# Comparative Analysis of Religious Texts: NLP Approaches to the Bible, Quran, and Bhagavad Gita

**A D Mahit Nandan, Ishan Godbole, Pranav Kapparad, Dr. Shrutilipi Bhattacharjee**
Department of Information Technology
National Institute of Technology Karnataka
Surathkal, India
mahitnandanad.211ai001@nitk.edu.in, ishangodbole.211ai020@nitk.edu.in,
pranavkapparad.211ai026@nitk.edu.in, shrutilipi@nitk.edu.in

## Abstract

Religious texts have long influenced cultural, moral, and ethical systems, and have shaped societies for generations. Scriptures like the Bible, the Quran, and the Bhagavad Gita offer insights into fundamental human values and societal norms. Analyzing these texts with advanced methods can help improve our understanding of their significance and the similarities or differences between them. This study uses Natural Language Processing (NLP) techniques to examine these religious texts. Latent Dirichlet allocation (LDA) is used for topic modeling to explore key themes, while GloVe embeddings and Sentence transformers are used to comapre topics between the texts. Sentiment analysis using Valence Aware Dictionary and sEntiment Reasoner (VADER) assesses the emotional tone of the verses, and corpus distance measurement is done to analyze semantic similarities and differences. The findings reveal unique and shared themes and sentiment patterns across the Bible, the Quran, and the Bhagavad Gita, offering new perspectives in computational religious studies.

***Keywords*—** Religious texts, Natural Language Processing, topic modeling, sentiment analysis, corpus distance, Bible, Quran, Bhagavad Gita.

## 1 Introduction

Religious texts played a very crucial role in forming the moral and ethical frames of society and have, therefore, influenced the way societies changed with time. Such texts as the Bible, the Quran, and the Bhagavad Gita have been indispensable in providing core human conceptions of justice, morality, and common social values. These aspects have historical importance since they are in a position to shape the ethics of humans, introduce social norms, and provide philosophical and spiritual guidance that can be seen across various cultures and eras.

Despite their ancient origins, religious texts continue to adapt to modern interpretations, showing the shifting priorities and evolving ethics of societies. The need for analysis of these texts arises from the desire to further understand their long-lasting relevance, uncover hidden connections, and find how their teachings have influenced and been influenced by different historical and cultural contexts. Modern computational tools provide new ways to examine these texts, offering insights that go beyond traditional interpretations.

Natural Language Processing (NLP) techniques provide a systematic and data-driven approach to the analysis of religious literature. They further aid in exploring themes, patterns of sentiment, and semantic relationships that may not emerge from analyses conducted through conventional means. This is possible through advanced NLP tools that can draw cross-textual comparisons to trace the evolution of ideas across different religious traditions. This research applies NLP techniques, such as topic modeling, sentiment analysis, and corpus distance measurement, to investigate the Bible, the Quran, and the Bhagavad Gita. Topic modeling is used to identify key themes in texts; the distribution of sentiment within texts is explored; and semantic similarities and differences are analyzed through corpus distance analysis. This seeks to highlight the commonalities and unique aspects of these revered writings, providing a fuller understanding of their cultural and moral impacts.

In the subsequent sections, this paper studies the content of these scriptures and compares them to further establish their linguistic relationships. Such an approach aims to contribute to the wider field of computational religious studies by demonstrating how NLP techniques may help unveil new perspectives about religious thought and its cultural applicability.



*Proceedings of the New Horizons in Computational Linguistics for Religious Texts (Coling-Rel)*, pages 1–10
January 19, 2025. 

## 2 Literature Survey

Applying Natural Language Processing (NLP) to religious texts is now an important area of research, given the depth of context in these texts. Joulin et al. (Joulin et al., 2017) investigated the FastText model for text classification-a core method for representing words in continuous space. This technique has been found to be very effective in analyzing large religious corpora, such as the Bible and Quran, and is crucial in understanding word-level representations in sacred texts. In addition, Hutchinson et al. (Hutchinson, 2024) discusses the broader ethical considerations of applying NLP to religious corpora, such as the Bible and Quran, particularly focusing on tasks like machine translation and sentiment analysis. This therefore calls for sensitivity while handling the sacred writings in NLP.

