Computational Analysis of Quran Text using Machine Learning and Large Language Models

Usama Shahid
School of Business, Computing and Social Sciences
University of Gloucestershire
Cheltenham, United Kingdom
ushahid7@glos.ac.uk

Muhammad Zunnurain Hussain
Department of Computer Science
Bahria University
Lahore, Pakistan
zunnurain.bulc@bahria.edu.pk

Dr William Sayers
School of Business, Computing and Social Sciences
University of Gloucestershire
Cheltenham, United Kingdom
wsayers@glos.ac.uk

***Abstract*— The Quran verses are foundational for Muslims worldwide. Significant research has been dedicated to information retrieval (IR) from Quran; however, multiple studies have focused on descriptive analysis and topic modelling of the Quran in Arabic and translated versions. This study presents a comprehensive framework for analysing large textual data using an English translation of the Quran. Initially, it conducts a descriptive analysis of the verses to uncover various features, including readability, word clouds, significant n-grams, and network graphs illustrating word associations. The framework then applies machine learning techniques, specifically clustering models based on numerical vectors from *text-embedding-3-large*, to identify effective groupings of verses. Additionally, *GPT-4-turbo* is used for topic modelling within each cluster through prompt engineering, aiming to enhance the understanding of these clusters. The results include statistical information graphs and concise knowledge summaries that are beneficial to both domain experts and wider populace.**

Keywords*— *quran, data science, machine learning, large language models, natural language processing, text mining

I. INTRODUCTION

The Quran, regarded as the primary reference text for Muslims globally and originally composed in Arabic, holds significant value for scholars in religious studies and wider public. The analysis of Quranic verses provides substantial insights. Computational analysis, employing various statistical techniques and artificial intelligence (AI), facilitates the understanding of such data for different applications.

Information retrieval (IR) refers to the process of extracting pertinent information from an extensive repository. This process is crucial for accessing relevant data efficiently. Notably, significant research has been dedicated to information retrieval from the Quran [1], [2], [3], [4], [5], [6], [7], highlighting the importance of this field in understanding and utilising the Quran text.

Descriptive analysis, a branch of statistics that focuses on summarising and delineating the primary characteristics of a dataset, has been applied to the Quran's text. Research studies have conducted descriptive analysis of the Quran in Arabic [8], Malay [9], and Indonesian languages [10], [11], which enables the dissemination of knowledge for audience of respective language.

Topic modelling, a form of statistical analysis employed to discern abstract topics within a collection of documents. Research studies have employed the probabilistic topic modelling algorithm Latent Dirichlet Allocation (LDA) to identify topics in the Quran, primarily using base Arabic language [12], [13], [14], with one study in Indonesian translation [15]. This highlights LDA's capacity to manage

extensive text collections and uncover latent thematic structures within the documents. However, LDA demonstrates several drawbacks: it is sensitive to the number of topics chosen, it can be challenging to interpret the resulting topics, and it assumes that documents are mixtures of topics, which may not always accurately represent the true structure of the text.

Machine learning, a subset of AI, utilises mathematical techniques to learn from data. Within this domain, unsupervised methods automatically identify patterns in the data to create groupings, process referred to as clustering. In related work, Ahmad and associates undertook a series of clustering algorithms experiments on Quran verses from single chapter, suggesting k-means algorithm was notably effective [16]. Subsequently, they proceeded to analyse various modifications to the k-means algorithm, aiming to improve its performance [17]. While these studies offer insights into optimal algorithms, the understanding of the clusters themselves remains unclear.

Large language models (LLMs) are sophisticated statistical systems that learn from extensive datasets, demonstrating robust linguistic processing capabilities and performing various cognitive tasks [18], [19], [20]. Yousef and colleagues fine-tuned AraQA, an open-source language model trained on Arabic Islamic question-and-answer pairs sourced from reliable web platforms. This model answers queries related to Islamic topics in Arabic with a perplexity score of 2.3 [21]. This evidence suggests that LLMs can be effectively tailored to specific linguistic and cultural contexts, enhancing their relevance and accuracy in specialised domains.

To the best of current knowledge, no previous research has conducted computational analysis of the English translation of the entire Quran. This paper aims to develop a method for analysing large textual data and subsequently offering statistical information and abridged knowledge that may benefit researchers in this domain. Additionally, it seeks to establish a framework for future computational investigations. The study employs the Quran as a case study to achieve these objectives. Initially, descriptive analysis and clustering are conducted using various Python packages such as *py-readability-metrics*, *wordcloud, NetworkX* and *sklearn*, followed by the application of LLMs to examine the clusters.

The remainder of this paper is structured as follows: Section II elucidates the methodology employed in gathering verses of Quran and explores the analytical approaches applied therein. Section III presents the results obtained from the study. Finally, Section IV discusses results and Section V concludes the paper.

