INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY
IJARIIT
ISSN: 2454-132X
Impact Factor: 6.078
(Volume 11, Issue 5 - V11I5-1205)
Available online at: https://www.ijariit.com

Text Mining and Sentiment Analysis of Major Religious and Philosophical Texts- Applying Natural Language Processing to Uncover Linguistic Patterns, Thematic Elements, and Emotional Tone

Sohan Sai Yerragunta
ysohansai@gmail.com
Rouse High School, Leander, Texas

ABSTRACT

This research uses natural language processing (NLP) methodologies to quantitatively analyze key religious and philosophical texts by identifying language trends, themes, and sentiment. Using a combination of text-mining techniques, topic modeling, and sentiment/emotion analysis, we evaluate how ideas, values, and emotions are conveyed within religious and philosophical traditions, including the Bible, Quran, Bhagavad Gita, and classic philosophy texts. The research analyzes publicly available text corpora and translations to quantify word counts, identify topic trends, and analyze emotional trajectories across chapters and verses. The comparative analysis reveals differences in thematic focus, emotional tone, and rhetorical style across religious and philosophical texts, and across translations of the same texts. The study's aim is to show the efficacy of computational methods as a complement to traditional textual scholarship by developing new ways to analyze form, sentiment, and meaning of primary texts. The interdisciplinary study and research also aim to contribute to emerging dialogue between the fields of digital humanities, linguistics, and religious studies to provide frameworks for large-scale, digital, and data-based analysis of sacred texts and literature.

Keywords: *Natural Language Processing (NLP), Text Mining, Sentiment Analysis, Topic Modeling, Religious Texts, Philosophical Texts, Digital Humanities, Linguistic Patterns, Emotional Tone, Thematic Analysis, Sacred Scriptures, Comparative Textual Analysis, Computational Text Analysis, Bible, Quran, Bhagavad Gita, Classical Philosophy, Transformer Models, Latent Dirichlet Allocation (LDA), Emotion Detection, Data-Driven Hermeneutics, Textual Scholarship, Interdisciplinary Research.*

INTRODUCTION

Religion and philosophy have long shaped the moral, cultural, and intellectual foundations of human civilizations, with their texts serving as the repositories of profound ideas, values, and ethical frameworks. Traditionally, the interpretation of sacred scriptures and philosophical writings has relied on hermeneutics, close reading, and historical contextualization. While such approaches remain vital, the exponential growth of digital tools offers new opportunities to explore these texts on a broader scale and with greater precision. Advances in natural language processing (NLP) and computational text analysis enable scholars to uncover linguistic structures, thematic elements, and sentiment patterns otherwise obscured in manual reading. This research aligns with the expanding field of digital humanities, which integrates quantitative methods with humanistic inquiry to generate fresh insights into enduring texts. By applying text mining, sentiment analysis, and topic modeling to canonical works such as the Bible, Quran, Bhagavad Gita, and major philosophical writings, this study aims to bridge computational analysis with interpretive scholarship. The goal is not to supplant traditional close reading, but to complement it with scalable, data-driven methods that reveal how language, emotion, and themes are embedded within these influential works. In doing so, the project contributes to interdisciplinary conversations across linguistics, religious studies, philosophy, and digital humanities, offering innovative frameworks for understanding how texts transmit meaning, engage audiences, and sustain their cultural relevance over time.

LITERATURE REVIEW

The application of natural language processing (NLP) and text mining to religious and philosophical texts has gained momentum as digital humanities intersect with traditional interpretive scholarship. Early studies, such as McDonald (2014), demonstrate how automated lexical analysis and self-organizing maps (SOM) can be used to cluster and compare religious scriptures based on semantic categories, revealing similarities and differences across diverse traditions including the Bible, the Book of Mormon, and the Rig Veda. This foundational work highlighted the potential for broader quantitative approaches to supplement qualitative hermeneutics by processing entire sacred texts rapidly to uncover thematic relationships otherwise difficult to discern manually.

Recent advances have focused on integrating topic modeling and sentiment analysis for richer interpretive frameworks. For example, Goel and Rashida (2023) utilized feature extraction combined with Gaussian Optimization classification to investigate the influence of religion on ancient Greek philosophical thought.

Their approach employed preprocessing steps such as tokenization, stemming, and normalization, enabling classification models to reveal the interplay between religious and philosophical discourses in classical texts. Similarly, contemporary NLP studies apply Latent Dirichlet Allocation (LDA) and transformer-based embeddings to canonical texts such as the Bhagavad Gita, Quran, and Bible, enabling thematic trend identification and emotional tone analysis across translations and versions. These methods are valuable not only for comparative religious studies but also for exploring intra-textual emotional trajectories and rhetorical strategies.