Chandra et al. (Chandra and Ranjan, 2022) (Chandra and Kulkarni, 2022a) focused on the semantic and sentiment analysis of the Bhagwad Gita, and finding out its similarity with other texts in Hindu philosophy, like the Upanishads. Alhawarat et al (Alhawarat, 2015) used generative models to perform topic modeling on the Holy Quran, but failed to capture meaningful results. The research (Abbasi et al., 2022) on toxic language identification has explored the use of machine learning models to classify harmful comments in online communication. However, some models have shown bias by assigning high toxicity ratings to non-toxic comments containing identity-related descriptors. Studies have compared various word embeddings, such as GloVe, Word2Vec, and FastText, for multilabel toxic comment classification. These studies found that different embeddings influenced the classification results, highlighting their role in improving the accuracy of detecting toxic language while addressing potential biases. In (Chandra and Kulkarni, 2022b) explores the comparison of English translations of the Bhagavad Gita using deep learning-based semantic and sentiment analysis. Motivated by the limitations of traditional translations, the authors employ BERT, a state-of-the-art language model, to analyze sentiment and semantic similarity across different translations. By utilizing a hand-labelled sentiment dataset for tuning, the study demonstrates that despite variations in style and vocabulary, the core message conveyed in the translations remains largely consistent.

In recent times, advanced techniques like the Fréchet Inception Distance (FID), initially developed for image generation models, have been adapted to text corpora to quantify distribution differences between embeddings (Heusel et al., 2018). Similarly, the IBM CompCor framework has gained attention for its ability to compute corpus-level distances, offering a robust methodology for evaluating the overall divergence in linguistic features (Kour et al., 2022). State-of-the-art metrics, such as those based on contextual embeddings from transformer models like BERT, provide richer semantic comparisons by capturing nuanced meanings, enabling more accurate measurements of semantic mismatch (Reimers and Gurevych, 2019). These modern metrics show improved sensitivity in detecting distribution mismatches, while classical methods tend to be more sensitive to surface-level perturbations. The integration of both classical and modern approaches is still an open challenge; there is no standardized framework for interpreting and comparing the effectiveness of these measures. Thereby, this creates an incentive for evaluation measures that are interpretable for ranking semantic similarity and its underlying characteristics, an ongoing theme recently discussed in various works.

VADER (Valence Aware Dictionary and sEntiment Reasoner) (Hutto and Gilbert, 2014) was introduced as a robust, rule-based model for sentiment analysis, specifically designed for handling microblog like contexts. The foundational study demonstrated its effectiveness by combining a validated lexicon with syntactical rules, achieving high F1 classification accuracy that surpass human raters. While VADER's origin lies in social media analysis, its adaptability makes it suitable for examining complex texts, including religious scriptures. Subsequent studies have applied sentiment analysis techniques to religious texts to uncover patterns in sentiment and thematic frameworks. One study (Peuriekeu et al., 2021) employed NLP and machine learning techniques to classify nine sacred texts and found that methods like Multinomial Naive Bayes achieved significant accuracy. Another comparative analysis of the Bible, Quran, and Bhagavad Gita (Goel and Arsiwala, 2024), utilized NLP tools to perform sentiment analysis and topic modeling, shedding light on emotional distributions and similarities between these texts. These works illustrate how VADER and related sentiment analysis approaches can offer deeper insights into the emotional and philosophical nuances of religious literature.

## 3 Methodology

The following sections explain the series of analyses that, using different approaches, investigate linguistic, thematic, and semantic aspects of the Bhagavad Gita, Bible, and Quran and examine relationships and connections between them.

### 3.1 Word Count Analysis

To gain insights into the linguistic and thematic focus of the Bhagavad Gita, Bible, and Quran, a word frequency analysis was conducted. The texts were first preprocessed to ensure uniformity by converting all text to lowercase, removing punctuation, and filtering out common stop words. This preprocessing step was essential for eliminating noise and ensuring that the analysis concentrated on meaningful terms. Following this, tokenization was performed to break the text into individual words, which were then analyzed to compute their frequency. The top 20 most frequently occurring words in each text were determined, which highlights the recurring concepts and thematic emphasis. Visualizations, such as bar plots as shown in figures 6, 8 and 7, were created to show these findings, which provided a comparative view of word distribution across the three texts.

### 3.2 Topic Modeling Using Latent Dirichlet Allocation (LDA)

To uncover the thematic structure within the Bhagavad Gita, Bible, and Quran, Latent Dirichlet Allocation (LDA) was used for topic modeling. Each text underwent preprocessing to standardize the format, including steps like text normalization, converting to lowercase, and removing stop words. The LDA model was then applied for each text separately, set to extract 10 unique topics for each text. Each topic was represented by 10 prominent keywords that provided insight into the main themes as shown in figure 4. This approach aimed to reveal hidden structures and thematic patterns in the texts.

### 3.3 Topic Comparison Across Texts with GloVe Embeddings and Sentence Transformers

To explore the thematic similarities and differences between the Bhagavad Gita, Bible, and Quran, embeddings were used to quantify the semantic relationships among the topics generated through Latent Dirichlet Allocation (LDA). Each text was analyzed to extract 10 topics, represented by sets of 10 significant keywords. These keywords were then transformed into vector representations using two methods: GloVe embeddings and SentenceTransformer embeddings (paraphrase-MiniLM-L6-v2). The keywords for each topic were concatenated into a string and encoded into a single topic embedding. The semantic similarity between topics was assessed using cosine similarity, providing a measure of how closely related the topics were across the texts. Heatmaps were created to visualize these similarities, offering an intuitive representation of thematic overlap and divergence.