II. Methodology

The study utilises English translation of the Quran, authored by Talal Itani, scraped with the author's consent from clearquran website [22]. The study utilised the edition of the Quran in which the divine entity is referred to as 'God'. The research was conducted using the Python programming language due to the availability of open-source tools.

Fig. 1. Methodology

A. Data Pre-processing

The pre-processing procedure employed OpenAI's *text-embedding-3-large* model to generate numerical vectors for the verses [23]. Additionally, the research prepared an alternative version of the verses, wherein stopwords were removed according to the list provided by *nltk.corpus*, facilitating the frequency calculation of significant words.

B. Descriptive Analysis

The analysis utilised various visualisations with the help of open-source tools to disseminate the characteristics of the Quran verses.

1) Assessing Readability of Quranic Verses

The research employed the Flesch readability score from the *py-readability-metrics* library to assess the readability ease and grade level of the Quran verses [24].

2) Visualising Keywords in Quran with WordCloud

Word Cloud facilitate a comprehensive understanding of the most frequently occurring words. The research utilized the *wordcloud* library for visualization of word importance for entire Quran verses.

3) Analysing Word Frequency Pattern

The research employed the *TfidfVectorizer* from *sklearn* tools to calculate Unigram, Bi-gram, Trigram, and Four-gram analyses due to its ability to capture different levels of n-gram features, determining the frequency of the most commonly occurring words. The findings are presented using horizontal bar graphs.

4) Mapping Word Relationships with Network Graphs

The research entailed constructing network graphs to illustrate word associations within Quran verses. By analyzing bigram frequencies, a graph was devised to represent the relationships between frequently co-occurring words. The *NetworkX* library facilitated the creation of this graph from an edge list derived from the top 40 bigrams. The visualization, executed using a spring layout, elucidates the interconnectedness of key words. This approach underscores textual interrelations, thereby augmenting the interpretative analysis of the verses.

C. Grouping Similar Verses using Clustering Techniques

Clustering utilised embeddings created during the pre-processing phase. Principal Component Analysis (PCA) was employed to reduce the dimensions of the 3072 embeddings into a 2-dimensional space for visualisation and clustering purposes. PCA was selected due to its efficiency in dimensionality reduction for large datasets. The clustering algorithms applied included Spectral Clustering, Gaussian Mixture Model, and Agglomerative Clustering, chosen for their varied approaches to handling data complexity and structure.

Various clustering configurations, ranging from 2 to 10 clusters, were created, and the Silhouette Score was calculated for evaluation. This score aided in assessing the cohesion and separation of the clusters. The concept of the elbow diagram was then used to identify the optimal number of clusters, balancing simplicity, and explanatory power. Visualisations of the clustering results from all algorithms were compared side by side, evaluating both the Silhouette scores and the distribution created by the algorithms.

D. Exploring Clustered Topics with AI Analysis

The research utilized *GPT-4-Turbo* through OpenAI's paid API access, employing prompt engineering to model topics within each cluster of verses.

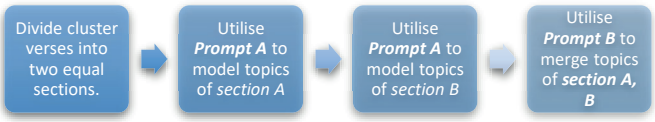

Fig. 2. Cluster Analysis Methodology

Fig 2. highlights cluster verses were divided into two equal sections because of the excessive length of the combined cluster text. This approach ensured that even lengthy cluster texts could be effectively managed and analysed, maintaining the coherence and interpretability of the results.

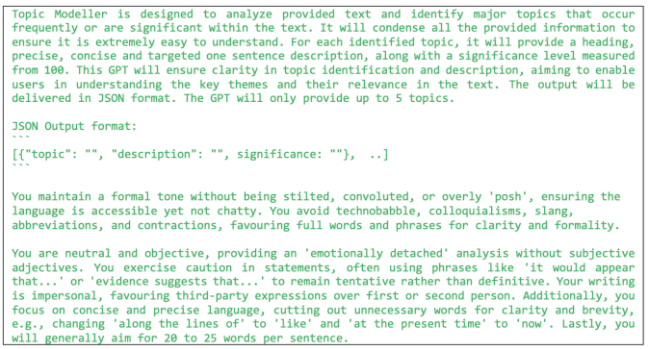

Fig. 3. Prompt A: Topic Modeller

Verses from both sections of the clusters were aggregated and presented to the model one by one using prompt A, as depicted in Figure 3. To enhance interpretability and organisation, the prompts instructed the language models to generate output in JSON format. This output consisted of a list of five topics, where each object included a topic, description, and significance.

