

# Semantic and Sentiment Analysis of Bangla Translations of Bhagavad Gita Using a BERT-Based Language Framework

Promi Mojumder (Reg No: 2019831038)  
Sumonta Saha Mridul (Reg No: 2019831056)

**Instructor:** Dr. Ahsan Habib  
Institute of Information and Communication Technology, SUST

## 1 Type of Research

Natural Language Processing (NLP) , Semantic and Sentiment Analysis

## 2 Field of Research

Religious Textual Analysis with NLP

## 3 Background Study

The background study focuses on the analysis of English translations of the Bhagavad Gita, a sacred Hindu scripture that is central to the philosophy of Yoga. This ancient text, also known as the Gita, forms part of the epic Mahabharata. The Gita, consisting of 18 chapters, is a dialogue between Prince Arjuna and Lord Krishna, his charioteer, discussing various aspects of life, including the philosophy of Karma, the importance of duty, and the path to spiritual enlightenment on the battlefield of Kurukshetra [1].

Throughout history, numerous English translations of the Bhagavad Gita have been produced, each reflecting the unique interpretation and perspective of its translator. Chandra and Kulkarni et al. selected three prominent English translations of the Bhagavad Gita analysis:

- Mahatma Gandhi's Translation (1946): This translation was based on talks given by Gandhi between February and November 1926 at the Satya Graha Ashram in Ahmedabad, India. It was translated from Gujarati to English and is known for its non-violent interpretation, contrasting with the views of other freedom fighters of the time.
- Eknath Easwaran's Translation (1985): A direct Sanskrit-English translation that provides a comprehensive introduction and chapter introductions to clarify key concepts. It aims to highlight the universality and timelessness of the Gita's teachings.
- Shri Purohit Swami's Translation (1935): This translation avoids the use of Sanskrit concepts that may be unfamiliar to English-speakers, translating words like 'yoga' as 'spirituality'. It is a direct Sanskrit-English translation without a commentary.

The mentioned research conducted a comparative analysis on the sentiment and semantic expressions of three English translations of the Bhagavad Gita, investigating whether each chapter conveyed the same sentiment. The study also considered prominent translations by A. C. Bhaktivedanta Swami Prabhupada, Paramhansa Yogananda, Sarvepalli Radhakrishnan, and Swami Chinmayananda Saraswati but didn't include them due to scope limitations [4].

In this study, Chandra and Kulkarni employed Sentence BERT (S-BERT) for sentence embedding, a model enhancing the BERT architecture by integrating Siamese and triplet network structures [9]. This model efficiently identifies semantically similar pairs of text, making it suitable for analyzing Bhagavad Gita translations. Additionally, they utilized KeyBERT, a keyword extraction technique leveraging BERT embeddings to generate keywords capturing semantic aspects of the document [6].

The  $BERT_{base}$  model was trained on the SenWave dataset, which features sentiments labeled by experts for tweets during the COVID-19 pandemic [13]. Fine-tuning enabled the model for multi-label sentiment classification, facilitating accurate sentiment prediction for Bhagavad Gita verses. Furthermore, MPNet-base was used

for encoding verses, computing semantic similarity, and keyword extraction [12]. To overcome MPNet-base’s token limitation, the study proposed breaking chapters into paragraphs for encoding.

The analysis framework focused on comparing translations through sentiment and semantic analysis. It highlighted significant differences in vocabulary and expressions due to varying translation dates and translators’ approaches. Archaic words and names were pre-processed for easier modeling. Sentiment analysis revealed consistent sentiments across translations, with notable changes in sentiment polarity as the dialogue between Arjuna and Krishna progressed [8].