### 3.4 Sentiment Analysis of Verses Using VADER

Sentiment analysis is applied to the corpus using Valence Aware Dictionary and sEntiment Reasoner (VADER), which is an unsupervised model that classifies text into positive, negative, and neutral sentiments. VADER first breaks down the text into individual words, and assigns a score to each word based on its polarity, with -4 being the most negative and +4 being the most positive. It also considers the intensity of the sentiment, which can be indicated by capitalization and punctuation. For example, an exclamation point can make a positive word even more positive. The overall sentiment score of the text is calculated based on the scores assigned to each word. The score ranges from -1 to 1, with -1 being very negative and 1 being very positive.

### 3.5 Corpus Distance Measurement

The following types of corpus distance analyses were conducted:

#### 3.5.1 Corpus Distance Analysis Using IBM CompCor

The corpus distance between the religious texts was measured using IBM CompCor, with the Fréchet Inception Distance (FID) applied through the corpus_metrics.fid_distance function. This metric was used to quantify the distribution differences between the embeddings of the corpora, effectively capturing semantic and linguistic variations. The embeddings were generated using STTokenizerEmbedder, ensuring that the representations of the texts retained essential semantic features. Pairwise comparisons were conducted between the Bible, Gita, and Quran to assess the relative distances, with higher FID values indicating greater divergence in thematic and linguistic content. This ap-

proach provided a robust quantitative basis for analyzing how these influential texts differ in their language and semantics.

#### 3.5.2 Semantic Similarity Analysis Using Cosine Similarity

To assess the semantic similarity between the religious texts, the STTokenizerEmbedder with the model all-MiniLM-L12-v2 was utilized to generate embeddings for each corpus. The embeddings were averaged to create a representative vector for each text set: the Bible, Gita, and Quran. Using these averaged embeddings, pairwise cosine similarities were computed with cosine_similarity from the sklearn.metrics.pairwise module. This method measured how closely aligned the corpora were in terms of semantic content, with values ranging from -1 (completely dissimilar) to 1 (identical). The results provided a direct comparison of the similarity between the religious texts, revealing the degrees of linguistic and thematic alignment between them.

#### 3.5.3 Structural Relationship Analysis Using KMeans Clustering

To further investigate the structural relationships between the religious texts, KMeans clustering was applied to the embeddings generated using STTokenizerEmbedder with the all-MiniLM-L12-v2 model. The embeddings for the Bible, Gita, and Quran were clustered into 5 groups, which allowed the identification of central semantic themes within each corpus. The methodology is as below:

1. Clustering: Each corpus was clustered separately using KMeans with 5 clusters (modifiable based on corpus characteristics). The cluster centroids represented the core semantic centers of the texts.
2. Centroid Similarity: Cosine similarity and Euclidean distance metrics were calculated between the centroids of different corpora to quantify their relative alignment and semantic differences.
3. Comparison Function: A function iterated over each centroid pair between two corpora to compute their similarity or distance, providing an average measure for each pairwise corpus comparison.

## 4 Experimental Results

The above mentioned mehthodology were applied to the English translations of three religious texts - the Bible, the Quran, and the Bhagavad Gita. The results for the various experiments are mentioned in the following sections.

### 4.1 Word Count Analysis

The analysis revealed distinct linguistic focuses for each of the religious texts, highlighting their unique thematic elements:

1. Bhagavad Gita: The analysis showed that terms like arjuna (119 occurrences), action (80), and krishna (68) were predominant, emphasizing the philosophical dialogues between Arjuna and Krishna centered around ethical dilemmas, self-realization, and duty. Words such as mind (63), desire (53), and supreme (54) further underscored the text's spiritual and psychological dimensions.

2. Bible: High-frequency terms included give (8847), lord (7887), and god (4558), highlighting themes of divine instruction, moral teachings, and covenant relationships. Other frequently appearing words such as king (3079), man (3055), and people (2779) reflected the narrative and historical aspects of the text, illustrating societal and human interactions.

3. Quran: The word allah appeared most frequently (2833 occurrences), followed by lord (1014) and believe (525), emphasizing the text's focus on faith and divine will. Additional terms such as day (551), messenger (361), and people (256) indicated eschatological themes and the role of the prophetic tradition.

These results were visualized using bar plots in figures 6, 7 and 8, enabling a clear comparison of the top 20 words from each text. The visual representations provided a comprehensive perspective on both shared and distinctive linguistic emphases among the texts, enhancing the understanding of their thematic structures.

