental Setup

- One-vs-One: In this setup, the model was trained to differentiate between one poet and another poet. This approach creates a binary classification problem, where the model learns to distinguish between two distinct writing styles.

- One-vs-Two: In this variation, the model was trained to differentiate between one poet and two poets in the opposite class. This introduces additional complexity, as the model must discern the target poet's style against a more varied background of two other authors.

- One-vs-Three: Similarly, this setup involved training the model to differentiate between one poet and three poets in the opposite class, further increasing the complexity of the classification task.

Results and Analysis

The performance of the model was evaluated using accuracy and F1 score across these different setups. The results indicated that the one-vs-one classification strategy consistently provided the best results, while the one-vs-three strategy yielded the worst outcomes. The F1 scores are 0.99, 0.97 and 0.94 respectively.

- One-vs-One Performance: This approach yielded the highest accuracy and F1 scores, demonstrating that the model was most effective when trained to distinguish between two individual authors. The simplicity of the binary classification allowed the model to focus on capturing the nuanced stylistic features of each poet, leading to more precise AA.

- One-vs-Two Performance: When the model was trained to differentiate between one poet and two poets in the opposite class, the performance metrics showed a decline compared to the one-vs-one setup. The increased variability in the training data made it more challenging for the model to accurately identify the distinctive features of the target poet, resulting in lower accuracy and F1 scores.

- One-vs-Three Performance: The one-vs-three setup resulted in the worst performance among the tested strategies. The added complexity of distinguishing one poet from three others led to significant confusion in the model. The overlapping styles and greater variability in the training data exacerbated the model's difficulty in isolating the unique characteristics of the target poet, leading to the lowest accuracy and F1 scores observed.

Voting Technique

To address the performance issues associated with the one-vs-many approaches, we implemented a voting technique. Instead of training the model with all the other poets at once in the opposite class, we decided to use multiple one-vs-one classifiers. Each one-vs-one classifier was trained to distinguish between the target poet and one other poet at a time. The final decision was then made by aggregating the outputs of these individual classifiers using a voting mechanism.

Conclusion

The experimentation with varying the number of authors in the classification strategy revealed that a one-vs-one approach is the most effective for AA in this context. The binary classification setup enables the model to better capture and differentiate the unique stylistic characteristics of individual poets, resulting in higher accuracy and reliability in AA. The implementation of the voting technique further enhanced the model's performance, as it allowed us to combine the strengths of multiple one-vs-one classifiers, avoiding the pitfalls of increased complexity seen in the one-vs-many setups. This finding underscores the importance of simplifying the classification task and using ensemble methods to enhance model performance, particularly in the nuanced field of AA.

8. How does the performance of the proposed model compare to traditional AA methods, such as Burrow's Delta?

In the context of evaluating our BERT model for the AA task, it was imperative to compare its performance against a well-established traditional method. For this purpose, we selected Burrow's Delta as the baseline. Burrow's Delta has long been recognized in the field of stylometry and AA for its simplicity and effectiveness in distinguishing between different authors based on stylistic features.

Burrow's Delta Methodology

Burrow's Delta operates by measuring the stylistic distance between texts. It calculates the difference in word frequency distributions between a disputed text and texts known to be authored by different candidates. This method relies on selecting a set of common function words and computing the standardized differences (deltas) in their frequencies. The author with the smallest delta is considered the most likely author of the disputed text.

Results

The results of our experiments demonstrated that the BERT model significantly outperformed Burrow's Delta across all evaluation metrics. Specifically, the BERT model exhibited higher accuracy and F1 scores, indicating a more reliable and precise identification of authorship. This performance improvement can be attributed to several factors inherent in the BERT architecture:

- Pre-trained Knowledge: BERT leverages extensive pre-training on large corpora, capturing a wide range of linguistic patterns and nuances. This pre-trained knowledge enables the model to perform well even with limited task-specific data.

- Contextual Understanding: Unlike traditional methods like Burrow's Delta, which rely solely on word frequency distributions, BERT captures the contextual relationships between words. This allows the model to better understand stylistic nuances that are indicative of an author's unique writing style.

- Fine-Tuning Capability: The ability to fine-tune BERT on the specific AA task ensures that the model adapts to the characteristics of the dataset, further enhancing its performance.

Conclusion

By surpassing the performance of Burrow's Delta by a significant margin, the BERT model demonstrates its efficacy and robustness in the AA task. This comparison not only validates the advantages of modern deep learning approaches over traditional methods but also provides a compelling case for the adoption of pre-trained language models in stylometric analysis and AA research.

9. Does the computational model effectively answer AA questions specifically in the context of pre-Islamic Arabic poetry?

The evaluation of the computational model's effectiveness in answering AA questions, particularly in the context of pre-Islamic Arabic poetry, involved a thorough testing process. Given the unique challenges posed by this corpus, including the scarcity of texts and the nuanced stylistic features of the poetry, it was crucial to validate the model against established scholarly doubts.

Testing on Highly Disputed Texts

To assess the model's accuracy and relevance, we first focused on texts that have long been subject to scholarly debate regarding their authorship. These texts are highly scrutinized within the academic community, with experts harbouring significant doubts about their attributed authors. This provided a robust benchmark for testing the model's ability to discern true authorship.

Model's Performance

The model's results were remarkably consistent with the scholarly consensus on these disputed texts. Specifically, for texts where scholars were confident of misattribution, the model similarly indicated uncertainty and pointed to the same potential issues in authorship. This alignment between the model's outputs and expert opinions is a strong indicator of its reliability and accuracy in this context.

- **Uncertainty Detection:** The model's ability to detect and reflect uncertainty in AA is a critical feature, especially for pre-Islamic poetry where many texts have ambiguous or contested origins. By flagging these texts, the model not only corroborates scholarly assessments but also provides a tool for further investigation.

- **Consistency with Scholarly Doubts:** The consistency between the model's predictions and scholarly doubts about certain texts underscores its effectiveness. It validates the model's capacity to capture the intricate stylistic and linguistic features that are pivotal in AA for this specific corpus.

Conclusion

The computational model has demonstrated a high degree of effectiveness in answering AA questions within the realm of pre-Islamic Arabic poetry. By aligning with scholarly assessments on highly disputed texts, the model has proven its capability to identify and reflect uncertainties in authorship. This not only enhances its credibility but also positions it as a valuable tool for researchers and scholars working in the field of classical Arabic literature.

The results highlight the model's potential to assist in resolving longstanding debates and contribute to a more accurate understanding of pre-Islamic poetic authorship. However, the scope of this study did not encompass all aspects of AA in Arabic poetry, indicating potential areas for future research and expansion. By addressing the remaining gaps, subsequent studies could further validate the effectiveness of computational approaches in answering complex literary questions within the realm of Arabic poetry.

4. Contributions to the Field

This research makes a significant contribution to the field of Arabic AA by pioneering the use of transformers, leading to improved accuracy in attribution tasks. Additionally, it advances the broader field of Arabic literary studies by integrating computational methods to address and resolve complex issues, offering a modern approach to challenges traditionally tackled through qualitative means.

5. Limitations of the Study

This study employs a highly complex form of deep learning that is challenging to interpret. While we implemented indirect methods to elucidate the model's decision-making processes, we still cannot fully unravel the inner workings of the model or identify which specific features played a key role in assigning the correct classes.