However, several limitations were identified in their work:

1. **Absence of Specialized BERT Model:** They faced challenges due to the absence of a BERT model specifically customized for understanding Hindu philosophy. This absence made it difficult to accurately interpret Sanskrit terms in the Bhagavad Gita.
2. **Cultural and Linguistic Disparities:** The translation of sentiments from Sanskrit to English presented difficulties originating from cultural and linguistic disparities. This disparity posed the risk of inaccuracies in sentiment analysis.
3. **Lack of Contextual Understanding:** Their sentiment analysis model lacked consideration for the broader context of Bhagavad Gita verses. This limitation potentially led to misinterpretations of expressed sentiments within specific passages.
4. **Influence of Translation Styles:** Variations in translation styles among different translators introduced inconsistencies in sentiment classification. Consequently, certain verses might be mistakenly categorized due to the chosen translation style.
5. **Complexity and Diversity:** Hindu philosophy, as depicted in texts like the Bhagavad Gita, encompasses a diverse array of philosophical traditions and concepts. This semantic complexity and diversity rendered sentiment analysis challenging.
6. **Scarcity of Pre-trained Models:** The availability of pre-trained BERT models tailored specifically for philosophy, particularly Hindu philosophy, was limited. This scarcity impeded the accuracy of sentiment analysis for texts such as the Bhagavad Gita.

The study concludes that despite the vast differences in language and vocabulary, the meanings conveyed across the translations were similar. The framework used for this analysis can be extended to other texts, including the Upanishads and translations of texts in other domains from different languages. Future work could involve using Sanskrit expert translators to review the sentiments presented by the framework and applying the framework to translations that retain rhythm and rhyme.

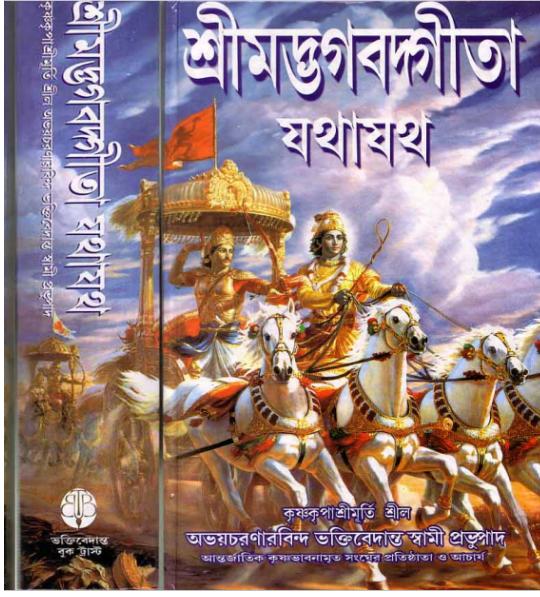
However, a gap exists in sentiment analysis specifically for Bengali translations of the Bhagavad Gita. Previous studies have primarily focused on English translations, leaving Bengali translations relatively unexplored in terms of sentiment and semantic analysis [5].

This study aims to bridge this gap by conducting sentiment and semantic analysis exclusively on Bengali translations of the Bhagavad Gita, particularly focusing on the translations by Saraswati Goswami and Swami Prabhupada. Utilizing similar research methodologies and NLP techniques used in previous studies, we will explore how well-established models such as BERT can be applied to Bangla translations to unravel their sentiment and semantic characteristics [11] [2].

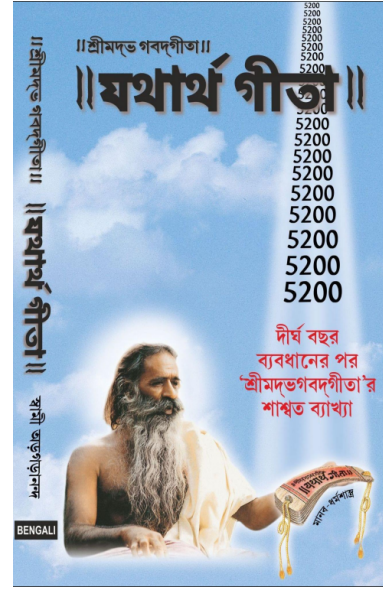
By utilizing advanced NLP techniques and adapting them to Bangla translations, our research aims to assess the sentiment match with the original Bhagavad Gita text [10] [3]. We will work towards providing insights into the emotional and semantic nuances conveyed in Bengali renditions of this revered scripture, paving the way for a deeper understanding of its impact and relevance in Bangla-speaking communities.