### 4.2 Topic Modeling Using Latent Dirichlet Allocation (LDA)

The application of LDA to the three religious texts yielded informative which can be seen in figure 4:

1. Bhagavad Gita: The topics derived included keywords such as duty, soul, wisdom, battle, and divine, which highlighted the philosophical and spiritual discourse central to

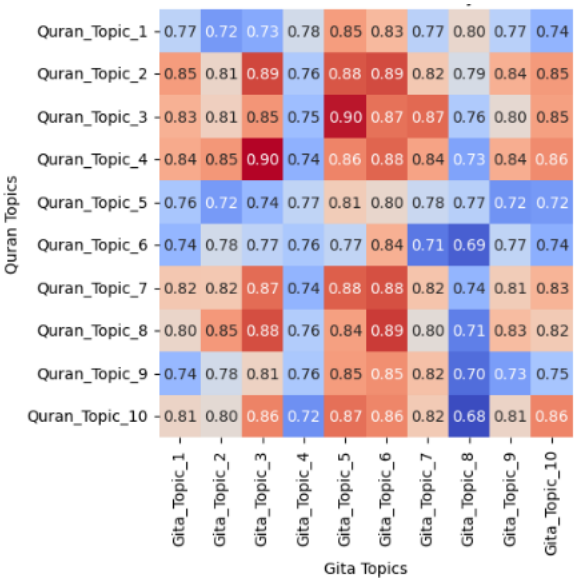


Figure 1: Topic Similarity between Quran and Gita using GLOVE

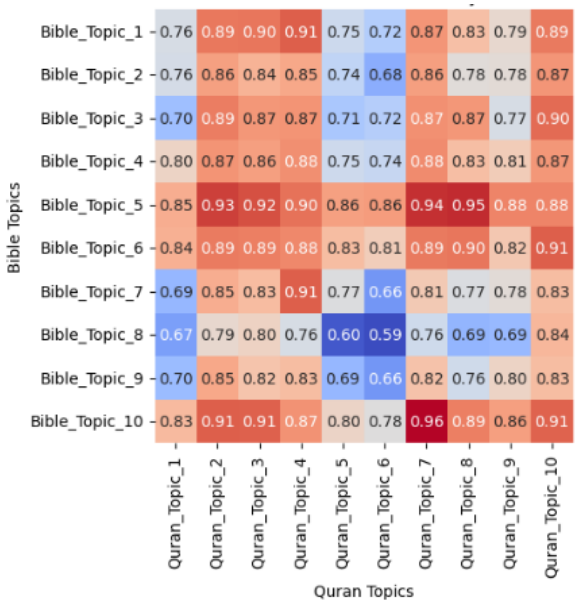


Figure 2: Topic Similarity between Bible and Quran using GLOVE

the dialogue between Krishna and Arjuna. The themes focused on moral duties, self-realization, and ethical dilemmas.

2. Bible: The identified topics featured words like covenant, kingdom, prophets, faith, and law, emphasizing its broad narrative of divine instructions, historical accounts, and religious teachings. This reflected the Bible's multi-faceted nature, combining guidance, historical context, and spiritual lessons.

3. Quran: Topics contained terms such as mercy, believers, guidance, prophet, and judgment, showcasing its focus on the divine message, moral conduct, and eschatological themes. The recurring emphasis was on faith, righteous living, and the role of prophets in guiding believers.

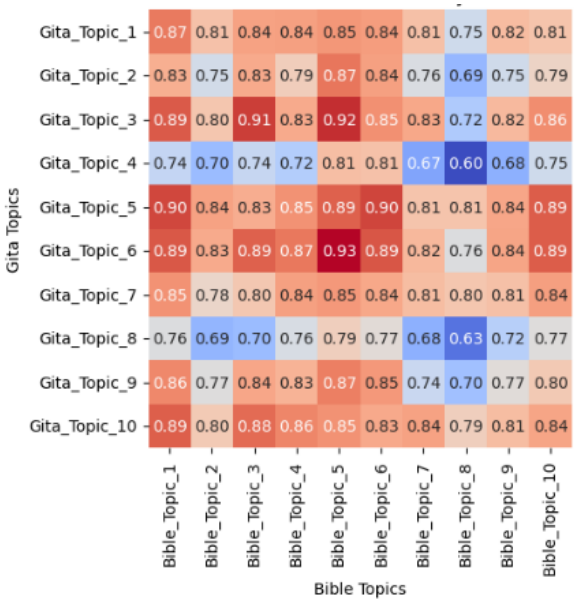


Figure 3: Topic Similarity between Gita and Bible using GLOVE

### 4.3 Topic Comparison Across Texts with GloVe Embeddings and Sentence Transformers

In this section, we discuss the results obtained form the pairwise comparison of the texts using Glove Embeddings and Sentence Transformers.