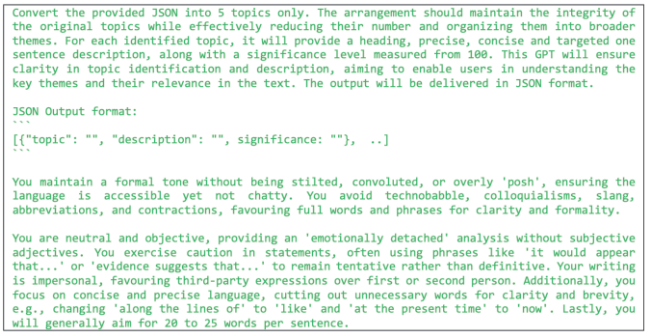

Fig. 4. Prompt B: Merged Modelled Topics

Finally, the modelled topics for both sections were merged with the help of *GPT-4-Turbo* using Prompt B, as depicted in Figure 4. Most instructions from Prompt A were incorporated into Prompt B to maintain consistency.

III. Results

The dataset comprised 6,236 rows, corresponding to the number of verses in the Quran, and included 113 chapters.

A. Assessing Readability of Quranic Verses

TABLE I. READABILITY ANALYSIS

Score	***Ease***	***Grade Level***
77.55	fairly_easy	7

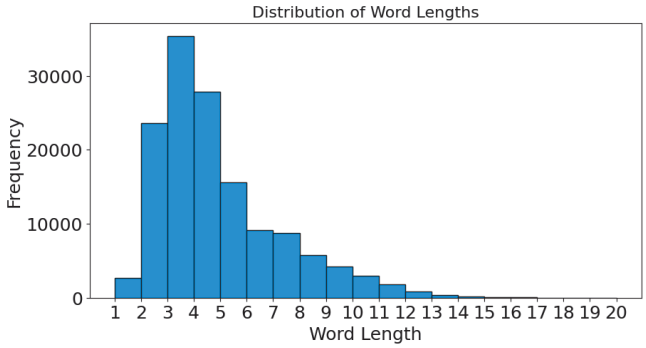

Fig. 5. Distribution of Word Length

B. Visualising Keywords in Quran with WordCloud

Fig. 6. Word Cloud for Quran Text

C. Analysing Word Frequency Pattern

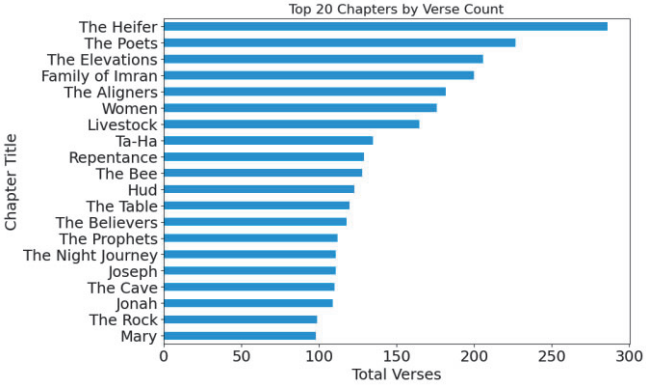

Fig. 7. Top 20 Chapters by Verse Count

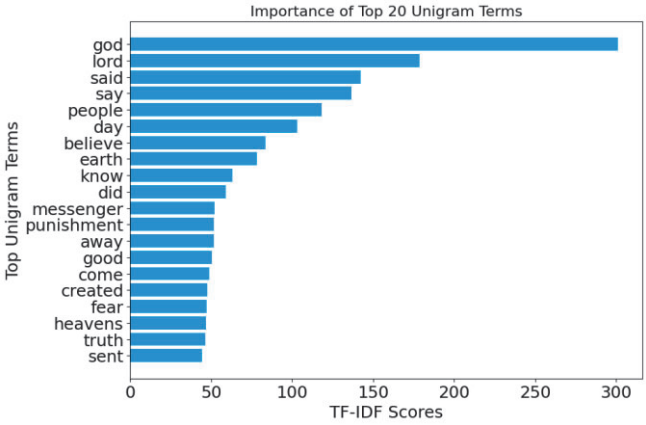

Fig. 8. Importance of Top 20 Uni-gram Terms

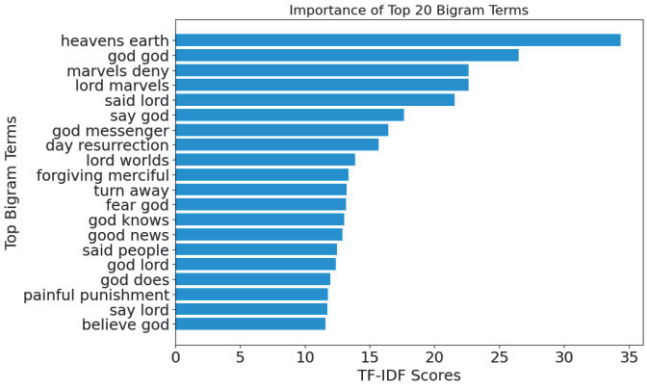

Fig. 9. Importance of Top 20 Bi-gram Terms

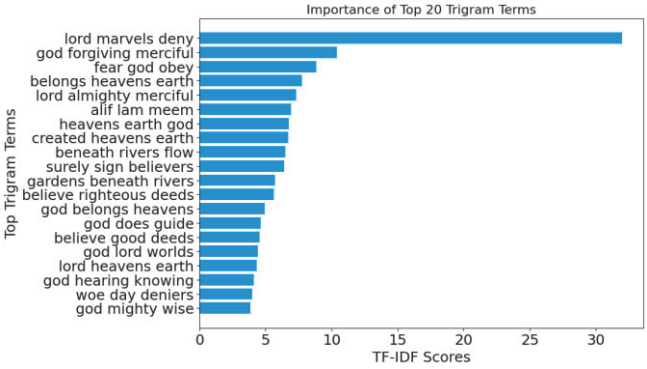

Fig. 10. Importance of Top 20 Tri-gram Terms

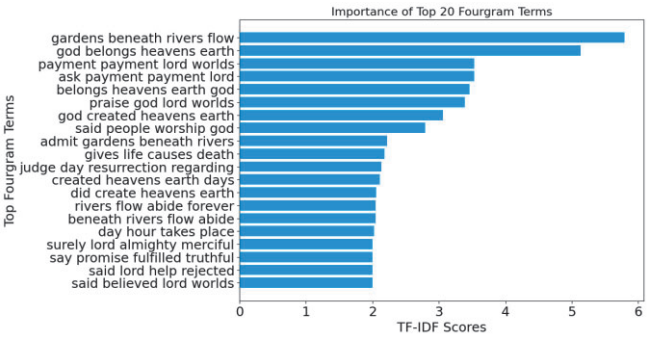

Fig. 11. Importance of Top 20 Four-gram Terms

D. Mapping Word Relationships with Network Graphs

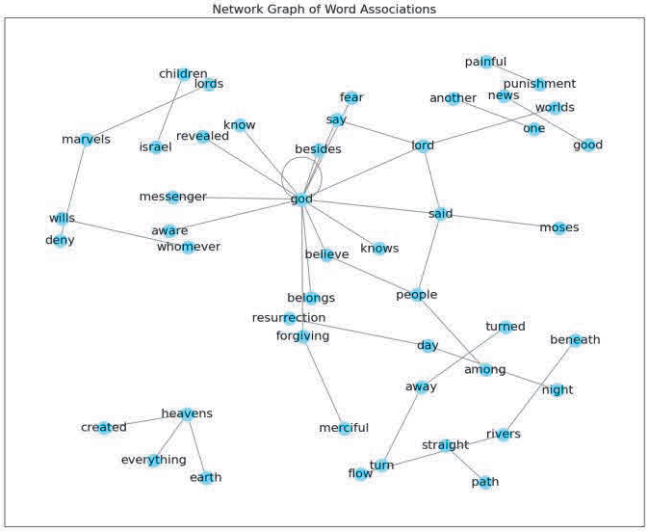

Fig. 12. Network graph of word associations

E. Grouping Similar Verses using Clustering Techniques

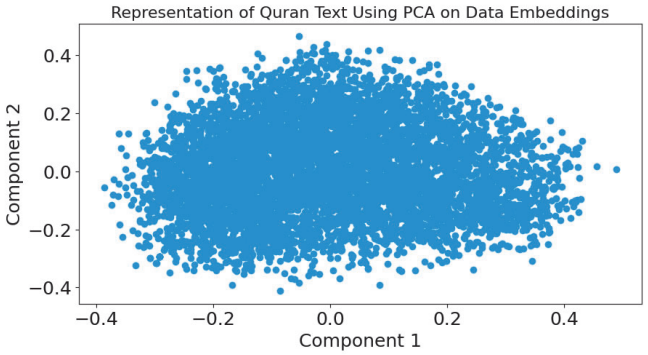

Fig. 13. PCA embeddings

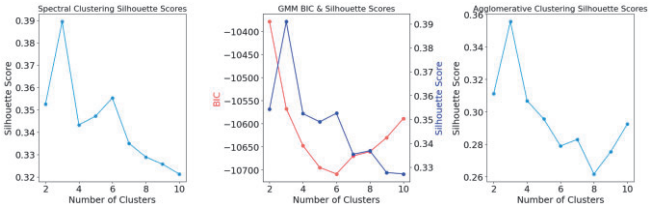

Fig. 14. Elbow digram with different cluster sizes and algorithms

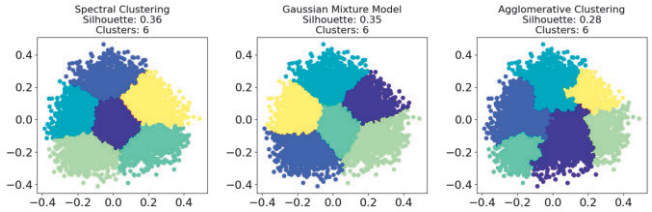

Fig. 15. Clustering outcome with six clusters

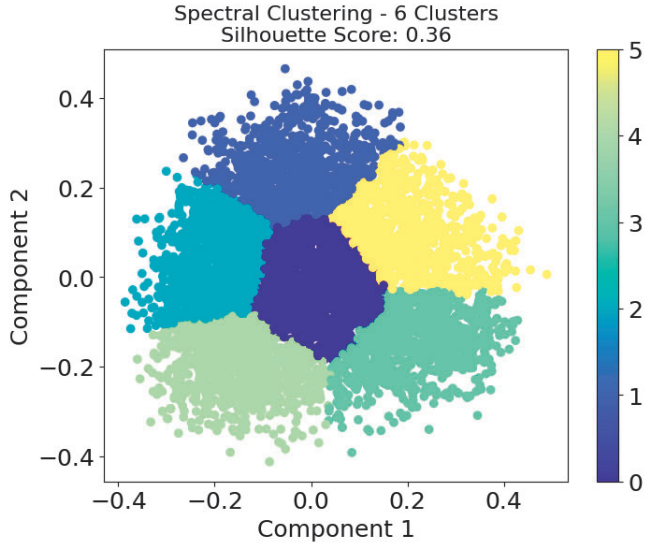

Fig. 16. Best clustering outcome with Spectral Clustering

TABLE II. DISTRIBUTION OF VERSES

Cluster Name	***Number of Rows***
First cluster	1366
Second cluster	934
Third cluster	1158
Fourth cluster	912
Fifth cluster	960
Sixth cluster	906

F. Exploring Clustered Topics with AI Analysis

Each cluster was processed in chronological order, with the results presented exactly as generated by the language models, without any alterations.