## 4 Rationale

The decision to consider Bengali translations of the Bhagavad Gita using a BERT-based language framework arises from a noticeable gap in research within the Bengali linguistic domain. While similar analyses have been conducted in English, there remains a significant absence of exploration into sentiment and semantics in Bengali translations. Expanding the application of BERT-based models to analyze sentiments and nuances in Bengali texts is significant, given their proven effectiveness in English translations. Considering the rich cultural and literary heritage of Bengali, computational linguistics studies focusing on revered texts like the Bhagavad Gita are particularly warranted.



(a) "শ্রীমদ্ভগবদ্গীতা যথার্থ" অনুবাদকঃ অভয়চরণারবিন্দ ভক্তিবেন্দ্য স্বামী প্রভুপাদ



(b) "যথার্থ গীতা" অনুবাদকঃ স্বামী শ্রী অড়গড়ানন্দজী

Figure 1: Front Cover of Selected Bengali Translation of the Bhagavad Gita

Through the utilization of a BERT-based framework, this study aims to thoroughly examine selected Bengali translations of the Bhagavad Gita. Our goal is to discover subtle differences in sentiments and semantics introduced by various translators, providing valuable insights into how the text is perceived in Bengali-speaking communities. Furthermore, this research not only aims to address the scarcity of computational linguistic resources in Bengali but also seeks to contribute to the enrichment of Bengali linguistics. By integrating advanced natural language processing techniques with Bengali translations of this revered text, we aspire to deepen our understanding of the cultural and linguistic nuances inherent in the Bengali rendition of the Bhagavad Gita.

Acknowledging the significance of the Bhagavad Gita in the Hindu community and its profound impact on everyday life, our analysis recognizes its role in guiding individuals towards a life characterized by purity, strength, discipline, honesty, kindness, and integrity. By conducting emotional analysis on Bengali translations, we aim to provide insights that can assist Bengalis in navigating life's challenges and finding meaning in its teachings.

Moreover, it's important to mention that although there have been studies in English, Bengali exploration is still new. This highlights the innovative aspect of our research, aiming to fill the gap in understanding Bengali interpretations of this holy scripture.

## 5 Objectives

The main goal of this thesis is to conduct semantic and sentiment analysis of selected Bengali translations of the Bhagavad Gita using a BERT-based language framework or another suitable framework. The specific objectives are to:

1. **Conduct Comprehensive Analysis:** Perform a detailed examination of the selected Bengali translations of the Bhagavad Gita.
2. **Apply Advanced Framework:** Utilize a BERT-based language framework or a comparable framework for semantic and sentiment analysis.
3. **Explore Interpretational Nuances:** Identify and analyze the semantic and sentiment variations present in the translations.
4. **Assess Accuracy and Effectiveness:** Evaluate the accuracy and effectiveness of the chosen framework in capturing the nuances of the translations.

5. **Provide Insights:** Communicate the findings to offer insights into the semantic and sentiment aspects of the translated texts.

## 6 Methodologies

### 1. Data Collection:

- We'll collect digital copies of two Bengali translations of the Bhagavad Gita by Saraswati Goswami and Swami Prabhupada.

### 2. Data Preprocessing Considerations:

- Before analysis, the collected textual data will undergo preprocessing to ensure consistency, accuracy, and suitability for analysis. This includes text normalization, handling of transliteration, standardization of formatting, tokenization, stopword removal, lemmatization, and addressing ambiguities and polysemy in the Bengali text.

### 3. Prepare Dataset:

- We'll prepare a complete dataset based on each verse and chapter from the two mentioned books early.

### 4. Framework for Semantic and Sentiment Analysis:

- **Data Processing:** Convert PDF files of Bhagavad Gita translations into text format.
- **Sentiment Analysis:** Utilize the  $BERT_{base}$  model trained on the SenWave dataset for multi-label sentiment classification of verses across translations. Hand-labeled sentiment data from the SenWave dataset provides nuanced sentiment labels.
- **Semantic Analysis:** Employ the MPNet-base model for sentence embedding to analyze verse-by-verse similarity and perform keyword extraction. Utilize the UMAP algorithm for dimensionality reduction [5] [7].