#### 4.3.1 GLOVE

We can draw the following inference from the heatmaps show in figures 1, 2 and 3.

1. Quran vs. Gita: The Quran and the Gita both emphasize themes of spiritual guidance and moral conduct. Topics in the Quran like Quran Topic 2 and Quran Topic 7 align with Gita Topic 4 and Gita Topic 3, where terms such as "lord," "believe," and "path" reflect a shared focus on enlightenment and ethical living. The strongest overlap occurs between Quran Topic 4 and Gita Topic 4, where keywords like "evil," "heaven," and "path" suggest common views on cosmic order and morality. Differences emerge in Quran Topic 5 and Gita Topic 5, with the Quran addressing themes of earth and doom, while the Gita centers on

| Topic Number | Gita Topics | Bible Topics | Quran Topics |
|---|---|---|---|
| 1 | arjuna, mind, world, krishna, great, attain, self, work, thus, body | death, man, without, power, dead, right, take, life, world, evil | allah, lord, seek, ever, upon, forgive, mercy, merciful, wherein, near |
| 2 | action, worship, among, path, self, know, attachment, yoga, wisdom, desire | go, land, come, glory, town, great, jerusalem, waste, king, take | lord, turn, good, save, away, none, hath, believe, work, favour |
| 3 | sense, mind, without, free, even, good, meditation, offer, action, evil | take, food, give, true, good, man, need, wealth, much, fruit | lord, verily, say, among, another, folk, believe, man, destroy, see |
| 4 | arjuna, attain, brahman, desire, path, knowledge, krishna, time, remember, pleasure | make, holy, take, body, priest, part, lord, place, unclean, every | evil, create, heaven, ward, light, like, use, whose, path, hear |
| 5 | arjuna, among, krishna, path, time, death, seek, god, know, life | give, god, lord, thing, faith, say, word, make, spirit, keep | earth, allah, doom, lord, heavens, fire, convey, good, disbelieve, owner |
| 6 | life, arjuna, spiritual, even, wisdom, free, offer, every, action, god | father, give, love, make, brother, desire, god, name, clear, christ | allah, messenger, religion, whoso, keep, duty, hath, disbeliever, well, promise |
| 7 | arjuna, world, among, creature, three, krishna, describe, listen, divine, every | like, fire, make, foot, earth, though, heaven, water, beast, round | say, worship, know, would, ease, lord, surely, moses, send, hand |
| 8 | arjuna, supreme, creature, krishna, goal, attain, lord, self, wise, action | child, number, israel, son, thousand, hundred, four, little, family, twelve | allah, hath, give, heart, reveal, concern, believe, scripture, good, make |
| 9 | supreme, self, within, lord, without, desire, free, selfish, attachment, knowledge | day, time, righteousness, first, come, year, light, till, rule, seven | day, allah, bring, night, people, disbelieve, scripture, naught, hell, forth |
| 10 | every, selfless, body, give, service, other, bear, without, creature, arjuna | say, lord, come, give, jesus, king, word, go, take, send | say, give, orphan, wife, find, make, garden, day, thereof, father |

Figure 4: Comparison of topics from the Gita, Bible, and Quran by topic number

personal spiritual pursuits involving Arjuna and Krishna.

2. Bible vs. Quran: The Bible and the Quran share strong themes of faith and devotion. Bible Topic 5 and Quran Topic 5 both highlight "god," "faith," and "lord," reflecting a mutual emphasis on submission and belief. Bible Topic 10 and Quran Topic 7 also align, with figures like "jesus" and "moses" underscoring themes of divine communication through prominent religious figures. However, Bible Topic 8, focused on lineage (e.g., "child," "israel"), differs from Quran Topic 8, which emphasizes universal faith themes, highlighting the Bible's specific cultural focus.

3. Gita vs. Bible: The Gita and the Bible similarly stress mental discipline, faith, and virtue. Gita Topic 3 and Bible Topic 5 share themes of mental control and faith, with terms like "mind," "free," and "spirit." The strongest overlap is between Gita Topic 6 and Bible Topic 6, which emphasize spirituality and love through terms such as "life," "wisdom," and "brother." Some topics differ, such as Gita Topic 4 and Bible Topic 8: the Gita focuses on

self-realization, while the Bible emphasizes lineage, showcasing different cultural narratives.

#### 4.3.2 Sentence Transformer

We can draw the following inference from the heatmaps show in figures 9, 10 and 11, as presented in the appendix.

1. Quran vs. Gita: The Quran and Gita intersect on themes of divine guidance and spiritual reflection. Quran Topic 1 and Gita Topic 4 focus on seeking and attaining divine wisdom. Quran Topic 6 and Gita Topic 6 both mention duty and spiritual practices. Differences arise with Quran Topic 8's focus on scripture compared to Gita Topic 5's take on life and spiritual transformation.