A prominent characteristic of poetry is its metre, yet this aspect was not incorporated into the research due to the challenges and time constraints associated with extracting metric features.

6. Suggestions for Future Research

The research addresses a selection of AA issues in Arabic poetry, but the body of Arabic literary studies contains numerous unresolved cases. A more in-depth examination of the model's decision-making process is required to offer comprehensive linguistic explanations and bolster the validation of the model's outcomes. As a future direction for this thesis on AA, interpreting model results warrants further investigation. Notably, a relevant technique for this purpose is outlined in a paper that presents "a framework for interpreting the results of automatic genre classification using linguistic features" (Sharoff , 2021) . This approach could potentially be adapted to enhance the interpretability of AA models.

A promising direction for future research involves examining the possibility of co-authorship in pre-Islamic poetry, reflecting the collaborative and communal nature of its creation. The oral-formulaic theory, proposed by Milman Parry (1928) and Albert Lord (1960), highlights how oral traditions often rely on shared formulas and iterative processes, challenging the notion of singular authorship. By applying this perspective, future studies could investigate how multiple contributors shaped poetic works, either through collective composition or successive modifications. Integrating insights from oral-formulaic studies with computational methods could provide a more nuanced understanding of authorship in oral traditions, offering new tools to analyze the interplay between collective creativity and individual expression in early Arabic poetry. Additionally, exploring the overlap between oral and written modes of composition could further illuminate the dynamics of poetic transmission and transformation across historical contexts.

Additionally, incorporating metric features by developing a model trained on all metre rules would allow for the transformation of text into its abstract metre form, providing a more nuanced analysis of Arabic poetry and a new aspect to consider. Achieving this requires precise diacritization for each letter, which is currently unavailable in any existing poetry data sources. Undertaking this task from scratch would demand a significant investment of time and effort.

7. Conclusion and Final Remarks

The primary objective of this research was to determine whether computational methods can effectively address literary questions, particularly those related to AA in Arabic poetry. The concluding section of Chapter 7 demonstrated that our methodology was able to resolve key AA questions with a satisfactory level of confidence.

Existing Computational Methods for AA in Arabic Poetry

Through our literature review, we traced the evolution of AA methods from simple statistical techniques to sophisticated deep learning approaches. Early methods relied on feature frequencies and distance measures, such as Euclidean and cosine similarity. More recent advancements have favoured machine learning (ML) techniques, including k-Nearest Neighbours (kNN), Naive Bayes (NB), Support Vector Machines (SVM), Decision Trees, Random Forests, and Neural Networks (NNs). Pre-trained models like BERT have achieved state-of-the-art results in AA. This research represents the first attempt to employ BERT for AA in Arabic poetry.

Optimal Data Sources and Pre-processing Steps

We compiled the largest known dataset for this research by merging comprehensive digital sources, Adab and AlDiwan, and ensuring no duplication. Standard data cleaning procedures were applied, and temporal segmentation was achieved by labelling dates of birth and death. The final dataset comprises 784 poets, 77,850 poems, and a total word count of 6,609,495, surpassing all known Arabic AA studies in scope.

Influence of Poetry Topics on AA Model Performance

Topical overlap can confound stylistic analysis, leading to misclassifications. To address this, we applied topic modelling using Embedded Topic Modelling (ETM), resulting in five primary topics. Our results indicated an improvement in model performance when controlling for topic, although the minimal difference between the topic-controlled and mixed cases suggests limited confounding between topics and authors in our dataset.

Suitability of BERT Model Variants

We compared three Arabic BERT models: AraBERT, CAMeLBERT, and multilingual BERT. CAMeLBERT demonstrated superior performance in handling our specific data and task requirements, making it the most suitable for our AA task.

Evaluation and Validation Methodologies

To ensure high accuracy and reliability, we conducted multiple iterations of the experiment, applied the model to diverse historical periods, and tested it against additional examples. This comprehensive approach confirmed the robustness and consistency of our findings.

Metrics for Assessing Model Performance

Accuracy and F1 score were selected as the primary metrics. Accuracy provides a high-level overview of model performance, while the F1 score offers insights into the precision-recall trade-off, crucial for AA tasks.

Impact of Training Data Quantity

We experimented with datasets containing 40, 60, and 80 examples per author. Using 60 examples yielded the best results, striking a balance between capturing stylistic nuances and avoiding excessive complexity. This finding is particularly relevant for periods like the pre-Islamic age, where data is scarce.

Classification Strategies

A one-vs-one classification approach consistently provided the best results compared to one-vs-many strategies. This simplified approach allowed the model to focus on distinguishing between two authors, enhancing accuracy and reliability. Implementing a voting technique further improved performance by combining multiple one-vs-one classifiers.

Comparison to Traditional AA Methods

The BERT model significantly outperformed Burrow's Delta, a well-established traditional method. BERT's pre-trained knowledge, contextual understanding, and fine-tuning capability contributed to its superior performance.

Effectiveness in Pre-Islamic Arabic Poetry

The model's results were consistent with scholarly assessments of highly disputed pre-Islamic texts, indicating its reliability and accuracy in this context. The ability to detect and reflect uncertainty in AA further validated the model's effectiveness.

Final Remarks

This research has demonstrated the potential of computational methods, particularly pre-trained language models like BERT, in addressing AA questions in Arabic poetry. While the study has achieved significant milestones, it also highlighted areas for future research. Expanding the scope to encompass all aspects of AA in Arabic poetry and addressing remaining gaps will further validate and enhance the effectiveness of computational approaches in this field. The findings underscore the importance of combining traditional literary scholarship with modern computational techniques to achieve more accurate and nuanced understandings of authorship in classical Arabic literature.

References:

Abbas, M., Lichouri, M., & Zeggada, A. (2019). Classification of Arabic Poems: From the 5th to the 15th Century. Springer International Publishing *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): Vol. 11808 LNCS*. https://doi.org/10.1007/978-3-030-30754-7_18

Abbasi, A., & Chen, H. (2005). Applying authorship analysis to extremist-group web forum messages. *IEEE Intelligent Systems*, 20(5), 67-75.

Abū al-Faraj al-Aṣfahānī (1927). *Kitāb al-Aghānī*. Cairo: Dār al-Kutub al-Miṣriyyah.

Abuhaiba, I. S., & Eltibi, M. F. (2016). Author attribution of Arabic texts using extended probabilistic context free grammar language model. *International Journal of Intelligent Systems and Applications*, *8*(6), 27.

Ahmed, A. F., Mohamed, R., & Mostafa, B. (2017). Machine learning for authorship attribution in Arabic poetry. *Int. J. Future Comput. Commun*, 6(2), 42-46.

Ahmed, H. (2017). Dynamic similarity threshold in authorship verification: evidence from classical Arabic. *Procedia Computer Science*, *117*, 145-152.

Ahmed, H. (2018). The role of linguistic feature categories in authorship verification. *Procedia Computer Science*, *142*, 214–221. https://doi.org/10.1016/j.procs.2018.10.478

Ahmed, H. (2019). Distance-based authorship verification across modern standard Arabic genres. In *Proceedings of the 3rd Workshop on Arabic Corpus Linguistics*, 89-96.

Ahmed, H. (2019). Sample size in Arabic authorship verification. In *Proceedings of the 3rd International Conference on Natural Language and Speech Processing*, 84-91.