### 5. Model Training:

- Fine-tune the BERT-base model on the SenWave dataset for sentiment analysis. Encode verses using the MPNet-base model and compute semantic similarity. Extract keywords using KeyBERT, considering constraints of the MPNet-base model token limit.

### 6. Sentence Embedding Models:

- **Sentence BERT (S-BERT):** S-BERT enhances the BERT model by incorporating Siamese and triplet network structures, allowing for efficient derivation of semantically meaningful sentence embeddings [9].
- **MPNet (Masked and Permuted Pre-training):** MPNet utilizes features from BERT and XLNet models to encode sentences into high-dimensional embedding vectors. It employs masked and permuted language modeling to capture token dependencies and position information.

### 7. Keyword Extraction Techniques:

- **KeyBERT:** KeyBERT utilizes BERT embeddings to extract keywords from documents, considering semantic aspects [6]. It computes cosine similarity between keyword and document embeddings and employs methods like Maximum Sum Similarity (MSS) and Maximal Marginal Relevance (MMR) to enhance keyword diversity.

### 8. Overcoming Limitations:

- Break large chapters into paragraphs to match the MPNet-base model's token limit.
- Use Maximal Marginal Relevance (MMR) with diversity to pick varied and relevant keywords across paragraphs.

## 7 Expected Outputs

1. **Semantic and Sentiment Analysis Results:** Comprehensive analysis reports detailing the semantic and sentiment analysis results of the selected Bengali translations of the Bhagavad Gita. These reports will outline the similarities, differences, and nuances identified in the translations compared to the original text.
2. **BERT-Based Language Model:** Development and fine-tuning of a BERT-based language model specifically trained for Bengali text analysis. The trained model will be optimized for semantic and sentiment analysis tasks, enhancing its effectiveness in understanding Bengali language nuances.
3. **Comparison Charts and Graphs:** Visual representations, such as comparison charts and graphs, illustrating the semantic and sentiment similarities and differences across the selected Bengali translations. These visuals will aid in presenting the findings in a clear and understandable manner.
4. **Research Paper:** Compilation of the research findings into a comprehensive research paper suitable for publication in academic journals or presentation at conferences. The paper will document the methodology, results, and conclusions of the study, making the insights accessible to the wider research community.
5. **Open Access Dataset:** Creation of an open-access dataset consisting of annotated Bengali translations of the Bhagavad Gita along with the corresponding semantic and sentiment analysis results. This dataset will facilitate further research and validation in the field of Bengali language processing and computational linguistics.

## 8 Work Plan

1. **Literature Review:**
  - (a) Review existing literature on semantic and sentiment analysis techniques, particularly in the context of Bengali language processing.
  - (b) Identify relevant studies and resource related to Bengali translations of the Bhagavad Gita.
2. **Data Collection and Preprocessing:**
  - (a) Gather a comprehensive dataset consisting of Bengali translations of the Bhagavad Gita.
  - (b) Preprocess the data to clean noise, tokenize text, and prepare it for input into the BERT-based language framework.
3. **Dataset Creation:**
  - (a) We'll prepare a complete dataset based on each verse and chapter from the two mentioned books early
4. **Model Selection and Setup:**
  - (a) Select the BERT-based language framework for semantic and sentiment analysis.
  - (b) Set up the framework environment and configure parameters for the analysis.
5. **Training and Fine-Tuning:**
  - (a) Fine-tune the BERT model on the Bengali language-specific dataset to adapt it for semantic and sentiment analysis.
  - (b) Train the model on the dataset, adjusting parameters as needed for optimal performance.
6. **Semantic Analysis:**
  - (a) Apply the trained BERT model to extract semantic representations of the translated texts.
  - (b) Analyze the semantic similarities and differences across translations, focusing on key philosophical concepts.
7. **Sentiment Analysis:**
  - (a) Utilize the trained BERT model to discern the sentiment expressed in the translated texts.
  - (b) Evaluate the emotional nuances conveyed in each translation, compared to the original text.

