BERT-based Classical Arabic Poetry Authorship Attribution

Lama Alqurashi, Janet Watson, Jacob Blakesley, Serge Sharoff
University of Leeds
{mlla, j.c.e.watson, j.blakesley, s.sharoff}@Leeds.ac.uk
Algurashi.lama@gmail.com

Abstract

This study introduces a novel computational approach to authorship attribution (AA) in Arabic poetry, using the entire Classical Arabic Poetry corpus for the first time and offering a direct analysis of real cases of misattribution. AA in Arabic poetry has been a significant issue since the 9th century, particularly due to the loss of pre-Islamic poetry and the misattribution of post-Islamic works to earlier poets. While previous research has predominantly employed qualitative methods, this study uses computational techniques to address these challenges. The corpus was scraped from online sources and enriched with manually curated Date of Death (DoD) information to overcome the problematic traditional sectioning. Additionally, we applied Embedded Topic Modeling (ETM) to label each poem with its topic contributions, further enhancing the dataset's value. An ensemble model based on CAMeLBERT was developed and tested across three dimensions: topic, number of poets, and number of training examples. After parameter optimization, the model achieved F1 scores ranging from 0.97 to 1.0. The model was also applied to four pre-Islamic misattribution cases, producing results consistent with historical and literary studies.

1 Introduction

Poetry holds a central place in all cultures and particularly Arab culture, functioning not only as an art form but also as a vital historical record and a powerful social tool. Its influence on public opinion was comparable to modern media, with verses that could incite or resolve conflicts, often accepted as undeniable truth by the masses.

However, the field of Arabic literature faces significant challenges in the attribution and authentication of texts, especially concerning pre-Islamic poetry. The lack of a written tradition during that time has led to significant scholarly efforts aimed at verifying and establishing accurate attributions. This issue was already noted as early as the 9th century in works like Ibn Sallām's *Ṭabaqāt fuḥūl al-shuʿarāʾ* (2003), which criticized fabricated poetry for lacking authenticity and value.

Previous studies, such as (Alanazi, 2015) and (Ahmed et al., 2017) explored AA in Arabic poetry. However, these studies primarily relied on conventional machine learning techniques like Random Forest (RF), Support Vector Machines (SVM), Naïve Bayes (NB), and Linear Discriminant Analysis (LDA), which are no longer considered state-of-the-art.

Transformer models have demonstrated state-of-the-art (SOTA) performance across various NLP tasks, including authorship attribution (AA). (Fabien et al., 2020) highlighted how BERT outperformed existing methods in AA tasks. In the context of fabricated or fake texts, (Silva et al., 2024) employed a BERT-base model to tackle forgery detection in novels.

This study represents the first application of Authorship Attribution (AA) analysis directly addressing actual cases of attribution problems in Arabic literature, with a comprehensive focus on linguistic, literary, and technical aspects.

Leveraging a vast dataset that includes the entire corpus of Classical Arabic poetry and advanced computational methods, this work employs BERT, a state-of-the-art transformer-based model, fine-tuned on Arabic poetry. Additionally, to enhance the accuracy of authorship attribution and control for thematic variation, topic modeling techniques are integrated into the classification process.

The remainder of this paper is structured as follows: Section 2 reviews related work, Section 3 describes the proposed methodology, Section 4

Proceedings of the 31st International Conference on Computational Linguistics, pages 6105–6119
January 19–24, 2025. ©2025 Association for Computational Linguistics

discusses experimental results, Section 5 presents overall discussion, Section 6 provides the conclusion and Section 7 outlines limitations and suggests future work directions.