Al-Abraṣ, ʻUbayd Ibn Al-Abraṣ Ibn ʻAwf. (1958). *Dīwān of ʻUbayd Ibn Al-Abraṣ.* Beirut: Dār Sader.

Al-Asad, N. A.-D. (1988). *Maṣādir al-shiʻr al-Jāhilī.* Beirut: Dār al-Jīl.

Al-Ayyoub, M., Jararweh, Y., Rabab'ah, A., & Aldwairi, M. (2017). Feature extraction and selection for Arabic tweets authorship authentication. *Journal of Ambient Intelligence and Humanized Computing*, *8*, 383-393.

Al-Bāqillānī, Abū Bakr Muhammad ibn al-Ṭayyib. (1951). *Iʻjāz al-Qur'ān* (Aḥmad Shalabī, Ed). Cairo: Dār al-Maʿārif.

Al-Harbi, S., Almuhareb, A., Al-Thubaity, A., Khorsheed, M. S., & Al-Rajeh, A. (2008). Automatic Arabic text classification. *Text*, *August*, 77–84. http://eprints.ecs.soton.ac.uk/22254/

Al-Ḥārith ibn Ḥulzah. (1991). *Dīwān of Al-Ḥārith ibn Ḥulzah* (M. Al-ʻAṭīyah, Ed.). Amman: Dār Al-Imām Al-Nawawi for Publishing.

Al-Jurjānī, Abd al-Qahir. (1960). *Dalā'il al-i'jāz*. (Muhammad Abd al-Mun'im Khafaji, Ed). Cairo: Dār al- Ma'arif.

Al-Marzūqī. (1991). *Sharḥ al-Ḥamāsah.* Beirut: Dār Al-Jīl.

Al-Nabigha al-Dhubyani. (2016). *Dīwān Al-Nābighah Al-Dhubyānī.* Beirut: Dār al-Kutub al-'Ilmīyah.

Al-Sarem, M., & Emara, A. H. (2019). The effect of training set size in authorship attribution: Application on short Arabic texts. *International Journal of Electrical and Computer Engineering*, *9*(1), 652.

Al-Sarem, M., Alsaeedi, A., & Saeed, F. (2020). A deep learning-based artificial neural network method for instance-based Arabic language authorship attribution. *Int. J. Advance Soft Compu. Appl*, *12*(2), 1–15.

Al-Sarem, M., Cherif, W., Wahab, A. A., Emara, A. H., & Kissi, M. (2018). Combination of stylo-based features and frequency-based features for identifying the author of short Arabic texts. *Proceedings of the 12th International Conference on Intelligent Systems: Theories and Applications*, 1-6.

Al-Sarem, M., Saeed, F., Alsaeedi, A., Boulila, W., & Al-Hadhrami, T. (2020). Ensemble methods for instance-based Arabic language authorship attribution. *IEEE Access*, *8*, 17331-17345.

Alanazi, S. (2015). Classical Arabic authorship attribution using simple features. *Natural Language Processing and Cognitive Science*, *4551*.

Albadarneh, J., Talafha, B., Al-Ayyoub, M., Zaqaibeh, B., Al-Smadi, M., Jararweh, Y., & Benkhelifa, E. (2015, December). Using big data analytics for authorship authentication of arabic tweets. In *2015 IEEE/ACM 8th International Conference on Utility and Cloud Computing (UCC)*, 448-452.

Alqahtani, F., & Dohler, M. (2023). Survey of authorship identification tasks on Arabic texts. *ACM Transactions on Asian and Low-Resource Language Information Processing*, 22(4), 1-24.

Altakrori, M. H., Iqbal, F., Fung, B. C., Ding, S. H., & Tubaishat, A. (2018). Arabic authorship attribution: An extensive study on twitter posts. *ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP)*, *18*(1), 1-51.

Altheneyan, A. S., & Menai, M. E. B. (2014). Naïve Bayes classifiers for authorship attribution of Arabic texts. *Journal of King Saud University-Computer and Information Sciences*, *26*(4), 473-484.

Alwajeeh, A., Al-Ayyoub, M., & Hmeidi, I. (2014, April). On authorship authentication of Arabic articles. In *2014 5th International Conference on Information and Communication Systems (ICICS)*, 1-6.

‘Amr ibn Kulthūm. (1992). *Dīwān of ‘Amr ibn Kulthūm al-Taghlibī* (A. M. A. Maidan, Ed.). Jeddah: Jeddah Cultural Club.

‘Antarah ibn Shaddād ibn ‘Amr ibn Mu‘āwiyah al-‘Absī. (1958). *Dīwān ‘Antarah al-‘Absī.* Beirut, Lebanon: Dār Sader.

Antoun, W., Baly, F., & Hajj, H. (2020). Arabert: Transformer-based model for Arabic language understanding. *Proceedings of the Twelfth International Conference on Language Resources and Evaluation (LREC 2020), Marseille, France*. arXiv preprint arXiv:2003.00104.

Argamon, S., Dhawle, S., Koppel, M., & Pennebaker, J. W. (2005, June). Lexical predictors of personality type. *Proceedings of the 2005 Joint Annual Meeting of the Interface and the Classification Society of North America*, 1-16.

Argamon, S., Koppel, M., Fine, J., & Shimoni, A. R. (2003). Gender, genre, and writing style in formal written texts. *Text & Talk*, 23(3), 321-346.

Baayen, H., Van Halteren, H., & Tweedie, F. (1996). Outside the cave of shadows: Using syntactic annotation to enhance authorship attribution. *Literary and Linguistic Computing*, 11(3), 121-132.

Badirli, S., Borgo Ton, M., Gungor, A., & Dundar, M. (2021). Open set authorship attribution toward demystifying Victorian periodicals. In *Document Analysis and Recognition–ICDAR 2021: 16th International Conference, Lausanne, Switzerland, September 5–10, 2021, Proceedings, Part IV 16*, 221-235.

Bakly, A. H. Baraka, R., Salem, S., Hussien, M. A., Nayef, N., & Shaban, W. A. (2014). Arabic text author identification using support vector machines. *Journal of Advanced Computer Science and Technology Research*, *4*(1), 1-11.

Barlas, G., & Stamatatos, E. (2020). Cross-domain authorship attribution using pre-trained language models. In *IFIP Advances in Information and Communication Technology: Vol. 583 IFIP*. https://doi.org/10.1007/978-3-030- 49161-1_22

Benjamin, V., Chung, W., Abbasi, A., Chuang, J., Larson, C. A., & Chen, H. (2013, June). Evaluating text visualization: An experiment in authorship analysis. In *2013 IEEE International Conference on Intelligence and Security Informatics*, 16-20.

Benzebouchi, N. E., Azizi, N., Hammami, N. E., Schwab, D., Khelaifia, M. C. E., & Aldwairi, M. (2019). Authors’ writing styles based authorship identification system using the text representation vector. *16th International Multi-Conference on Systems, Signals and Devices, SSD 2019*, 371–376. https://doi.org/10.1109/SSD.2019.8894872

Berkani, A., Holzer, A., & Stoffel, K. (2020). Pattern matching in meter detection of Arabic classical poetry. *Proceedings of IEEE/ACS International Conference on Computer Systems and Applications, AICCSA*, *2020-Novem*, 1-8. https://doi.org/10.1109/AICCSA50499.2020.9316497

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

Britannica, E. (1993). Encyclopædia Britannica. Chicago: *University of Chicago Press*.