1) First Cluster

TABLE III. FIRST CLUSTER TOPICS

Topic	***Description***	***Sig***
Divine Judgment and Afterlife	The text underscores the certainty of resurrection, divine judgment, and the afterlife, emphasizing accountability for one's actions with corresponding rewards or punishments.	95
Prophetic Guidance and Revelations	Narratives of prophets like Noah, Abraham, Moses, and Jesus are presented, illustrating moral and spiritual lessons and the importance of obedience to divine directives.	90
Moral and Ethical Directives	The text offers explicit moral guidelines on justice, charity, honesty, and behaviour, aiming to shape individual conduct and societal norms.	85
Divine Omnipotence and Sovereignty	References to God's creation and control over the universe highlight divine omnipotence and sovereignty, reinforcing the purpose of human existence and accountability.	80
Consequences of Human Actions	The narrative contrasts the outcomes for piety versus disbelief, detailing the adverse consequences for denial and disobedience, including divine retribution.	75

2) Second Cluster

TABLE IV. SECOND CLUSTER TOPICS

Topic	***Description***	***Sig***
Divine Judgment and Consequences	The text underscores the severe repercussions for disbelief and sin, including divine retribution and eternal punishment in Hell.	95

Guidance and Spiritual Conduct	It provides comprehensive guidance on moral and spiritual rectitude, advocating for adherence to divine revelations and righteous living.	90
Prophetic Messages and Rejection	The narrative discusses the historical rejection of prophets and the dire consequences for communities that denied their messages.	85
Divine Mercy and Protection	Amidst warnings, the text highlights God's mercy and forgiveness for believers, promising divine protection and rewards in Paradise.	80
Faith Integrity and Accountability	The text distinguishes between true believers and hypocrites, emphasizing the importance of sincere faith and individual accountability.	75