2 Related work

2.1 General overview of AA

AA presents diverse challenges across different contexts and data types, with no single method universally applicable. This variety of methods complicates comprehensive evaluation of their effectiveness. Early AA research primarily utilized straightforward algorithms that estimated authorship probabilities based on feature frequencies within known-author texts, attributing authorship to the author whose text featured the closest distribution (Mosteller, 1964; Zhao, 2005; Clement, 2003; Madigan, 2005). Distance measures also played a significant role, comparing text vectors using methods such as Euclidean and Manhattan distances (Ahmed, 2018; Halvani et al., 2020) or cosine similarity, which is particularly effective with raw frequency data (Bakly et al., 2014; Koppel & Winter, 2014; Ramezani, 2021). A key advancement was the development of the CNG measure for assessing text dissimilarity (Selj, 2003). Subsequent variations, such as those proposed by (Potha & Stamatatos, 2014) and (Stamatatos, 2017), adapted CNG for unbalanced documents. Other distance measures, including min-max (Koppel & Winter, 2014), Hel (Ramezani, 2021), and L1-norm (Kocher, 2017), have been employed, particularly when training machine learning models with limited data. Establishing a distance threshold is crucial for determining whether two documents should be attributed to the same author. In recent years, machine learning (ML) techniques have become dominant in AA research. Some ML methods, like k Nearest Neighbours (kNN) (Abbas et al., 2019; Maurya et al., 2016; Ramezani, 2021), bear similarities to distance measures, while others, such as Naive Bayes (NB) and its variants, use probabilistic approaches (Koppel et al., 2013; Ramezani, 2021). Support Vector Machines (SVM) are notably popular in natural language processing (NLP) and AA due to their effectiveness with high-dimensional data, which is common in text analysis (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).

Decision Trees (Koppel et al., 2013) and Random Forests (Kumar et al., 2017) are also utilized, offering the benefit of decision graphs for result interpretation. Additional ML algorithms, including 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), are also applied, though they are generally less favored compared to the aforementioned techniques.

In recent years, researchers have increasingly utilized neural networks (NNs) to tackle authorship attribution (AA) challenges. NNs have gained prominence in artificial intelligence for their ability to learn complex relationships within data. Recurrent Neural Networks (RNNs) and their variants, such as Long Short-Term Memory (LSTM) networks, are particularly effective for sequential data, making them well-suited for language tasks (Hosseinia & Mukherjee, 2018). Simpler NN architectures like Feedforward Neural Networks (FNNs) have also shown their effectiveness, often surpassing traditional machine learning techniques by learning intricate patterns between inputs and outputs (Al-Sarem et al., 2020; Benzebouchi et al., 2019; Sari et al., 2018).

Convolutional Neural Networks (CNNs), initially designed for image processing, have been adapted for textual data and are especially useful for analyzing short texts such as tweets (Hitschler et al., 2018; Shrestha et al., 2017). One of the significant advantages of NNs is their reduced need for extensive feature engineering, although they require substantial amounts of data to reach their full potential.

2.2 AA in Arabic texts and poetry

In their study, (AlZahrani & Al-Yahya, 2023) demonstrate how fine-tuning pretrained transformer models can be effectively applied to address authorship attribution tasks in specialized domains of Arabic texts.

(Bsir et al., 2024) fine-tune a customized transformer-based model, specifically BERT, to enhance author profiling accuracy for Arabic and other languages, achieving approximately 79%

accuracy by retraining on the PAN 2018 authorship dataset.

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 (Ouamour and Sayoud, 2012, Howedi and Mohd, 2014, Altheneyan and Menia, 2014, Abbasi and Chen, 2005, Benjamin et al., 2013, Alwajeeh et al., 2014, Baraka et al., 2014, Otoom et al., 2014, Abuhaiba and Eltibi, 2016, Al-Sarem and Emara, 2019, Khalil et al., 2020, Al-Sarem et al., 2020, Albadaneh et al., 2015, Alsager, 2020, Ahmed, 2017, Ahmed, 2018, Al-sarem et al., 2018, Ahmed, 2019a, Ahmed, 2019b), function words (Shaker and Corne, 2010, Altakrori et al., 2018), characters (Ouamour and Sayoud, 2012, Ouamour and Sayoud, 2013, Howedi and Mohd, 2014, Kumar, 2012), syntax (Howedi and Mohd, 2014, Altheneyan and Menia, 2014, Abbasi and Chen, 2005, Benjamin et al., 2013, Otoom et al., 2014, Abuhaiba and Eltibi, 2016), structure (Altheneyan and Menia, 2014, Alanazi, 2015, Abbasi and Chen, 2005, Benjamin et al., 2013, Otoom et al., 2014), poetic features (Alanazi, 2015), HTML features (Benjamin et al., 2013), morphological features (Baraka et al., 2014, Al-Ayyoub et al., 2017b, Al-Sarem and Emara, 2019, Omar and Hamouda, 2020, Ahmed, 2018, Ahmed, 2019a, Ahmed, 2019b), content-specific features (Altheneyan and Menia, 2014, Otoom et al., 2014), stylometric features (Abbasi and Chen, 2005, Al-Ayyoub et al., 2017b, Al-Sarem et al., 2020, Altakrori et al., 2018, Al-sarem et al., 2018), and diacritics (Ahmed, 2018, Ahmed, 2019a, Ahmed, 2019b).