2. Bible vs. Quran: The Bible and the Quran share thematic overlaps, particularly in themes of divine principles and moral guidance. Bible Topic 5 and Quran Topic 6 highlight "lord," "faith," and "god," reflecting devotion to a higher power. Bible Topic 10 and Quran Topic 1 align on divine will and revelations through prophets. However, Bible Topic 8's focus on lineage contrasts with the Quran's universal themes, as seen in Quran Topic 3, centering on communal beliefs.

3. Gita vs. Bible: The Gita and Bible align on spiritual conduct and duty. Gita Topic 1 and Bible Topic 5 share themes of moral responsibility and divine wisdom. Gita Topic 3 and Bible Topic 3 both touch on "mind" and "action," highlighting the pursuit of righteousness. Distinctions include Bible Topic 8's lineage emphasis versus Gita Topic 7's philosophical discussions on divine duties.

### 4.4 Sentiment Analysis of Verses Using VADER

Using the VADER sentiment analysis, we obtain the sentiment scores for each verse of each book. Figure 5 shows the distribution of sentiment scores for the three religious texts. Here, we should note that the sentiment scores do not represent the moral positive or negative sentiment, but rather the use of positive or negative words, regardless of the context they are used in. From the figure, we see that the Bible and the Quran have a higher percentage of neutral verses, while the Gita has a higher percentage of positively classified verses. This can be interpreted the Gita uses a higher percentage of words that have positive sentiments, regardless of the context they are used in.

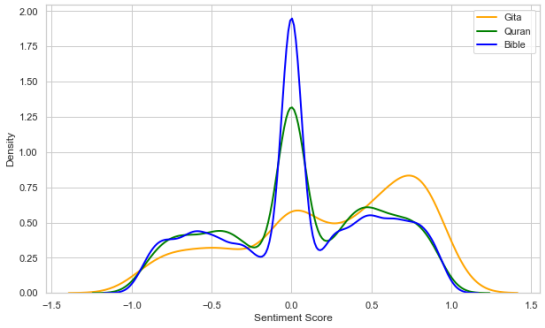


Figure 5: Distribution of sentiment scores of religious texts

### 4.5 Corpus Distance Measurement

The analysis of corpus distances and similarities between the religious texts revealed several key insights. The FID distance (Table 1), indicated that the Bible vs Quran pair had the closest distribution similarity (0.1664), while the Bible vs Gita and Gita vs Quran comparisons showed higher values, suggesting more divergence. In terms of semantic similarity, cosine similarity of averaged embeddings (Table 2) showed the Bible vs Quran pair had the highest similarity (0.8980), with the Bible vs Gita and Gita vs Quran pairs demonstrating lower similarities. KMeans clustering further reinforced these findings, with centroid based cosine similarity (Table 3, 4) being highest for Bible vs Quran (0.7035) and the lowest Euclidean distance (0.4170), indicating closer semantic structures, while the Bible vs Gita comparison showed greater dissimilarity.

Table 1: FID distance (no embeddings)

| Comparison | FID Distance |
|---|---|
| Bible vs Gita | 0.3868 |
| Bible vs Quran | 0.1664 |
| Gita vs Quran | 0.3801 |

Table 2: Cosine similarity (embeddings averaged)

| Comparison | Cosine Similarity |
|---|---|
| Bible vs Gita | 0.7278 |
| Bible vs Quran | 0.8980 |
| Gita vs Quran | 0.7757 |

Table 3: Average cosine similarity between Centroids

| Comparison | Avg. Cosine Sim. |
|---|---|
| Bible vs Gita | 0.5423 |
| Bible vs Quran | 0.7035 |
| Gita vs Quran | 0.6226 |

Table 4: Average euclidean distance between Centroids

| Comparison | Avg. Euclidean Dist. |
|---|---|
| Bible vs Gita | 0.5284 |
| Bible vs Quran | 0.4170 |
| Gita vs Quran | 0.4911 |

## 5 Conclusion and Future Work

Natural Language Processing (NLP) techniques have proven to be highly effective in analyzing and comparing the thematic and emotional dimensions of the Bible, Quran, and Bhagavad Gita. By utilizing methods such as Latent Dirichlet Allocation (LDA) for topic modeling, VADER for sentiment analysis, and semantic comparison through embeddings, distinct linguistic focuses and shared themes have been uncovered across these sacred texts. The analysis highlights unique philosophical dialogues and moral imperatives in each text while also revealing common ethical frameworks and human values that connect them. This research offers valuable insights into how religious texts continue to influence and shape cultural, ethical, and moral systems in contemporary society.