3) Third Cluster

TABLE V. THIRD CLUSTER TOPICS

Topic	***Description***	***Sig***
Divine Guidance and Accountability	The text underscores the role of divine guidance through scriptures, stressing the importance of adhering to God's commands and the consequences of human actions in the afterlife, including rewards and punishments.	95
Moral and Ethical Instructions	It details moral and ethical behaviors such as charity, justice, and patience, advocating for adherence to these virtues as outlined in religious texts to promote societal harmony and personal salvation.	90
Principles of Worship and Piety	The significance of worship, prayer, and remembrance of God is emphasized as central practices that foster spiritual growth and provide	88

	a moral compass in believers' lives.	
Social Justice and Community Conduct	Guidance is provided on maintaining justice and proper conduct within the community, including fair treatment of vulnerable groups, which is deemed essential for a righteous society.	85
Prophets, Revelations, and Spiritual Growth	The text highlights the importance of prophets and their revelations as divine guides for mankind, linking past prophets to the continuity of God's message and emphasizing spiritual and moral conduct for communal well-being.	88

4) Fourth Cluster

TABLE VI. FOURTH CLUSTER TOPICS

Topic	***Description***	***Sig***
Cosmic and Divine Order	The text underscores a structured cosmic order and divine omnipotence, detailing celestial operations and the meticulous design of the universe.	95
Divine Communication and Prophetic Guidance	It highlights divine communication through prophets and scriptures, which guide humanity with wisdom and command obedience and faith.	90
Human Existence and Moral Testing	Human beings are portrayed as undergoing moral and ethical tests, with their choices having significant consequences, guided by divine wisdom.	88
Judgment and Eternal Afterlife	The narrative focuses on themes of judgment and the afterlife, emphasizing accountability and eternal outcomes based on earthly life.	90