Burrows, J. F. (1987). Word-patterns and story-shapes: The statistical analysis of narrative style. *Literary & Linguistic Computing*, 2(2), 61-70.

Clement, R., & Sharp, D. (2003). Ngram and Bayesian classification of documents for topic and authorship. *Literary and linguistic computing*, 18(4), 423-447.

Devlin, J. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.

Dieng, A. B., Ruiz, F. J., & Blei, D. M. (2020). Topic modeling in embedding spaces. *Transactions of the Association for Computational Linguistics*, *8*, 439-453.

El Bakly, A. H., Darwish, N. R., & Hefny, H. A. (2020). Using ontology for revealing authorship attribution of Arabic text. *Int. J. Eng. Adv. Technol.(IJEAT)*, *4*, 143-151.

Fabien, M., Villatoro-Tello, E., Motlicek, P., & Parida, S. (2020). BertAA: BERT fine-tuning for authorship attribution. In *Proceedings of the 17th International Conference on Natural Language Processing (ICON)*, 127-137.

Peng, F., Schuurmans, D., Keselj, V., & Wang, S. (2003). Language independent authorship attribution using character-level language models*. Proceedings of the 10th Conference on European Chapter of the Association for Computational Linguistics*.

Gamon, M. (2004). Linguistic correlates of style: authorship classification with deep linguistic analysis features. In Coling 2004: *Proceedings of the 20th International Conference on Computational Linguistics*, 611-617.

Gamon, M. (2004). Sentiment classification on customer feedback data: noisy data, large feature vectors, and the role of linguistic analysis. *COLING 2004: Proceedings of the 20th international conference on computational linguistics*, 841-847.

Goncalves, L., & Vimieiro, R. (2021). Approaching authorship attribution as a multi-view supervised learning task. *Proceedings of the International Joint Conference on Neural Networks*, *2021-July*. https://doi.org/10.1109/IJCNN52387.2021.9533360

Halvani, O., Graner, L., & Regev, R. (2020). A step towards interpretable authorship verification. *The 15th International Conference on Availability, Reliability and Security (ARES 2020)*. 1–21. http://arxiv.org/abs/2006.12418

Hitschler, J., van den Berg, E., & Rehbein, I. (2018). Authorship attribution with Convolutional Neural Networks and POS-Eliding. *In Proceedings of the Workshop on Stylistic Variation*, 53-58.

Hosseinia, M., & Mukherjee, A. (2018). *Experiments with Neural Networks for Small and Large Scale Authorship Verification*. http://arxiv.org/abs/1803.06456

Howedi, F., & Mohd, M. (2014). Text classification for authorship attribution using Naive Bayes classifier with limited training data. *Computer Engineering and Intelligent Systems*, *5*(4), 48-56.

Ibn Khaldūn, 'Abd al-Raḥmān ibn Muḥammad. (2001). *Al-Muqaddimah*. Damascus: Dār al-Fikr.

Ibn Rabī ʿah.(1880).*Dīwān of Labid Al- ʿAmirī* (T. al-Tusi, Ed.). Germany: Adolph Helsehusen Press.

Ibn Sallām al-Jumaḥī, Muḥammad ibn Sallām (2003*). Ṭabaqāt fuḥūl al-shu'arā'*. Cairo: Maṭba'at Dār al-Kutub wa-al-Wathā'iq al-Qawmīyah.

Imru' al-Qais. (2014). *Dīwān Imru' al-Qais* (M. Abu al-Fadl Ibrahim, Ed.). Cairo, Egypt: Dār al-Ma'arif.

Inoue, G., Khalifa, S., & Habash, N. (2021). Morphosyntactic tagging with pre-trained language models for Arabic and its dialects. *arXiv preprint arXiv:2110.06852*.

Keselj, V., Peng, F., Cercone, N., & Thomas, C. (2003). N-gram-based author profiles for authorship attribution. *Proceedings of the Pacific Association for Computational Linguistics*. 255- 264.

Kestemont, M. (2014, April). Function words in authorship attribution. From black magic to theory?. In *Proceedings of the 3rd Workshop on Computational Linguistics for Literature (CLFL)*, 59-66.

Khalil, H. M., Ahmed, T., & El-Shistawy, T. A. R. E. K. (2020). Authorship authentication of political Arabic articles based on modified TF-IGF algorithm. *Journal of Theoretical and Applied Information Technology*, *98*(17).

Kilgarriff, A. (2005). Language is never, ever, ever, random. *Corpus Linguistics and Linguistic Theory. 1*(2), 263-276. https://doi.org/10.1515/cllt.2005.1.2.263

Kocher, M. &. (2017). A simple and efficient algorithm for authorship verification. *Journal of the Association for Information Science and Technology*, *68*(1), 259-269

Koppel, M. &. (2007). Measuring differentiability: unmasking pseudonymous authors. *Journal of Machine Learning Research*, 8(6).

Koppel, M., & Schler, J. (2004). Authorship verification as a one-class classification problem. *Proceedings, Twenty-First International Conference on Machine Learning, ICML 2004*, 489–495. https://doi.org/10.1145/1015330.1015448

Koppel, M., & Winter, Y. (2014). Determining if two documents are written by the same author. *Journal of the American Society for Information Science and Technology*, *65*(1), 178–187. https://doi.org/10.1002/asi.22954

Koppel, M., Schler, J., & Argamon, S. (2013). Authorship attribution: what's easy and what's hard? *SSRN Electronic Journal*, *21*(2). https://doi.org/10.2139/ssrn.2274891

Koppel, M., Schler, J., Argamon, S., & Winter, Y. (2012). The "fundamental problem" of authorship attribution. *English Studies*, *93*(3), 284–291. https://doi.org/10.1080/0013838X.2012.668794

Kumar, P. J., Reddy, G. S., & Reddy, T. R. (2017). Document weighted approach for authorship attribution. *Int. J. Comput. Intell. Res*, *13*(7), 1653-1661.

Kumar, S., & Chaurasia, M. A. (2012). Assessment on stylometry for multilingual manuscript. *IOSR Journal of Engineering*, *2*(9), 1-6.
language understanding. arXiv preprint arXiv:2003.00104.

Leech, Geoffrey, and Michael H. Short. (2007). *Style in Fiction: A Linguistic Introduction to English Fictional Prose*. 2nd ed., United Kingdom: Pearson Longman.

Lord, A. B. (1960). *The Singer of Tales*. Harvard University Press.

Madigan, D. G. (2005). Author identification on the large scale. *In Proceedings of the 2005 Meeting of the Classification Society of North America* (CSNA).

Maimūn ibn Qays. (2012). *Dīwān of Al-Aʿshā al-Kabīr* (M. Ḥusayn, Ed.). Cairo: Maktabat Al-Adab for Printing, Publishing, and Distribution.

Markov, I., Stamatatos, E., & Sidorov, G. (2018). Improving cross-topic authorship attribution: The role of pre-processing. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): Vol. 10762 LNCS* (Issue Cic). Springer International Publishing. https://doi.org/10.1007/978-3-319-77116-8_21

Maurya, R. K., Saxena, M. R., & Akhil, N. (2016). Intelligent Systems Technologies and Applications. *Advances in Intelligent Systems and Computing*, *384*(January 2016), 247- 257. https://doi.org/10.1007/978-3-319-23036-8

Mechti, S., & Almansour, F. (2021). An Orderly Survey on Author Attribution Methods: From Stylistic Features to Machine Learning Models. *Int. J. Adv. Res. Eng. Technol*, *12*, 528-538.