For further understanding of religious texts, future research could expand the scope by including additional texts from diverse religious traditions. Advanced NLP techniques, such as transformer-based models for sentiment analysis and topic modeling, could offer more nuanced insights into the complex language and meanings within these scriptures. Additionally, alternative text comparison metrics like BLEU and METEOR could be explored to enhance analysis. Furthermore, longitudinal studies examining the evolving interpretations of these texts over time, particularly in response to societal changes, could provide a broader view of their continued relevance. Interdisciplinary collaborations combining theology, linguistics, and cultural studies would also enhance the depth and scope of computational religious studies, offering richer perspectives on the texts' ongoing influence.

## 6 Limitations and Ethical Considerations

This study relies on the English translations of the three religious texts, which may introduce inherent biases due to potential variations in translation. Additionally, the use of pre-trained models (such as VADER, GloVe, Sentence Transformer, etc.) may further contribute to biases in the obtained results. The sentiment values obtained using VADER are also not representative of the moral sentiment of the verses in the text, but only based on the polarity of the words in them. Hence, while the findings provide meaningful insights, they should be interpreted with caution, considering these limitations.

In conducting research on religious texts using NLP techniques, it is imperative to approach the work with cultural sensitivity and respect for the diverse beliefs and values these texts embody. Efforts have been made to ensure that the analyses, including topic modeling and sentiment analysis, are presented objectively, without bias or misrepresentation of the texts' meanings or contexts. The results are intended solely for academic exploration and understanding, not for promoting ideological agendas or controversial interpretations. Researchers acknowledge the profound significance of these texts to their adherents and have prioritized ethical use and dissemination of findings to avoid any misuse or miscommunication.

## References


Ahmed Abbasi, Abdul Rehman Javed, Farkhund Iqbal, Natalia Kryvinska, and Zunera Jalil. 2022. Deep learning for religious and continent-based toxic content detection and classification. *Scientific Reports*, 12(1):17478.

Mohammad Alhawarat. 2015. Extracting topics from the holy quran using generative models. *International Journal of Advanced Computer Science and Applications*, 6(12):288–294.

Rohitash Chandra and Venkatesh Kulkarni. 2022a. Semantic and sentiment analysis of selected bhagavad gita translations using bert-based language framework. *IEEE Access*, 10:21291–21315.

Rohitash Chandra and Venkatesh Kulkarni. 2022b. Semantic and sentiment analysis of selected bhagavad gita translations using bert-based language framework. *IEEE Access*, 10:21291–21315.

Rohitash Chandra and Mukul Ranjan. 2022. Artificial intelligence for topic modelling in hindu philosophy: Mapping themes between the upanishads and the bhagavad gita. *Plos one*, 17(9):e0273476.

Pramit Goel and Rashida Arsiwala. 2024. A comparative study of religious scriptures using natural language processing. *Journal of Student Research*, 13(2).

Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. 2018. Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Preprint*, arXiv:1706.08500.

Ben Hutchinson. 2024. Modeling the sacred: Considerations when using religious texts in natural language processing. In *Findings of the Association for Computational Linguistics: NAACL 2024*, pages 1029–1043.

Clayton Hutto and Eric Gilbert. 2014. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the international AAAI conference on web and social media*, volume 8, pages 216–225.

Armand Joulin, Edouard Grave, Piotr Bojanowski, and Tomas Mikolov. 2017. Bag of tricks for efficient text classification. *arXiv preprint arXiv:1607.01759*.

George Kour, Samuel Ackerman, Eitan Farchi, Orna Raz, Boaz Carmeli, and Ateret Anaby-Tavor. 2022. Measuring the measuring tools: An automatic evaluation of semantic metrics for text corpora. In *Proceedings of the 2nd Workshop on Natural Language Generation, Evaluation, and Metrics (GEM 2022)*. Association for Computational Linguistics.

Younous Mofenjou Peuriekeu, Victoire Djimna Noyum, Cyrille Feudjio, Alkan Goktug, and Ernest Fokoue. 2021. A text mining discovery of similarities and dissimilarities among sacred scriptures. *arXiv preprint arXiv:2102.04421*.

Nils Reimers and Iryna Gurevych. 2019. Sentence-BERT: Sentence embeddings using Siamese BERT-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 3982–3992, Hong Kong, China. Association for Computational Linguistics.

# A Appendix

Figures 6, 7, 8 show the word counts of the three religious texts.