Miracles and Divine Signs	Miraculous events and signs are depicted as divine interventions that affirm faith, challenge disbelief, and illustrate divine power.	85

5) Fifth Cluster

TABLE VII. FIFTH CLUSTER TOPICS

Topic	***Description***	***Sig***
Divine Sovereignty and Control	The text underscores God's absolute sovereignty and omnipotence, highlighting His role as the creator, sustainer, and ultimate authority over all existence.	95
Attributes and Actions of God	God is characterized by wisdom, mercy, and justice, engaging actively in the universe through guidance, revelations, and interactions with prophets.	90
Human Obligations and Relationships with God	The narrative explores human responsibilities like worship, obedience, and seeking forgiveness, emphasizing a reciprocal relationship of guidance and submission with God.	85
Concept of Judgment and Afterlife	The text discusses the Day of Judgment and the afterlife, stressing moral accountability and the consequences of one's actions on earth.	80
Prayer, Devotion, and Moral Conduct	Prayer and devotion are portrayed as vital for spiritual growth, with an emphasis on the importance of moral righteousness and the rejection of idolatry.	75

6) Sixth Cluster

TABLE VIII. SIXTH CLUSTER TOPICS

Topic	***Description***	***Sig***
Divine Judgment and Retribution	The text underscores themes of divine judgment and retribution, illustrating punishment for disbelief and	95

	wrongdoing with significant emphasis.	
Prophetic Guidance and Historical Lessons	Narratives focus on prophetic warnings, historical examples, and miracles to convey lessons and the consequences of ignoring divine messages.	90
Moral and Ethical Instructions	The text imparts moral and ethical lessons through parables and the actions of past communities, emphasizing righteousness and divine obedience.	90
Concept of Afterlife and Eternal Repercussions	Exploration of the afterlife, detailing judgment day events and eternal consequences for individuals based on their earthly actions.	88
Human Disbelief and Divine Omnipotence	Discussions highlight human skepticism towards divine signs and the omnipotence of the divine, stressing the folly of human hubris.	85

IV. Discussion

Table I indicates that the translation of the Quran employed in the study is comprehensible to a broader audience with limited English proficiency. Additionally, Figure 5 reveals that the most frequent word lengths range between three and five characters, while words exceeding ten characters are negligible. This information is useful for optimising text processing algorithms and enhancing computational efficiency in natural language processing tasks.

The word cloud of Quranic verses in Figure 6 facilitates the visual interpretation of frequently occurring words, aiding in the comprehension of themes within the text. Figure 7 presents the 20 longest chapters in the Quran, measured by the number of verses. This metric can potentially serve as a criterion for future research on specific chapters of Quran. The unigrams, bigrams, trigrams, and four-grams presented in Figures 8-11 demonstrate the most common n-grams in the Holy Quran. These terms and the word cloud are crucial for future research in semantic search because of insights provided into semantic structure and thematic elements of the verses. Their importance is underscored by the prior removal of stop words before extraction. Lemmatization and stemming techniques, often employed to reduce words to their base or root form for textual analysis, were deliberately not used to maintain the original context of the verses.

Figure 12, a network graph of word associations, visually illustrates the interconnectedness of key terms within the Quranic verses. Central to this graph is the word "God," which is directly linked to other significant concepts such as "believes," "knows," and "forgiving." This centrality suggests that these terms are crucial for understanding the thematic and semantic structure of the Quran. It is particularly useful for identifying core themes and relationships within the Quran, aiding researchers and readers in comprehending the complex interplay of concepts.

Figure 13 presents a two-dimensional representation of Quranic verses, utilising PCA on numerical embeddings. The scatter plot reveals a broad distribution with considerable overlap among data points, suggesting both diversity in the embeddings and significant commonality in their features. This visualisation underscores the complexity and rich, varied nature of the verses data.

Figure 14 displays the silhouette scores for Spectral Clustering, GMM, and Agglomerative Clustering across various cluster numbers. The figure suggests that the optimal number of clusters is 6, as this number results in the highest silhouette scores for both Spectral Clustering and GMM. In contrast, Agglomerative Clustering consistently yields lower silhouette scores compared to the other methods, indicating its reduced effectiveness in this context. Figures 15 and 16 further clarify these findings by presenting the silhouette scores and cluster distributions for the selected algorithms. Spectral Clustering achieves the highest silhouette score of 0.36, thus emerging as the most effective clustering method. The visual distribution of clusters supports this conclusion, showing distinct and well-separated clusters under Spectral Clustering.