Monroe, J. (1972). Oral composition in pre-Islamic poetry. *Journal of Arabic Literature, 3*, 1–53.

Mosteller, F., & Wallace, D. L. (1964). Inference in an authorship problem: A comparative study of discrimination methods applied to the authorship of the disputed Federalist Papers. *Journal of the American Statistical Association*, 58(302), 275-309.

Nasser Alsager, H. (2021). Towards a stylometric authorship recognition model for the social media texts in Arabic. *Arab World English Journal (AWEJ) Volume*, *11*.

Omar, A., & Hamouda, W. I. (2020). The effectiveness of stemming in the stylometric Authorship Attribution in Arabic. *International Journal of Advanced Computer Science and Applications*, *11*(1).

Omer, A. I. (2017). Arud, the metrical system of Arabic poetry, as a feature set for Authorship Attribution. IEEE/ACS 14th International Conference on Computer Systems and Applications, 431-436.

Otoom, A. F., Abdullah, E. E., Jaafer, S., Hamdallh, A., & Amer, D. (2014). Towards author identification of Arabic text articles. In *2014 5th International conference on information and communication systems (ICICS)*, 1-4.

Ouamour, S., & Sayoud, H. (2013). Authorship attribution of short historical Arabic texts based on lexical features. In *2013 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery*, 144-147.

Ouamour, S., & Sayoud, H. (2018). A Comparative survey of authorship attribution on Short Arabic Texts. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): Vol. 11096 LNAI*. Springer International Publishing. https://doi.org/10.1007/978-3-319-99579-3_50

Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, 29(1), 24-54.

Parry, M. (1928). L'Épithète traditionnelle dans Homère: Essai sur un problème de style homérique. *Collection d'études anciennes*. Société Éditrice «Les Belles Lettres».

Posadas-Durán, J. P., Gómez-Adorno, H., Sidorov, G., Batyrshin, I., Pinto, D., & Chanona-Hernández, L. (2017). Application of the distributed document representation in the authorship attribution task for small corpora. *Soft Computing*, *21*(3), 627–639. https://doi.org/10.1007/s00500-016-2446-x

Potha, N., & Stamatatos, E. (2014). A profile-based method for authorship verification. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *8445 LNCS*, 313–326. https://doi.org/10.1007/978-3-319-07064-3_25

Ramezani, R. (2021). A language-independent authorship attribution approach for author identification of text documents. *Expert Systems with Applications*, *180*(May 2020), 115139. https://doi.org/10.1016/j.eswa.2021.115139

Ramnial, H., Panchoo, S., & Pudaruth, S. (2016). Authorship attribution using stylometry and machine learning techniques. In *Intelligent Systems Technologies and Applications: Volume 1*, 113-125.

Rocha, M. A. D., Nóbrega, G. Â. S. D., De Medeiros Valentim, R. A., & Alves, L. P. C. F. (2020). A text as unique as fingerprint: AVASUS text analysis and authorship recognition. *PervasiveHealth: Pervasive Computing Technologies for Healthcare*, *Part F1667*. https://doi.org/10.1145/3401895.3401935

Rogers, A., Kovaleva, O., & Rumshisky, A. (2021). A primer in BERTology: What we know about how BERT works. *Transactions of the Association for Computational Linguistics*, *8*, 842-866.

Sapkota, U., Bethard, S., Montes-Y-Gómez, M., & Solorio, T. (2015). Not all character N-grams are created equal: A study in authorship attribution. *NAACL HLT 2015 - 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Proceedings of the Conference*, 93–102. https://doi.org/10.3115/v1/n15-1010

Sari, Y., Stevenson, M., & Vlachos, A. (2018). Topic or style? exploring the most useful features for authorship attribution. *27th International Conference on Computational Lingustics*, 343–353. https://www.aclweb.org/anthology/C18-1029

Shaker K. (2012). *Investigating features and techniques for Arabic authorship attribution.* (Doctoral dissertation, Heriot-Watt University).

Shakir, M. (1994). *Namaṭ Ṣa'b wa-namaṭ Mukhīf.* Jeddah: Dār al-madani.

Sharoff, S. (2021). Genre annotation for the web: text-external and text-internal perspe**ct**ives. *Register Studies*, 3(1), 1-32.

Sowayan, S. A. (1985). *Nabati Poetry: The Oral Poetry of Arabia*. University of California Press.

Argamon, S., Whitelaw, C., Chase, P., Hota, S. R., Garg, N., & Levitan, S. (2007). Stylistic text classification using functional lexical features. *Journal of the American Society for Information Science and Technology*, 58(6), 802-822.

Shrestha, P., Sierra, S., González, F. A., Rosso, P., Montes-Y-Gómez, M., & Solorio, T. (2017). Convolutional neural networks for authorship attribution of short texts. *15th Conference of the European Chapter of the Association for Computational Linguistics, EACL 2017 - Proceedings of Conference*, *2*, 669–674. https://doi.org/10.18653/v1/e17- 2106

Stamatatos, E. (2006). Intrinsic plagiarism detection using character. *Proceedings of 3rd PAN Workshop. Uncovering Plagiarism, Authorship and Social Software Misuse*, 38-46.

Stamatatos, E. (2009). A survey of modern authorship attribution methods. *Journal of the American Society for Information Science and Technology*, *60*(3), 538–556. https://doi.org/10.1002/asi.21001

Stamatatos, E. (2017). Authorship attribution using text distortion. *15th Conference of the European Chapter of the Association for Computational Linguistics, EACL 2017 - Proceedings of Conference*, *1*, 1138–1149. https://doi.org/10.18653/v1/e17-1107

Sundararajan, K., & Woodard, D. L. (2018). What constitutes " style " in authorship attribution? *27th International Conference on Computational Linguistics*, 2814–2822.

Stamatatos, E., Fakotakis, N., & Kokkinakis, G. (2001). Computer-based authorship attribution without lexical measures. *Computers and the Humanities*, 35(2), 193-214.

Houvardas, J., & Stamatatos, E. (2006, September). N-gram feature selection for authorship identification. *International Conference on Artificial Intelligence: Methodology, Systems, and Applications* (pp. 77-86). Berlin, Heidelberg: Springer Berlin Heidelberg.

Ṭarafah ibn Al-‘Abd. (2016) Dīwān Ṭarafah ibn al-‘Abd. Beirut: Dār al-Arqām ibn Abī al-Arqām.

Tyo, J., Dhingra, B., & Lipton, Z. (2021). Siamese BERT for authorship verification. *CEUR Workshop Proceedings*, *2936*, 2169–2177.

Wang, X., & Iwaihara, M. (2021). Integrating RoBERTa Fine-Tuning and User Writing Styles for authorship attribution of Short Texts. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): Vol. 12858 LNCS*. Springer International Publishing. https://doi.org/10.1007/978-3-030-85896-4_32

Zhao, Y., & Zobel, J. (2005, October). Effective and scalable authorship attribution using function words. In *Asia Information Retrieval Symposium*, 174-189. Berlin, Heidelberg: Springer Berlin Heidelberg.

Zuhayr ibn Abī Sulma. (2008). *Dīwān of Zuhayr ibn Abī Sulma*. Dār Al Kotob Al Ilmiyah.