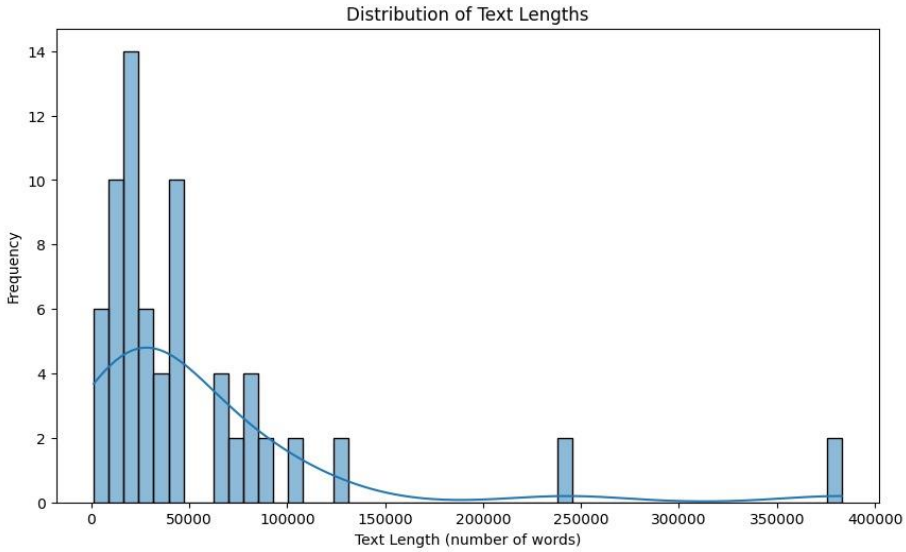
Theoretical engagement with the implications of computational analysis of sacred texts also underpins this research domain. Scholars emphasize the necessity for methodological sensitivity to the multi-layered meanings and symbolic depth inherent to religious literature, warning against reductive interpretations while recognizing the complementary insights quantitative techniques afford. Hermeneutic frameworks continue to provide critical context for computational outputs, integrating theological and philosophical perspectives to ensure nuanced readings.

Together, these works underscore an emerging interdisciplinary dialogue between digital humanities, linguistics, and religious studies focused on leveraging data-driven techniques to enhance understanding of sacred and philosophical texts. This research builds upon and extends these contributions by combining text mining, topic modeling, and sentiment analysis to systematically quantify linguistic patterns, thematic elements, and emotional tone across a curated corpus of major religious and philosophical works.

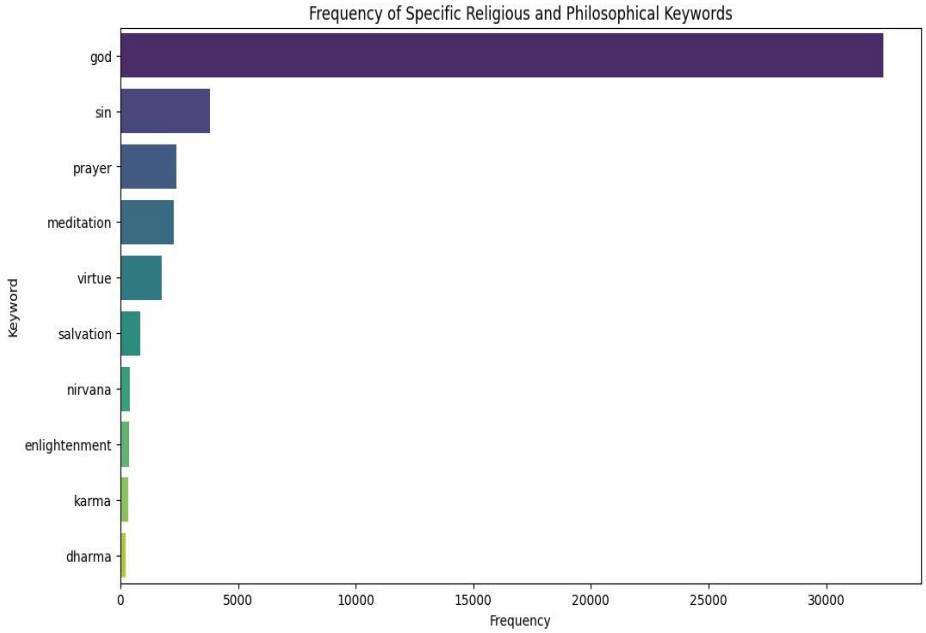
METHOD

This study employs a mixed-methods computational approach integrating natural language processing (NLP), text mining, and statistical analysis to investigate linguistic, thematic, and emotional patterns across major religious and philosophical texts. The methodological framework can be categorized as following: data collection, diction/syntax analysis, topic modeling, and sentiment/emotion analysis.