Table II indicates low variation in the number of verses among clusters, suggesting homogeneity. However, it is anticipated that clusters should exhibit diverse topics with a substantial number of verses. Tables III-VIII identify key themes within each cluster identified by LLMs rather than using traditional methods like LDA to maximise interpretability and ensure consideration of contextual relevance of verses. Each topic offers a broad understanding, while the descriptions provide detailed insights. The significance (Sig) scores indicate the relative importance within each cluster. Although there are recurring themes across clusters, the descriptions highlight distinct nuances, enriching the overall analysis. The tables collectively underscore recurrent themes of divine judgment, moral and ethical guidance, prophetic teachings, and the significance of divine omnipotence. Each cluster offers a nuanced perspective, highlighting the depth and complexity of the themes within the verses.

V. Limitations

This study's limitations arise from its reliance on an English translation of the Quran, which may not fully convey the nuances inherent within the original Arabic. This risks losing critical meanings necessary for accurate interpretation, especially within religious contexts. Moreover, the study does not seek input from Islamic scholars to verify translation accuracy or methodology from a Sharia perspective, which raises questions regarding its scholarly rigour.

Moreover, the text-embedding model used for clustering was trained on general linguistic data, potentially lacking the domain-specific depth required for a thorough analysis of religious texts. This generalisation may produce clusters that inadequately capture the theological and contextual complexities of Quranic verses. Although machine learning models were employed for clustering, the interpretability of the resulting clusters could be improved. Furthermore, the study assumes statistic groupings of verses, thereby potentially overlooking fluid or overlapping themes across different clusters.

VI. Future work

Future research could investigate the use of fine-tuned language models trained specifically on religious texts to enhance clustering accuracy and interpretability. Integrating the original Arabic text with multilingual models may yield deeper insights into thematic structures. Moreover, employing dynamic topic models and temporal analysis could provide a more nuanced understanding of evolving Quranic themes across various clusters or chapters. Expanding the dataset to encompass additional translations and interpretations might also facilitate cross-linguistic comparisons, thereby enriching the overall findings.

VII. Conclusion

This study presents a computational framework for future investigations that is related to the application of natural language processing, data mining and qualitative analysis. Initially, this paper employs descriptive analysis to summarise and delineate the characteristics of Quranic verses from all chapters, utilising readability analysis, word cloud, n-grams, and network graph visualisations. Subsequently, clustering techniques are applied using numerical embeddings to group verses from all chapters, and then utilising capabilities of LLMs to identify topics within each cluster rather than relying on traditional methods like LDA to maximise interpretability and ensure consideration of contextual relevance of verses.

By generating word clouds and network graphs and extracting key terms—following the removal of stop words—researchers obtain valuable insights into the text's semantic structure. These insights are crucial for developing effective semantic search methodologies. This process helps identify core themes

and relationships within the Quran, facilitating a deeper understanding of the intricate interplay of concepts for both researchers and readers.

Each identified cluster within the text provides a nuanced perspective, highlighting the depth and complexity of Quranic themes. Although this study encompasses the entire Quran, the methodologies and findings can be applied effectively to individual chapters, enabling detailed analysis at a granular level. This approach will be instrumental in future research endeavours, allowing for targeted exploration of specific thematic elements within the Quran's chapters.

REFERENCES

[1] H. Ullah Khan, S. Muhammad Saqlain, M. Shoaib, and M. Sher, 'Ontology Based Semantic Search in Holy Quran', *IJFCC*, 2013, doi: 10.7763/IJFCC.2013.V2.229.

[2] N. Shahzadi, A. Atta-ur-rahman, and M. Jamil Sawar, 'Semantic Network based Classifier of Holy Quran', *IJCA*, vol. 39, no. 5, Feb. 2012, doi: 10.5120/4820-7069.

[3] M. Shoaib, M. N. Yasin, U. K. Hikmat, M. I. Saeed, and M. S. H. Khiyal, 'Relational WordNet model for semantic search in Holy Quran', in *2009 International Conference on Emerging Technologies*, Islamabad, Pakistan: IEEE, Oct. 2009. doi: 10.1109/ICET.2009.5353208.

[4] A. R. Yauri, R. A. Kadir, A. Azman, and M. A. A. Murad, 'Quranic Verse Extraction base on Concepts using OWL-DL Ontology', *RJASET*, vol. 6, no. 23, Dec. 2013, doi: 10.19026/rjaset.6.3457.

[5] M. A. Yunus, R. Zainuddin, and N. Abdullah, 'Semantic query for Quran documents results', in *2010 IEEE Conference on Open Systems (ICOS 2010)*, Kuala Lumpur: IEEE, Dec. 2010. doi: 10.1109/ICOS.2010.5719959.

[6] S. M. Alam, M. Ali, A. Khan, J. Khan, and F. Khan, 'A Framework for Unit of Interest Meta Data Analysis for Understanding the Message of Quran', in *2013 Taibah University International Conference on Advances in Information Technology for the Holy Quran and Its Sciences*, Madinah, Saudi Arabia: IEEE, Dec. 2013. doi: 10.1109/NOORIC.2013.70.

[7] M. Alqarni, 'Embedding Search for Quranic Texts based on Large Language Models', *IAJIT*, vol. 21, no. 2, 2024, doi: 10.34028/iajit/21/2/7.