Zwettler, M. (1978). *The Oral Tradition of Classical Arabic Poetry: Its Character and Implications*. Ohio State University Press.

Appendix

Century	Nature	Love	Faith	Honor	Praise
6th	كالبيض في الروض المنور قد أفضى إليه إلى الكثيب فغر	يا أيها الراكب الغادي لطيته عرج أبثك عن بعض الذي أجد ما عالج الناس من وجد تضمنهم إلا ووجدي به فوق الذي أجد حسبي رضاه وأني في مسرته ووده آخر الأيام أجتهد	أنت الرسول رسول الله نعلمه عليك نزل من ذي العزة الكتب	سئمت وأمسيت رهن الفرا ش من جرم قومي ومن مغرم ومن سفه الرأي بعد النهى وعيب الرشاد ولم يفهم فلو أني قومي أطاعوا الحليم لم يتعدوا ولم يظلم ولكن قومي أطاعوا الغوا ة حتى تعكس أهل الدم فأودى السفيه برأي الحل يم وانتشر الأمر لم يبرم	غلبت ملوك الأرض بالحزم والنهى فأنت امرؤ في سورة المجد ترتقي وأنجب به من آل نصر سميدع أغر كلون الهندواني رونق
7th	ونار كسحر العود رفع ضوؤها مع الصبح هبات الرياح الزعازع	ألا أيها القلب الذي لج هائما بليلى وليدا لم تقطع تمائمه أفق قد أفاق العاشقون وقد أنى لك اليوم أن تلقى طبيبا تلائمه فما لك مسلوب العزاء كأنما ترى نأي ليلى مغرما أنت غارمه أجدك لا تنسيك ليلى ملمة تلم ولا عهد يطول تقادمه	كفى حزنا أنا عصينا إمامنا عليا وأن القوم طاعوا معاويه وأن لأهل الشام في ذاك فضلهم علينا بما قالوه فالعين باكيه فسبحان من أرسى ثبيرا مكانه ومن أمسك السبع الطباق كماهيه أيعصى إمام أوجب الله حقه علينا وأهل الشام طوع لطاغيه	فلم أر منزولا به بعد هجعة ألذ قرى لولا الذي قد نحاذره وكلا أحاذر بوابين قد بها وأسمر من ساج تاط مسامره فقلت لها كيف النزول فإنني أرى الليل قد ولى وصوت طائره	لبشر بن مروان على كل حالة من الدهر فضل في الرخاء وفي الجهد قريع قريش والذي باع ماله ليكسب حمدا حين لا أحد يجدي ينافس بشر في السماحة والندى ليحرز غايات المكارم بالحمد فكم جبرت كفاك يا بشر من فتى ضريك وكم عيلت قوما على عمد وصيرت ذا فقر غنيا ومثريا فقيرا وكلا قد حذوت بلا وعد
8th	لا تمزج الخمر على حال وسقنيها بنت أحوال عتقها الكردي في مجلس بين بساتين وأجبال ثم أتانا ناكسا رأسه منحدرا من مرقب عال إبريقه في كفه مترع مغترف من ذوب جريال نأخذها من كف ذي غنة كأنما خط بتمثال يسقيك بالعينين خمرا إذا ناغاك بالكأس بإعجال	فديتك ليس لي عنك انصراف ولا لي في الهوى منك انتصاف وصالك عندي الشهد المصفى وهجرك عندي السم الزعاف وقائلة متى عنها تسلى فقلت لها إذا شاب الغداف أطوف بقصركم في كل يوم كأن لقصركم خلق الطواف ولولا حبكم للزمت بيتي ففي بيتي لي الراح السلاف أنا العبد المقر بطول رق وليس عليك من عبد خلاف	الشيء محروص عليه إذا امتنع ولقل من يخلو هواه من ولع والمرء متصل بخير صنيعه وبشره حتى يلاقي ما صنع والدهر يخدع من ترى عن نفسه إن ابن آدم يستريح إلى الخدع ولمن يضيق عن المكارم ضيقة ولمن تفسح في المكارم متسع والناس بين مسلم ربح الرضى	ولو أن غير الموت لا قى عدبسا وجدك لم يسطع له أبدا هضما فتى لم يكن فقر يضعضع متنه ويبدي الغنى منه لنا خلقا ضخما فتى لو يصاغ الموت صيغ كمثله إذا الخيل جالت في مساحلها قدما ولو أن موتا كان سالم رهبة من الناس إنسانا لكان له سلما	وأحببت من حبها الباخلي ن حتى ومقت ابن سلم سعيدا إذا سيل عرفا كسا وجهه ثيابا من المنع صفرا وسودا يغير على المال فعل الجواد وتأبى خلائقه أن تجودا

	ليس بمحتاج إلى مكحل ولا دماليج وخلخال خال به في خده واضح وابأبي ذلك من خال		فيما يمض وبين من خسر الجزع والحق متصل ومتصل به فإذا سمعت بميت فقد انقطع		
9th	إذا رحت من سكر غدوت إلى سكر وأنفقت في الهوى وفي لذتي عمري ولم لا أجر الذيل في ساحة الصبا وشرخ شبابي قائم لي بالعذر ومعشوقة الألحاظ تهتز للصبا كما اهتز غصن البان في الورق الخضر مهفهفة صفراء إلا لآلئا إذا ابتسمت بيضا يلحن من الثغر قطعت بها ليل التمام وبدره إذا ما رآها ظنها غرة البدر	بعينيك ما ألقى وتعلم أن لي فؤادا سيلقى بالهوى آخر الدهروتزجرني العذال عن طاعة الهوى وقد علموا أن لا سبيل إلى الصبروقد كنت أخفي الشوق ما ستر الهوى علي فنمت عبرة بدم يجري فيا حسن أيامي وطول سلامتي ولذة عيشي لو علمت من الهجر	أيها المغرور مهلا فلقد أوتيت جهلا كم إلى كم تحسن القو ل ولا تحسن فعلا ظاهر يجمل والبا طن لا يخفى على ربك كلا	لا يبعد الله أسلافا لنا سبقوا ولو بقوا للقوا ما لا يحبونا كيف العزاء وما في العيش مغتبط ولا اغتباط لأقوام يموتونا متى تعش قبلي الأحياء يدركنا وإن تمت قبلي الأموات يعفونا لا بد من ميتة للمرء أو هرم يظل منه جليد القوم موهونا والبيض والجون لا نهوى فراقهما ولا نزال نذم البيض والجونا	أيا من له الشرف المستقل ومن جوده العارض المستهل ويا من أضاء كشمس الضحى فأضحى عليه به نستدل
10th	كأن في كل عضو لي وجارحة قلبا يحن وعينا تشتهي النظرا شوقا إلى ناعم اللذات لو لمست كفاه جلمود صخر أنبت الصخرا غصن تشرب من ماء النعيم فلو بزت سرابيله عن جسمه لجرى قد صاغه من نسيم الطيب خالقه وصانه في حجاب النور فاستترا رقت حواشيه حتى لو يمر به وهم الضمير لفرط اللين لاعتصراره انظر ترى ملكا ذ الشخص أم بشرا لو تعصر الخمر من خديه	إني لأحسد مقلتي عليكا حتى أغض إذا نظرت إليكا وأراك تنظر في شمائلك التي هي فتنتي فأغار منك عليكا من لطف إشفاقي ورقة غيرتي إني أغار عليك من عينيكا ولو استطعت جرحت لفظك غيرة أني أراه مقبلا شفتيكا خلص الهوى لك واصطفتك مودتي حتى حذرت عليك من أبويكا	اسم الذي أنا طائعا أفديه خاف ولكن فطنتي تدنيه مقدار ثالثه إذا حصلته مضروب حاشيتيه في ثانيه	أتخشى بنو حمدان ميلي عليهم وقد صفرت أحلامهم يوم غرب وما كل من يجني علي يخافني فيا خوف لا تذهب بهم كل مذهب وإني لأعفو عن ذنوب كثيرة لو اخترت ما طالت بها يد مذنب تجنت بنو حمدان كل جناية علي فلم أغضب ولم أتعتب وكادت من التأنيب تطلق خبوتي متى كان عرضي قبلها لمؤنب وصار التلاقي والمزار يريبها فداويت وجدي بب	أسمعتما أن الجبال تضام وعلمتما من غالت الأيام فجع تطير له على أحشائنا شعل وتسقط في القلوب سهام ورزية أخذ الردى ما يبتغي منا ونال بها الذي يستام شهدت بتحليل الدموع وخبرت أن العزاء على اللبيب حرام كنا نعد الحصن دار إقامة فاليوم وقفتنا به إلمام يبكي الغمام المستسير بأرضها ونقول جاد بذي الغميم غمام إن يفترق أحبابنا أيدي سبا عنها فقد يتفرق
11th	كأن الشقائق إذ أبرزت غلالة لاذ وثوب أحم قطاع من الجمر مشبوبة بأطرافها لمع من حمم	أصبت بعيني من أصاب بعينه فؤادي ولم يعقل دمي يوم طله لقد ثأرت عيني بقلبي ولم يكن حلالا له من مهجتي ما استحله فأهلا بعينيه وإن طلتا دمي فكم مالك لم يرزق العبد عدله	جحدت ولاء مولانا علي وقدمت الدعي على الوص يمتى ما قلت إن السيف أمضى من اللحظات في قلب الشجي لقد فعلت جفونك في فؤادي كفعل يزيد في آل النبي	دع الدنيا فقد غدرت بأقوام بها وثقوا وكانوا يجمعون لها كأنهم لها خلقوا	كان قضاء الإله مكتوبا لولاك كان العزاء مغلوبا ما بقيت كفك الصناع لنا فكل كسر يكون مرؤوبا ما احتسب المرء قد يهون وما أوجع ما لا يكون محسوبا نهضا بها صابرا فأنت لها والثقل لا يعجز المصاعيبا فقد أرتك الأسى وإن قدمت عن يوسف كيف صبر يعقوبا