In terms of algorithms, Some studies employed distance or dissimilarity measures (Kumar, 2012, Abbasi and Chen, 2005, Ahmed, 2019b, Al-Sarem and Emara, 2019, Ouamour and Sayoud, 2013, Ahmed, 2018, Ahmed, 2019a, Ahmed, 2019b), while others utilized traditional machine learning algorithms such as Support Vector Machines (Ouamour and Sayoud, 2012, Ouamour and Sayoud, 2013, Howedi and Mohd, 2014, Alwajeeh et al., 2014, Baraka et al., 2014, Otoom et al., 2014, Al-Ayyoub et al., 2017b, Altakrori et al., 2018, Ahmed, 2019b), Linear Regression (Al-Sarem and Emara, 2019), Naive Bayes (Howedi and Mohd, 2014, Altheneyan and Menia, 2014, Alwajeeh et al., 2014, Al-Ayyoub et al., 2017b, Albadaneh et al., 2015, Altakrori et al., 2018), Decision Tree (Al-Ayyoub et al., 2017b, Altakrori et al., 2018) and Random Forest (Altakrori et al., 2018, Alanazi, 2015). Additionally, two studies applied deep learning techniques, specifically using multilayer perceptrons (Ouamour and Sayoud, 2013, Al-Sarem and Emara, 2019).

In the field of AA for Arabic texts, there has been a few studies that applied a transformer-based model, focusing on regular Arabic prose (AlZahrani and Al-Yahya, 2023 and Bsir et al., 2024). Our work represents a notable step forward by being among the first to apply transformers to Arabic poetry and to incorporate the entire corpus of Classical Arabic poetry into an authorship attribution study.

2.3 State-of-the-art techniques

Earlier approaches to AA relied on manually crafted stylistic features, but recent advancements have shifted towards using text embeddings from pre-trained language models like BERT. BERT has achieved state-of-the-art performance across various language tasks, including AA (Barlas & Stamatatos, 2020; Tyo et al., 2021; Wang & Iwaihara, 2021), and is particularly effective in cross-topic and cross-genre contexts (Barlas & Stamatatos, 2020).

One notable method, "BertAA: BERT Fine-Tuning for AA" (Fabien et al., 2020), demonstrates how fine-tuning BERT with additional layers enhances authorship classification, resulting in up to a 5.3% performance improvement on datasets like Enron Email, Blog Authorship, and IMDb. Incorporating stylometric and hybrid features into an ensemble approach further boosts performance by 2.7%.

Expanding on these advances, Wang and Iwaihara (2021) introduce a hybrid model that combines a RoBERTa-based language model with a CNN for feature embeddings. This model achieves state-of-the-art results in tweet authorship attribution by effectively capturing both stylistic features and contextual nuances.

Kumarage and Liu (2023) explore neural authorship attribution by analyzing stylometric features in various large language models, improving classifiers to trace AI-generated text back to its source.

(Uchendu et al., 2023) present TopFormer, which integrates Transformer-based architectures with Topological Data Analysis (TDA), enhancing authorship attribution of deepfake texts and achieving up to a 7% increase in Macro F1 score.

Silva et al. (2024) introduce Forged-GAN-BERT, a GAN-BERT-based model that improves the classification of forged novels through data augmentation with ChatGPT and GAN generators, achieving F1 scores of 0.97 for single-author and 0.71 for multi-author settings.

Finally, Sarwar et al. (2024) present AGI-P, a solution for author gender identification using a balanced dataset of 1,944 samples, achieving an accuracy of 92.03%. This approach outperforms existing classifiers and fine-tuned multilingual models through a customized fine-tuning strategy.

3 Methodology

3.1 Data collection and pre-processing

The data source consists of Arabic poetry spanning all periods, sourced from the AlDiwan website, which provides the data in HTML format. This dataset is significant due to its richness, including a notably large number of poets and poems. The dataset is labeled with information related to topics, metres, genre, and rhyme.

This dataset comprises over 100,000 poems from various historical periods. The classification is organized according to the classical chronological periods: 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), and Ottoman (1299 - 1922 AD). However, the current segmentation lacks precision, with varied time windows that do not contribute effectively to meaningful analysis.

To extract this dataset, a web-scraping script was created to retrieve and restructure the data into a dataframe table. 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.