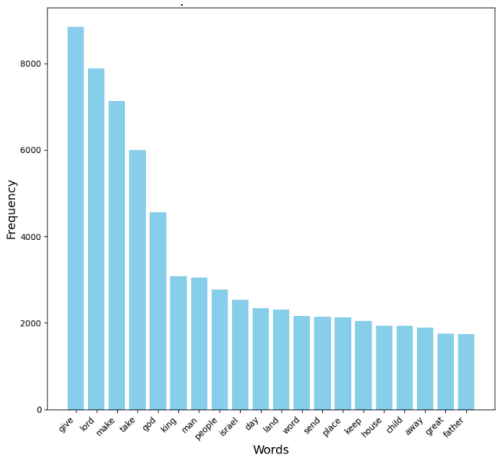


Figure 6: Word count for Bible

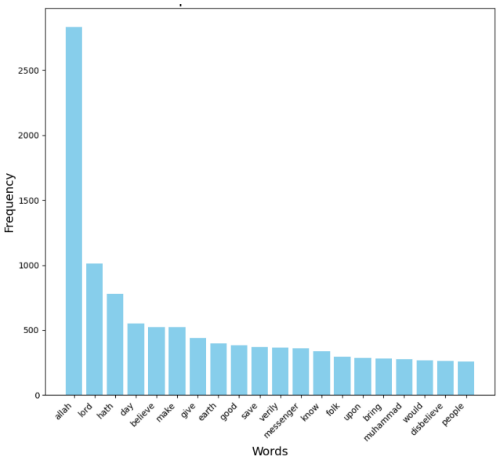


Figure 7: Word count for Quran

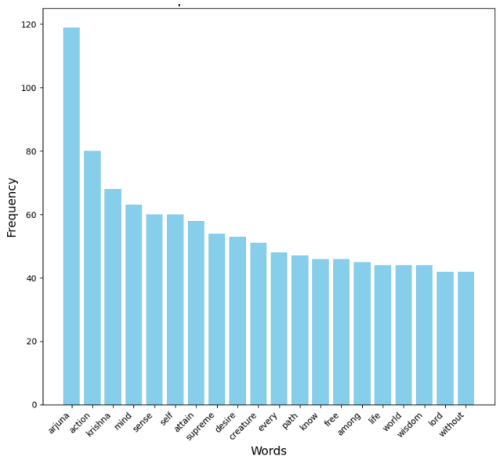


Figure 8: Word count for Bhagavad Gita

Figures 9, 10, 11 show the topic similarity between religious texts using Sentence Transformer

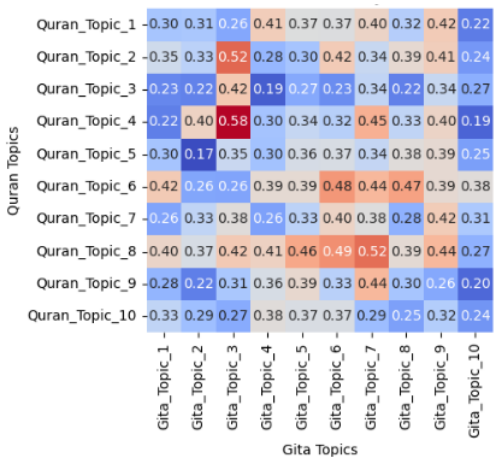


Figure 9: Topic similarity between Quran and Gita using Sentence Transformer

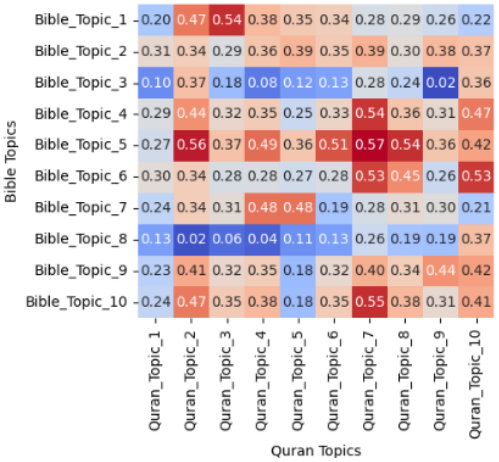


Figure 10: Topic similarity between Bible and Quran using Sentence Transformer

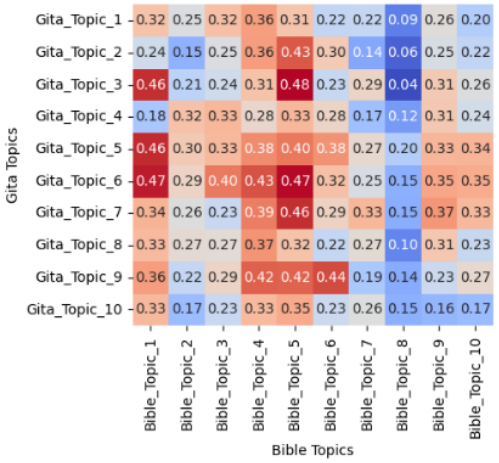


Figure 11: Topic similarity between Gita and Bible using Sentence Transformer