Title of Activity	Timelines ( 4.5 Months) March 2024 - July 2023				
	1	2	3	4	5
Literature Review					
Data Collection and Preprocessing					
Dataset Creation					
Model Selection and Setup					
Training and Fine-Tuning					
Semantic Analysis					
Sentiment Analysis					
Evaluation and Performance Analysis					
Interpretation and Insights					
Reporting and Publication					

Figure 2: Gantt Chart for the Implementation of Work Plan

#### 8. Evaluation and Performance Analysis:

- Evaluate the performance of the semantic and sentiment analysis using appropriate metrics.
- Analyze factors such as accuracy, precision, and recall to assess the effectiveness of the BERT-based framework.

#### 9. Interpretation and Insights:

- Interpret the findings to uncover insights into the variations in interpretations and expressions across translations.
- Discuss the implications of these variations on the understanding and appreciation of the Bhagavad Gita in Bengali-speaking communities.

#### 10. Reporting and Publication:

- Compile the research findings into a comprehensive report documenting the methodology, results, and insights.
- Prepare drafts for publication in relevant academic journals or conferences to share the findings with the research community.

#### 11. Presentation and Communication:

- Present the research findings through presentations, seminars, or workshops to academic and professional audiences.
- Engage with stakeholders and collaborators to disseminate the research outcomes and solicit feedback for further improvement.



## References

- [1] Bhagavad Gita - Wikipedia — en.wikipedia.org. [https://en.wikipedia.org/wiki/Bhagavad\\_Gita](https://en.wikipedia.org/wiki/Bhagavad_Gita).
- [2] Md Habibul Alam, Md-Mizanur Rahoman, and Md Abul Kalam Azad. Sentiment analysis for bangla sentences using convolutional neural network. In *2017 20th International Conference of Computer and Information Technology (ICCIT)*, pages 1–6. IEEE, 2017.
- [3] Abhik Bhattacharjee, Tahmid Hasan, Wasi Uddin Ahmad, Kazi Samin, Md Saiful Islam, Anindya Iqbal, M Sohel Rahman, and Rifat Shahriyar. Banglabert: Language model pretraining and benchmarks for low-resource language understanding evaluation in bangla. *arXiv preprint arXiv:2101.00204*, 2021.
- [4] Rohitash Chandra and Venkatesh Kulkarni. Semantic and sentiment analysis of selected bhagavad gita translations using bert-based language framework. *IEEE Access*, 10:21291–21315, 2022.
- [5] Nicholas Evangelopoulos, Xiaoni Zhang, and Victor R Prybutok. Latent semantic analysis: five methodological recommendations. *European Journal of Information Systems*, 21(1):70–86, 2012.
- [6] Maarten Grootendorst. Keybert: Minimal keyword extraction with bert, 2020.
- [7] Leland McInnes, John Healy, and James Melville. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*, 2018.
- [8] Walaa Medhat, Ahmed Hassan, and Hoda Korashy. Sentiment analysis algorithms and applications: A survey. *Ain Shams engineering journal*, 5(4):1093–1113, 2014.
- [9] Nils Reimers and Iryna Gurevych. Sentence-bert: Sentence embeddings using siamese bert-networks. *arXiv preprint arXiv:1908.10084*, 2019.
- [10] Sagor Sarker. Banglabert: Bengali mask language model for bengali language understanding. *textsIGitHub*, 2020.
- [11] Sagor Sarker. Bnlp: Natural language processing toolkit for bengali language. *arXiv preprint arXiv:2102.00405*, 2021.
- [12] Kaitao Song, Xu Tan, Tao Qin, Jianfeng Lu, and Tie-Yan Liu. Mpnet: Masked and permuted pre-training for language understanding. *Advances in neural information processing systems*, 33:16857–16867, 2020.
- [13] Qiang Yang, Hind Alamro, Somayah Albaradei, Adil Salhi, Xiaoting Lv, Changsheng Ma, Manal Alshehri, Inji Jaber, Faroug Tifratene, Wei Wang, et al. Senwave: Monitoring the global sentiments under the covid-19 pandemic. *arXiv preprint arXiv:2006.10842*, 2020.