As demonstrated, the classic division proves to be imprecise and has the potential to introduce noise to our model. To address this limitation, we 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 dataset encompasses a total of 784 poets, 77,850 poems, and 6,609,495 words. The pie chart breaks down the distribution of poets and poems across different centuries Figure 1. This breakdown is crucial for understanding the quantitative aspects of the dataset and its distribution over time, and it is of the same importance for AA classification.

Our dataset surpasses the size of other datasets used in Arabic language AA research Figure 2. While the same corpus is available from other sources online, our version has been enhanced with more accurate century-based segmentation. Additionally, we applied topic modeling and included topic labels as an extra feature. This dataset will soon be made publicly available.

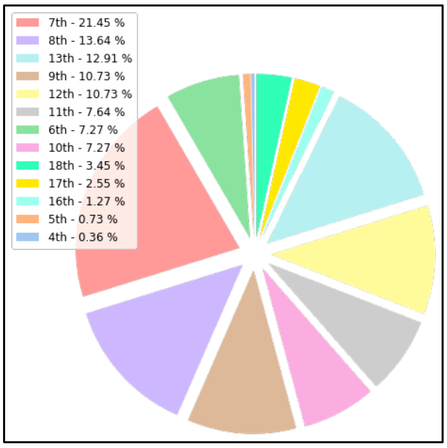

Figure 1: Data distribution across centuries.

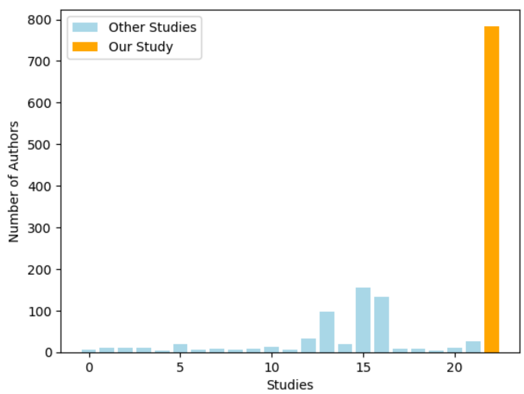

Figure 2: Number of authors in our study and other Arabic AA studies.

Epochs	Topics	Diversity	Coherence
15	5	0.768	0.231
	10	0.44	0.219
	15	0.381	0.223
50	**5**	**0.872**	**0.236**
	10	0.712	0.230
	15	0.554	0.231
70	5	0.864	0.230
	10	0.752	0.234
	15	0.562	0.235

Table 1: Topic modeling configurations.

3.2 Topic modeling

In this study, we employed the Embedded Topic Model (ETM) (Dieng et al., 2020) to identify and categorize latent topics within our corpus of Arabic poetry. This model was selected due to its ability to leverage word embeddings in topic modeling, allowing it to capture semantic relationships among words, a capability traditional topic models lack. Unlike conventional models, which primarily rely on word co-occurrence within documents, the ETM provides more coherent and interpretable topics by accounting for word semantics. Given the complex and nuanced nature of Arabic poetry, the ETM offered a fitting solution for our research objectives.

To prepare the corpus for ETM, we transformed the dataset into a bag-of-words (BoW) format[1], where each document was represented by word frequencies without regard to word order. This preprocessing step is essential in topic modeling, aligning the data structure with ETM's input requirements and enabling a more efficient exploration of latent topics.

We then tuned the ETM's hyperparameters, including the number of topics, learning rate, and epochs, by performing a grid search across various parameter ranges. The model's performance was evaluated using diversity and coherence metrics to ensure the most accurate and meaningful topic distribution within the Arabic poetry corpus.

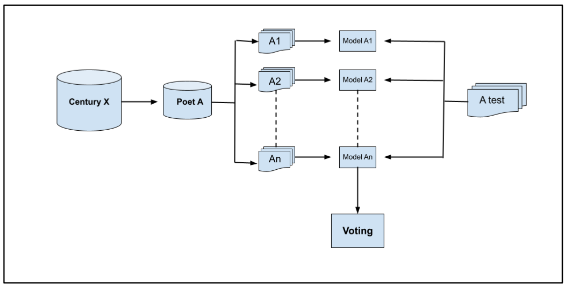

Figure 3: Flowchart of our ensemble model.