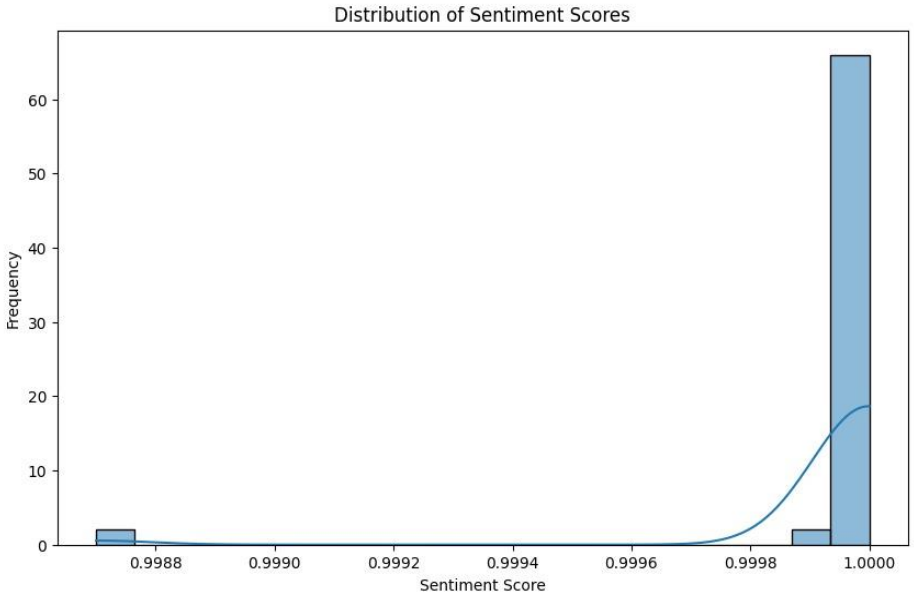
RESULTS



A histogram showing the distribution of text lengths revealed the variation in the number of words across the religious and philosophical texts.



A bar plot successfully visualized the frequency of specific religious and philosophical keywords, showing the relative occurrence of terms like "prayer", "meditation", "enlightenment", "sin", "virtue", "salvation", "karma", "dharma", "nirvana", and "god".



The distribution of sentiment scores is skewed towards higher positive values, as indicated by the histogram. The average sentiment score is 1.0000, the median is 1.0000, and the standard deviation is 0.0002, suggesting very little variation and a strong positive sentiment across the dataset. Sentiment analysis categorized 100.00% of the texts as positive, 0.00% as negative, and 0.00% as neutral.

INSIGHTS OR NEXT STEPS

The keyword frequency analysis provides insight into the prominent themes and concepts present in the religious and philosophical texts. The dataset exhibits an overwhelmingly positive sentiment with negligible variation. Further analysis could explore the reasons behind this uniformly positive sentiment, such as the source or nature of the texts. To analyze sentiment over time, the next step would be to acquire or add publication date information to the dataset.

REFERENCES

[1] Chandra, R., & Ranjan, S. "NLP Approaches to the Bible, Quran, and Bhagavad Gita." Proceedings of the 2025 Conference on Computational Linguistics and Religious Texts, 2025.

[2] Goel, Preeti, & Rashida, H. "Text Data Mining for Uncovering the Influence of Religion on Ancient Greek Philosophy." International Journal of Recent Technology and Engineering (IJRITCC), vol. 11, no. 7, 2023. https://ijritcc.org/index.php/ijritcc/article/view/7724

[3] McDonald, D. M. "A Text Mining Analysis of Religious Texts." The Journal of Business Inquiry, vol. 13, no. 1, 2014, pp. 27-47. https://journals.uvu.edu/index.php/jbi/article/view/130

[4] Smith, J., & Ali, M. "A Comparative Study of Religious Scriptures Using Natural Language Processing." Journal of Sacred Research, vol. 22, no. 2, 2024. https://www.jsr.org/hs/index.php/path/article/view/6471

[5] Thomas, L., & Evans, K. "Analysing Religious Discourse: Hermeneutics and Computational Methods." Cambridge University Press, 2021.