[8] M. Alhawarat, M. Hegazi, and A. Hilal, 'Processing the Text of the Holy Quran: a Text Mining Study', *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 6, no. 2, Art. no. 2, 28 2015, doi: 10.14569/IJACSA.2015.060237.

[9] N. Kamal, N. Abdul Rahman, Z. Bakar, T. Muhammad, and T. Sembok, 'Terms Visualization for Malay Translated Quran Documents', Jan. 2007, [Online]. Available: https://www.researchgate.net/publication/228985856_Terms_Visualization_for_Malay_Translated_Quran_Documents

[10] S. J. Putra, T. Mantoro, and M. N. Gunawan, 'Text mining for Indonesian translation of the Quran: A systematic review', in *2017 International Conference on Computing, Engineering, and Design (ICCED)*, Kuala Lumpur: IEEE, Nov. 2017. doi: 10.1109/CED.2017.8308122.

[11] S. Raharjo and K. Mustofa, 'Visualization of Indonesian translation of Quran Index', in *2014 International Conference on Smart Green Technology in Electrical and Information Systems (ICSGTEIS)*, Kuta, Bali, Indonesia: IEEE, Nov. 2014. doi: 10.1109/ICSGTEIS.2014.7038735.

[12] M. A. Siddiqui, S. M. Faraz, and S. A. Sattar, 'Discovering the Thematic Structure of the Quran using Probabilistic Topic Model', in *2013 Taibah University International Conference on Advances in Information Technology for the Holy Quran and Its Sciences*, Madinah, Saudi Arabia: IEEE, Dec. 2013. doi: 10.1109/NOORIC.2013.55.

[13] M. Alhawarat, 'Extracting Topics from the Holy Quran Using Generative Models', *ijacsa*, vol. 6, no. 12, 2015, doi: 10.14569/IJACSA.2015.061238.

[14] M. Naili, A. Chaibi, and H. Ghézala, 'Arabic topic identification based on empirical studies of topic models', *Revue Africaine de Recherche en Informatique et Mathématiques Appliquées*, vol. Volume 27-2017-Special..., p. 3102, Aug. 2017, doi: 10.46298/arima.3102.

[15] D. Rolliawati, I. Rozas, K. Khalid, and I. Rozas, 'Text Mining Approach for Topic Modeling of Corpus Al Qur'an in Indonesian Translation', presented at the Proceedings of the 2nd International Conference on Quran and Hadith Studies Information Technology and Media in Conjunction with the 1st International Conference on Islam, Science and Technology,

ICONQUHAS & ICONIST, Bandung, October 2-4, 2018, Indonesia, May 2020. doi: 10.4108/eai.2-10-2018.2295559.
[16] M. A. Ahmed, H. Baharin, and P. Nohuddin, 'Analysis of K-means, DBSCAN and OPTICS Cluster Algorithms on Al-Quran Verses', *International Journal of Advanced Computer Science and Applications*, vol. 11, Jan. 2020, doi: 10.14569/IJACSA.2020.0110832.
[17] M. A. Ahmed, H. Baharin, and P. N. E. Nohuddin, 'K-means variations analysis for translation of English Tafseer Al-Quran text', *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 13, no. 3, Art. no. 3, Jun. 2023, doi: 10.11591/ijece.v13i3.pp3255-3265.
[18] T. B. Brown *et al.*, 'Language Models are Few-Shot Learners', Jul. 22, 2020, *arXiv*: arXiv:2005.14165. doi: 10.48550/arXiv.2005.14165.
[19] S. Bubeck *et al.*, 'Sparks of Artificial General Intelligence: Early experiments with GPT-4', Apr. 13, 2023, *arXiv*: arXiv:2303.12712. doi: 10.48550/arXiv.2303.12712.
[20] J. Kocoń *et al.*, 'ChatGPT: Jack of all trades, master of none', *Information Fusion*, vol. 99, p. 101861, Nov. 2023, doi: 10.1016/j.inffus.2023.101861.
[21] Y. Adel *et al.*, 'AraQA: An Arabic Generative Question-Answering Model for Authentic Religious Text', in *2023 11th International Japan-Africa Conference on Electronics, Communications, and Computations (JAC-ECC)*, Alexandria, Egypt: IEEE, Dec. 2023. doi: 10.1109/JAC-ECC61002.2023.10479645.
[22] T. Itani, *Quran in English - Clear and Easy to Read.* Accessed: Jun. 05, 2024. [Online]. Available: https://www.clearquran.com
[23] OpenAI, 'OpenAI Platform'. Accessed: Jun. 05, 2024. [Online]. Available: https://platform.openai.com
[24] Flesch–Kincaid, 'Flesch–Kincaid readability tests', *Wikipedia*. Mar. 19, 2024. Accessed: Jun. 05, 2024. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Flesch%E2%80%93Kincaid_readability_tests&oldid=1214583693