3.3 Model development

To address the challenges of data scarcity and overfitting in the task of AA, we propose a multi-model ensemble approach Figure 3. This methodology leverages the diversity of multiple distinct models, each fine-tuned on CAMeLBERT (Go Inoue et al., 2021), to capture unique authorial patterns. By training individual models on different subsets of the data, we ensure that each model specializes in a distinct subset of authors, allowing for a more nuanced representation of stylistic differences. Instead of creating a single contrasting class composed of all non-target authors, we construct multiple datasets where each dataset excludes the target author's writings. This allows us to develop a series of models, each focusing on different groups of authors. Once trained, these models are evaluated on unseen test data.

To combine the predictions from these models, we employ a majority voting technique, wherein each model contributes a vote for its predicted author. The author who receives the most votes from the ensemble of models is selected as the final prediction. This method improves the overall robustness of the system by minimizing the influence of any one model’s incorrect prediction, leading to more accurate and stable results.

We evaluate the ensemble’s performance using accuracy as the primary metric, assessing its ability to correctly attribute unseen texts to the correct author. Additionally, we fine-tune the ensemble’s parameters to ensure optimal performance for the AA task, further validating the model through five cross-validation techniques to ensure generalizability to unseen data.

To ensure the robustness of our ensemble model, we conducted thorough validation, confirming that the chosen voting thresholds are the most effective for this specific authorship attribution task. Beyond

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

standard model training and testing, we devised an additional evaluation protocol. Each model was tested on a dataset of the same size as the training set but composed entirely of unseen data, providing a reliable measure of the model's generalizability.

Key Factors for Model Optimization:

One critical factor in optimizing the ensemble involves determining the appropriate number of poets in the "opposite" class (non-target authors). We rigorously tested different configurations, examining how this parameter affects model performance across poets from various historical periods. This testing allowed us to pinpoint the ideal configuration that maximizes accuracy.

Additionally, we explored the effect of varying data sizes. In the pre-Islamic context, for instance, there is no fixed number of available examples, with data often varying significantly in quantity. As a result, we tested the model's performance across a wide range of data sizes to identify how model accuracy responds to these fluctuations.

Analysis of Probability Distributions:

We also investigated the probability distributions of correctly classified versus misclassified instances to gain deeper insights into model performance Figure 4. By analyzing these distributions, we can identify a probability threshold that allows us to confidently interpret the model's outputs as highly reliable, providing a more nuanced understanding of the model's behavior.

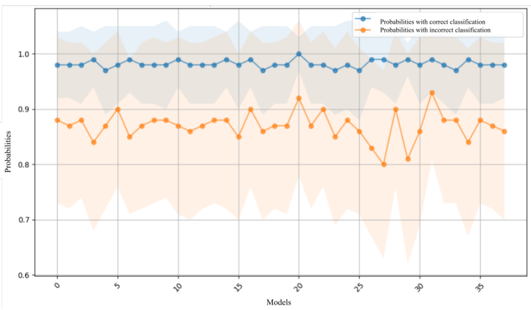

Figure 4: Comparison of probabilities with error bars.

4 Experiments and results

4.1 Parameter optimization

To optimize the parameters of our classification model, we employed a systematic approach involving parameter tuning and extensive experimentation. We utilized the Adam optimizer with its default learning rate and conducted experiments with varying epochs to identify the optimal settings. Our results indicated that using 100 epochs yielded the best performance for our model. In addition to optimizing the number of epochs, we experimented with different configurations for the number of poets included in the opposing class and the number of examples per class. Specifically, we tested scenarios with 1, 2, and 3 poets and with 40, 60 and 80 examples per poet. Our findings revealed that the best performance was achieved with a setup of 1 poet and 60 examples (see appendix 2). This configuration consistently produced the highest accuracy and robustness in our classification results.

4.2 Topic modeling

4.2.1 parameters

In terms of topic modeling, we assessed various parameters to find the optimal balance between diversity and coherence. Our results indicated that the best performance was obtained with 5 topics and 50 epochs Table 1. This configuration provided the most coherent and diverse topic representation for the dataset. In addition, we conducted a qualitative evaluation for the topic modeling by picking randomly poems for each topic from all centuries and then analyze their theme manually. this analysis not only strengthens the results we got but also showed the evolution of topics through centuries and their distribution (see appendix 1).