12th	أجميل بأن يبطل صوت ال عود من عندنا وعندك خمر فأتنا يا أبا المحاسن إنا كنجوم وأنت للشرب بدر وأتنا بالمدام إنا أناس ما لنا عن شرب المدامة صبر وتفضل كما عهدت ودع عن ك اعتذارا فليس يقبل عذر وإذا ما تأخر الخمر عنا كان للشعر فيك نهي وأمر	أهلا وسهلا بالكتاب الواصل ليهيج من ألم الفراق بلابلي وافى فواقى لي بطيب بشائر ودنا فأدنا لي سرور مخابلي فطفقت أدنيه على كبدي عسى يطفي حرارة مهجتي وبلابلي	قلت للحاسد نما عملت فيه البشاره والولايات جميعا عند أهليها معاره ما عبيد الله ممن نال بالعزل خساره هو لم يعزل ولكن عزلت عنه الوزاره وكأني بالليالي ولقد تكفي الإشاره + للجسم إن داخلته روحه فرع والروح بالجسم عند البعث يمتسك روحان هذا له وزن وذاك بلا وزن به تم في أجسادها الحرك هذا به تغلب	تقول ابنة السعدي وهي تلومني أما لك عن دار الهوان رحيل فإن عناء المستنيم إلى الأذى بحيث يذل الأكرمون طويل وما في الورى إلا لك البدر والد ولا لسواك النيرات قبيل وعندك محبوك السراة مطهم وفي الكف مطرور الشباة صقيل	كاد الكمال يعود ربعا بلقعا حتى رفعت مناره فترفعا ما كنت إلا البدر حجب ضوءه ستر الغمام ومن قريب أقشعا شكرا لدهر زاد عنك صروفه فثناك محروس المحل مرفعا يغدو من العار المدنس عاريا ويروح بالمجد الأثيل ملفعا أوضحت من سبل المكارم ما عفا
13th	لهفي على شادن في حسن طلعته وشعره صار إصباحي وإمسائي قد برد القلب في تموز مرشفه وظل يحرق في كانون أحشائي	يا ومضة البرق من ذات الأثيلات صفي لأهل الحمى وجدي ولوعاتي ويا نسيم فروع البان من إضم هل ما مضى في الحمى بعد النوى آت كأن فيك شذى من أرض كاظمةس فهل مررت على تلك الثنيات وحملوك رسالات فوا أسفا عليهم وعلى تلك الرسالات عوجوا وحيوا على الخلصاء دارهم علي أعلل قلبي بالتحيات واستخبروا دمنة الوادي أهل خطرت ظعونهم بالقدود السمهريات فقد تجيبكمث الخرساء ناطقة عنهم سريعا ولكن بالإشارات + أموسى ولم أهجرك والله إنما هجرت الكرى واللب والأنس والصبرا تركتك لا نقضا لعهدي بل أرى حياتي ذنبا بعد بعدك أو غدرا قنعت على رغمي بذكرك وحده أدير عليه الخمر والأدمع الحمرا أقبل من كأس المدير حبابها إذا مثلت عند المنى ذلك الثغرا	فإن نسي الإنسان ركنا فإنه يعيد ويقضي ما تضمن واحتوى وإن لم يكن ركن وعطل سنة فلم يأنس الزلفى ولم يبلغ المنى وذلك في كل العبادات سائر وليس جهول بالأمور كمن درى إذا كان هذا ظاهر الأمر فالذي توارى عن الأبصار أعظم منتشا وهذا طهور العارفين فإن تكن من أحزابهم تحظى بتقريب مصطفى	على م وفي م ظلما تلحياني ذراني لا أبا لكما ذراني وحسبكما فما سمعي بمصغ ولا واع لما تتحدثان فلي همم إذا جاشت أرتني قرى عمان ميلا من عمان إذا سولمتما فتناسياني وإن أسلمتما فتذكراني فمثلي من يقيم صغى الأعادي ويستعدى على نوب الزمان وما ذكر المنية عند أمر أحاوله بثان من عناني إذا يومي أظل فما أبالي بسيف كان حتفي أو سنان ومن يك عمره المكتوب تسعا فلا يخشى المنية في الثمان وبعد عن أخ لأب وأم إذا ما عق خير من تدان فلا يتوهم السفهاء أني أحن إلى غوان أو مغان ولا أني أرى والموت حتم بعين جلالة من لا يراني عظيم الناس في عيني حقير إذا بالمقلة الخوصا رآني	سواي في سلوته يطمع فعنفوا إن شئتم أو دعوا أوضحتم الرشد فمن يهتدي وقلتم الحق فمن يسمع + دمت مجد الدين يا من نفسه بالجود صبه لك أخلاق كرام وسجايا مستحبه ويد تخجل فينا ال غيث بل تفضل سحبه
14th		ما ضر إحسانك يا سيدي	لزمت انفرادي إذ قطعت العلائقا	سكنت وابني بدار قوم	