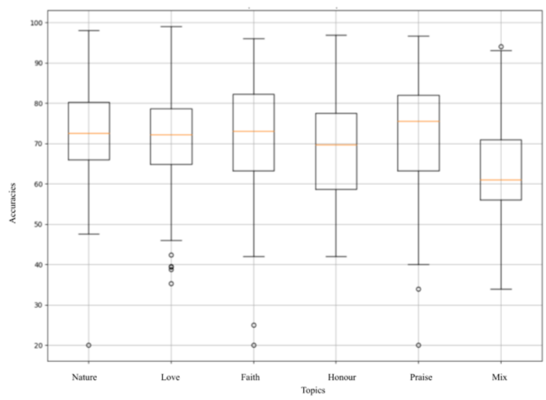

Figure 5: Control for topic across the corpus.

4.2.2 Topics

The model identified several key motifs for each topic based on word sets:

- Topic 0: Nature and Aesthetics - Words like "night," "sun," and "roses" suggest themes of beauty, nature, and appreciation for the natural world.
- Topic 1: Love and Longing - With terms such as "love," "heart," and "yearning," this topic centers around emotions of romantic love, desire, and heartache.
- Topic 2: Faith and Morality - Featuring words like "God," "Prophet Muhammad," and "morality," this topic explores religious devotion, moral values, and faith.
- Topic 3: Honour, Life and Death - Words like "life," "death," and "youth" reflect on human existence, mortality, and the meaning of life.
- Topic 4: Praise and Power - Focused on "royalty," "power," and "wealth," this topic deals with rulership, social status, and concepts of honour.

4.2.3 Topic control

Before constructing the authorship classifier, it was essential to examine the intersection of topics and author styles by establishing reliable topic labels. This is crucial because machine learning models learn both topic-specific language and unique author traits, and understanding their overlap is key.

This topic control approach resulted in slightly higher classifier accuracies, highlighting the importance of considering topics in authorship classification Figure 5. The absence of low-accuracy outliers in the mixed-topic category further underscores the stability achieved with a balanced dataset.

To further examine the effect of topics, we analyzed the misclassified instances and assigned topics to those examples, calculating the failure ratio for each topic Figure 6. This approach revealed insights into each poet's weaknesses concerning specific topics and highlighted the areas where the model struggled. Identifying these challenging topics helps us better understand the model's limitations and potential areas for improvement.

4.3 Performance evaluation

The analysis of average F1 scores for the tested models shows that the ensemble model achieved a score of 0.98, significantly outperforming the single model, which obtained a score of 0.75.

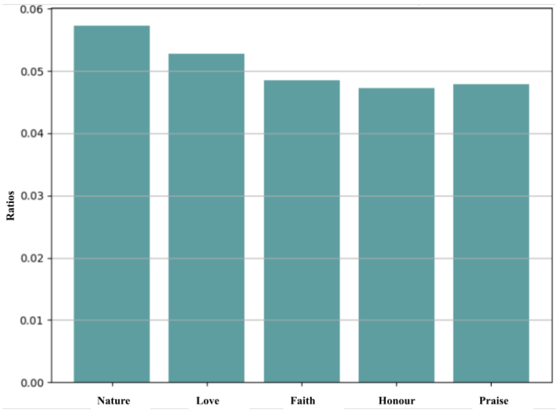

Figure 6: Ratios of classification fails for each topic.

4.4 Case studies of pre-Islamic AA

We present the results of our model's assessment for verifying the attribution of certain contested poems. The texts tested are initially ascribed to their assumed authors, although questions about their authenticity remain.

Poem 1 (Imru' al-Qays):

This poem, attributed to Imru' al-Qays, differs structurally from his other works, exhibiting linguistic anomalies and unusual vocabulary. When classified by the model, the poem showed a low confidence score of 0.71, contrasting with the usual 0.99 for other Imru' al-Qays works, reinforcing doubts about its authenticity.

Poem 2 (Al-Aʿshā):

Attributed to Al-Aʿshā and containing Islamic doctrinal themes, this poem raises doubts about authorship due to historical inconsistencies, particularly given Al-Aʿshā's known resistance to Islamic teachings. The model's confidence in attribution was 0.86, indicating some hesitancy.

Poem 3 (Al-Aʿshā):

This poem, regarded by historical scholars as fabricated, received a lower confidence score of 0.85 from our model. The philosophical content and divergence from Al-Aʿshā's known style further support the claim that this poem may not be authentic.

Poem 4 ('Ubayd Ibn Al-Abraṣ):

Although primarily attributed to 'Ubayd, this 21-line poem has been debated due to the absence of corroborating sources. However, our model provided a high confidence score of 0.99, strongly supporting its attribution to 'Ubayd, though the possibility of multi-author contributions cannot be detected by the model.

These case studies illustrate the model's capacity to evaluate disputed authorship with varying levels of confidence, aligning with historical scepticism in many instances.

5 Discussion

Our study demonstrates the effectiveness of advanced NLP techniques, specifically Embedded Topic Modeling (ETM) and CAMeLBERT-based ensemble models, for AA in classical Arabic poetry. The use of ETM allowed us to capture the semantic relationships between words, producing coherent and interpretable topics that aligned with the nuanced themes of Arabic poetry. Our parameter tuning experiments, particularly the use of 5 topics and 50 epochs, provided the best balance between topic coherence and diversity.

The ensemble approach significantly outperformed single models, achieving an F1 score of 0.98 compared to 0.75 for a single model. By incorporating different subsets of data in each model and employing majority voting, we enhanced robustness and accuracy in authorship classification. This method successfully captured stylistic variations across the dataset, proving especially effective for large, heterogeneous data.

The case studies highlight the model's ability to address disputed authorship, with confidence scores aligning with historical skepticism in many instances. Our refinement of historical data segmentation and manual annotation of poets' dates contributed to more precise modeling and analysis of stylistic evolution over time. Overall, this approach offers a reliable framework for AA in classical Arabic poetry.

6 Conclusion

In conclusion, this study presents a robust methodology for authorship attribution (AA) in classical Arabic poetry, combining modern NLP techniques with a carefully curated dataset. By leveraging Embedded Topic Modeling (ETM) and a CAMeLBERT-based ensemble model, we effectively captured both the thematic depth and how it can affect the stylistic classification of Arabic poetry across centuries. Our manual segmentation of poets by century, coupled with extensive parameter tuning, ensured accurate results, with the ensemble model achieving superior performance compared to single models.

This approach also proved effective in resolving authorship disputes, aligning well with historical evidence and skepticism. By exploring topics and stylistic traits in depth, we provided a clearer understanding of how authorial style evolved over time.

Overall, this study not only advances the field of AA in classical Arabic poetry but also offers valuable insights into the intersection of machine learning, literature, and historical analysis. Our findings lay a strong foundation for future research and potential applications in other domains of literary studies and historical authorship analysis.

7 Limitations and future work

Several limitations and avenues for future work are identified in this study. Firstly, incorporating metre features into the classification process could enhance model performance, but this requires a comprehensive analysis and sufficient data for effective training. Secondly, expanding the classification framework to include a distinct class for fabricated or fake poems might provide the model with more flexibility in distinguishing between authentic and inauthentic texts; careful development of this class is necessary to accurately represent inauthentic works. Lastly, exploring multi-author scenarios where attribution is shared between multiple authors could offer deeper insights into collaborative or disputed authorship cases.

8 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.

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. (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-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-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-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.

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

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.

AlZahrani, Fetoun Mansour, and Maha Al-Yahya. "A Transformer-Based Approach to Authorship Attribution in Classical Arabic Texts." *Applied Sciences* 13.12 (2023): 7255.

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.
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.
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

Bsir, Bassem, Nabil Khoufi, and Mounir Zrigui. "Prediction of Author's Profile basing on Fine-Tuning BERT model." *Informatica* 48.1 (2024).

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

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.

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.

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.

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 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.

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).

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., & 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

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.

Kumarage, Tharindu, and Huan Liu. "Neural Authorship Attribution: Stylometric Analysis on Large Language Models." In *2023 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC)*. IEEE, 2023.

language understanding. arXiv preprint arXiv:2003.00104.

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

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.

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).

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. (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.

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

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

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

Sarwar, R., Teh, P. S., Sabah, F., Nawaz, R., Hameed, I. A., & Hassan, M. U. "AGI-P: A Gender Identification Framework for Authorship Analysis Using Customized Fine-Tuning of Multilingual Language Model." *IEEE Access* (2024).

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

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

Silva, K., Frommholz, I., Can, B., Blain, F., Sarwar, R., & Ugolini, L. "Forged-GAN-BERT: Authorship Attribution for LLM-Generated Forged Novels." In *Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics: Student Research Workshop* (pp. 325-337). March 2024.

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

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

Uchendu, Adaku, Thai Le, and Dongwon Lee. "TopRoBERTa: Topology-Aware Authorship Attribution of Deepfake Texts." *arXiv preprint arXiv:2309.12934* (2023).