	ألا بأبي خل حمدت إخاءه رقيق الحواشي جامع الشمل والأنس قطعنا به ليلا كأن حديثه أزاهير روض صين عن بذلة اللمس وبتنا يعاطينا الحديث كأنه جنى النحل ممزوجا لنا بابنة الكأس حبيب إلى كل الأنام كأنه تقسم بين القلب والروح والنفس هو النجم جلى ظلمة الليل نوره فأوضح من شمس وأفصح من قس ولكنه إذ لاحت الشمس غائب ولا غرو إن النجم يخفى مع الشمس	لو حاد في أمري عن الاقتصاد يا عين آمالي إذا استجمعت إني إلى مورد لقياك صاد	وجالست من ذاتي الصديق الموافقا وآنسني فكري لبعدي عن الورى فلست إلى شيء سوى العلم تائقا	أوقاتنا تارة وتاره فإنها بالخصام نار وقودها الناس والحجارة	ملأت قلوب العالمين محبةإن القلوب صنائع الإحسان
15th	رعى الجدي المغير رياض زهر بها العذراء أسكنت القصورا وأسفرت الثريا إذ تهادت لطرف سهيلها فازداد نورا	أحبابنا خلفتموني لقى في الدار صبا كاد أن يهلكا لا تشتكي المحل ربوع لكم فإنني استغرقتها بالبكا	تجلت لوحدانية الحق أنوار فدلت على أن الجحود هو العار وأغرت لداعي الحق كل موحد بمقعد صدق حبذا الجار والدار	ولما رأيت الدهر يقتل أهله وأيقنت أني عن قريب سأقتل جعلت حجابي منزلي وتشاغلت يداي عن الدنيا بما هو أفضل	إذا تطاولت الأعناق للرتب أتتك تسعى وما أمعنت في الطلب
16th	علقته أسمرا حلو القوام وكم لي من حديث هواه سر من سمر مؤنث الجفن الا ان مقلته تزري اذا ما رنت بالصارم الذكر	حتام تخلف وعدي وكيف ضيعت عهدي يا اذكر الناس هجرا ما كان انساك ودي	الصبر من كرم الأخلاق فاصطبر يا نفس لو كان طعم الصبر كالصبر من ليس يرضى على الأقدار أرغمه على المكاره إرغاما قضا القدر	أبى الجسم إلا أن يزال لدائه على العز مالي والأذى يال نبع أضعت عظيمات الأمور ولم أكن لتدبيرها لولا أذى بالمضيع وأقسم لو يلقي امرؤ ما لقيته لبات بحال الأيس المتضعضع وإني لجلد في المصيبات صابر على كل فجاج من الخطب موجع	لعلياك في أفق الكمال مطالع وسعدك بالإقبال والعز طالع وغصن المعالي بالهنا عاد مثمرا وروض المعاني قد زها وهو يانع
17th	ولرب ليل برقه يبدي خفايا الأنجم يطوي وينشر تارة كصحيفة لمنجم	لي حبيب فداه كل حبيب ليس إلا به الحياة تطيب	قالوا ذكرت أبا بكر فقلت لهم لا غافلا أبدا عنه ولا لاهي قالوا كذاك أبا حفص وصاحبه عثمان قلت نعم والحمد الله	ومثقل وافاه يوم حمامه في غفلة وكذا الحياة غرور قد قلت إذ مروا على بنعشه يعلوه منه على الأكف ثبير ما كنت احسب قبل موتك أن أرى رضوى على أيدي الرجال يسير	شجاع إذا قام في معرك يزيد اشتعالا لديه العراك فكم دارع صاده في الوغى كأن الدروع عليه شباك

18th	وبكرة ما رأى الراؤن مشبهها كأنها سرقت سرا من الزمن غيم وطل وروض مونق وهوى يجري مع الروح مجرى الروح في البدن	سل عن فؤادك حين طاش بك الهوى إن كنت تملك في الغرام فؤادا هيهات عهدي يوم منعرج اللوى قادت أزمته الهوى فانقادا	تجددي تجددي يا أرشليم الأبدي فالنور وافى ربعك والرب يدعوك اقتدي هذا هو البيت الذي بناه آب سرمدي هذا هو البيت الذي بالإبن خير موطد هذا هو البيت الذي ال روح له بمجدد ذاك الذي ينيرنا بقدسه المشدد	تفكر بالمآل ودع طماعا فمالك هائما في حب مالك فمن يك زاهدا في المال أمسى غني النفس خير المال مالك	سلام على رب الفضائل والعلى على عالم الدنيا على علم الهدى
19th	لهفي لفاتر لحظه الفتاك رشأ به سلب الهوى إدراكي نصبت شراك لحاظه فاصابني وله نصبت فأخطأت أشراكي هو ثالث القمرين يحسد غيره فيه البسيطة رابع الأفلاك فإذا رنا الرشا وإذا انثنى منه القوام أراك غصن أراك يفتر عن منظوم مبسمه ولي عين بمنثور المدامع باك يا مهجتي أضناك رمح قوامه طعنا وسيف لحاظه بتاك	يا مانعي وصال من أحببته بزواجر قد خافها ونواه الله أعطاني حبيبا مخلصا فمنعتموني من عطاء الله	ولما قضت مني الحياة مأربا وقد تركوني في المقابر أعظما وقالوا قضى نحبا وسار لربه ومات بحمد الله إذ مات مسلما ومن عبد الرحمن سبعين حجة لقى الله باريه أبر وأرحما	أدمى الكنائس إن تكن عبثت بكم قوم بهمتهم تدك جبال أو ان تكن قد أوصلت لكم الأسى أيدي الحوادث أو تغير حال فلطالما سجدت لكن شمامس عبدوا الصليب وللظلالة مالوا فأبادهم رب السماء بفتية شم الأنوف عوابس أبطال	مولاي يهنك هذا الفتح مبتكرا من غير جيش ولاسيف به شهر
20th	طلعت من الافق الجديد عطارد للعاكفين على القديم تطارد طلعت بليل غاسق فتألقت والناس أما مؤمن أو جاحد	بعد أن فارقت ليلي مرغما حثني طيف لها ان ارجعا والتقينا بغتة بعد النوى فتعانقنا واجهشنا معا	لا تكون الأشياء خيرا على العل لات مما قد كانت الأشياء ليس تأتي يوما باحسن مما قد اتته الطبيعة العمياء	ذريني ألاقي الموت يا نفس واذهبي فما أنا من جدي أعز ومن أبي ذريني بموتي وهو آخر ملجئ أخفف من عبء المعيشة منكبي ذريني أنجو من حياة عذابها شديد ومن لم يحى لم يتعذب ذريني ذريني أستنيم إلى الردى فإن الردى يا نفس سؤلي ومطلبي لقد كبرت يا نفس أيام شقوتي وطال على جمر الحياة تقلبي	ضياء العيد من أفق السلام تبدى مشرقا في بدء عام بهاتيك الليالي البيض حلت فاني حسنها طيب المنام وأمت ربع اقوام كرام تدير عليهم كأس المدام مشعشعة كان بها خلالا تضاهي خلق كوكيل الهمام فلا زالت شموسك طالعات بافق السعد يا خير الكرام