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

Wang, Xiangyu, and Mizuho Iwaihara. "Integrating RoBERTa Fine-Tuning and User Writing Styles for Authorship Attribution of Short Texts." In *Web and Big Data: 5th International Joint Conference, APWeb-WAIM 2021, Guangzhou, China, August 23–25, 2021, Proceedings, Part I*, 5. Springer International Publishing, 2021.

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.

9 Appendices

9.1 Appendix 1

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.

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, it is worth highlighting that 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 analyzing 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 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 atmosphere and the fruits from which the wine is derived. 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.

Despite the fact that discussing 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.

This guides us to the next topic which is love. Mainly in this category, 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 the topic 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. These poems serve as a means of celebrating the achievements and virtues of prominent figures within society.

Additionally, within the praise category, 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 very same 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 Figure 7. 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 Figure 8. 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.

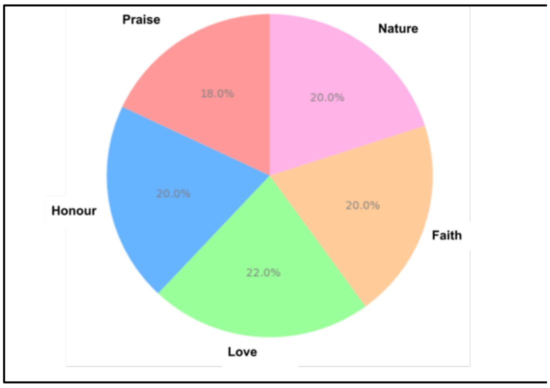

Figure 7: Topic distribution for the whole corpus.

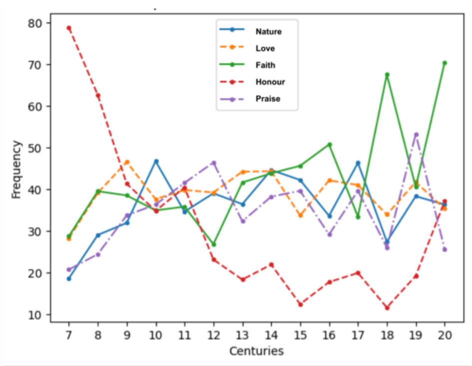

Figure 8: Topic distribution across centuries.

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. It's worth noting that 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.

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 regions 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.

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.

9.2 Appendix 2

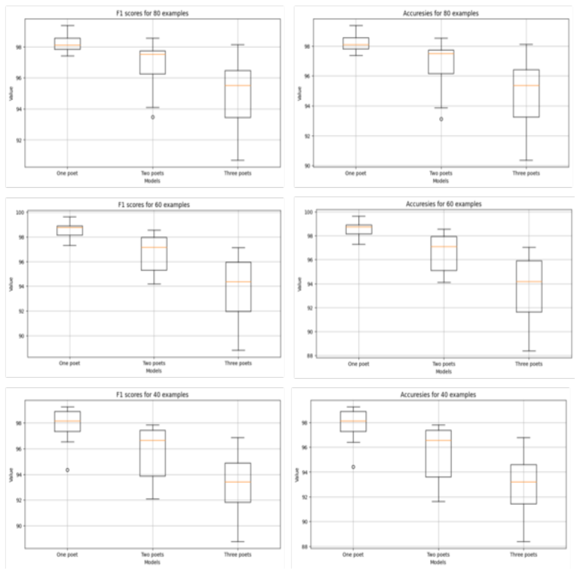

Figure 9: F1 scores and accuracies varying the number of poets in the opposite class.

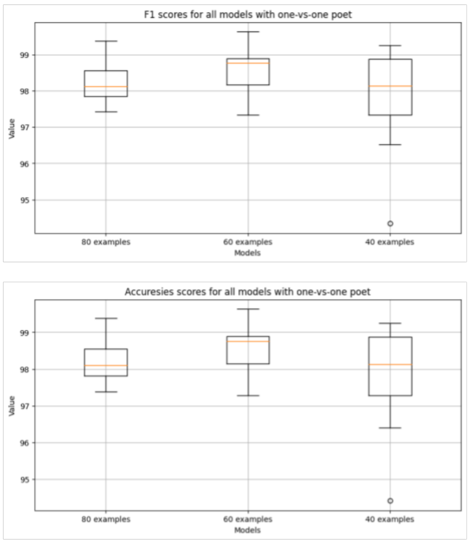

Figure 10: F1 scores and accuracies varying the number of training examples.