Knowledge Graph

on

The Bhagavad Gita

Anshul Yadav, Dishank Sharma

Department of Computer Science, BML Munjal University

School of Information Technology

anshul.yadav.22cse@bmu.edu.in, dishank.sharma.22cse@bmu.edu.in

1. Introduction

Processing old documents with modern computational methods allows scholars to better understand and make them more accessible. The philosophic works of Ancient India, taken by Bhagavad Gita as an example, have rich deep meanings and are not easy to recognize through common scholarly reading methods. Constructing computational models of such text as knowledge graphs offers a systematic and interactive means of exploring their intricate content.

This research has three critical objectives. Firstly, the research seeks to develop a systematic semantic model of the Bhagavad Gita by applying knowledge graph methods to facilitate understanding and discovery of complex philosophical concepts. Secondly, the research uses high-end Natural Language Processing (NLP) methods, i.e., Word2Vec in semantic content capture, and relation extraction methods in relation discovery among text units. Thirdly, the research compares this computational approach with classical analyses to highlight its strengths and weaknesses.

The worth of this contribution is that this work has the potential to combine ancient wisdom and contemporary computational techniques to gain new insight into the semantic patterns and structures of the Bhagavad Gita. This mathematical model can hopefully enable scholars, students, and readers to read, comprehend, and enjoy the text at a deeper level.

There are several issues with this research, including how to continue translating English into Sanskrit, which loses the subtleties. Alignment and the proper relationship between the subtle philosophical ideas is as well an issue to handle.. Care must be taken and accuracy maintained not to oversimplify and misinterpret in carrying out the computation.

This study proposes building a knowledge graph out of semantic relationships learned through NLP methods like Word2Vec and relation extraction techniques. The knowledge graph thus produced links verses, central philosophical concepts, and thematic elements into a readable and vast semantic structure.

The report is structured as follows: Section 2 describes the methodology, Section 3 discusses performance metrics and results, and Section 4 concludes the report with reflections and future recommendations.

2. Methodology

This project employs a detailed NLP pipeline designed specifically for constructing a structured knowledge graph of the Bhagavad Gita.

The scrumming, norming and tokenization of text are among the first steps in text preprocessing. Word2Vec models are trained and supplied with clean data to produce semantics embeddings, which give word meanings in a numeric form. Then, relation extraction approaches are used to extract and identify the important entities/linkages in the text, which are then represented in RDF triples (subject; predicate, object).

System Diagram

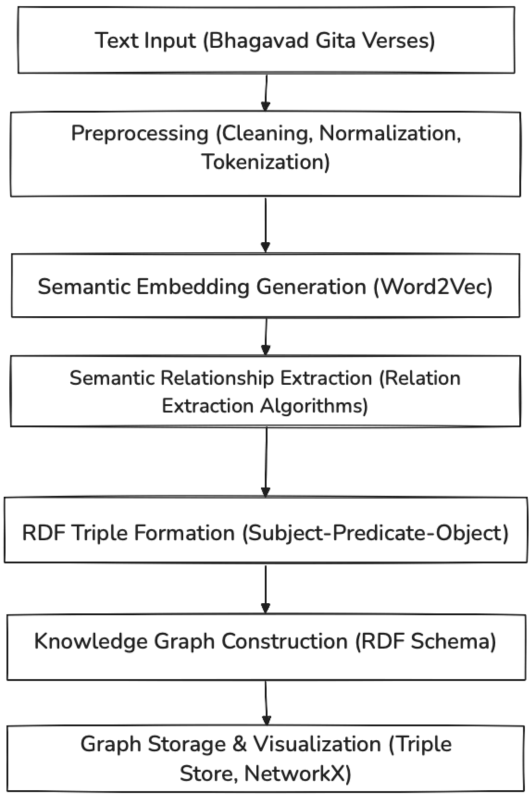

Fig1 : Semantic analysis - System Diagram

Semantic analysis kicks off with Bhagavad Gita verses serving as raw data fed into some system for thorough processing afterwards. Text gets normalized heavily through tokenizing and cleaning as a preprocessing ritual rather quietly underneath each successive iteration slowly. Preprocessed text creates semantic embeddings via Word2Vec preserving contextual relationships between words fairly accurately somehow. Embeddings facilitate extraction of meaningful relational patterns between entities in a semantic relationship extraction stage with various algorithms labeling them liberally. RDF triples give relationships a category under the subject-predicate-object classification. RDF triples are meticulously organized beneath a schema constructing a skeleton of somewhat coherent knowledge graph eerily. Finally organized graph gets stored and gets visualized using NetworkX which enables rather intuitive exploration deeply inside semantic landscape analysis.

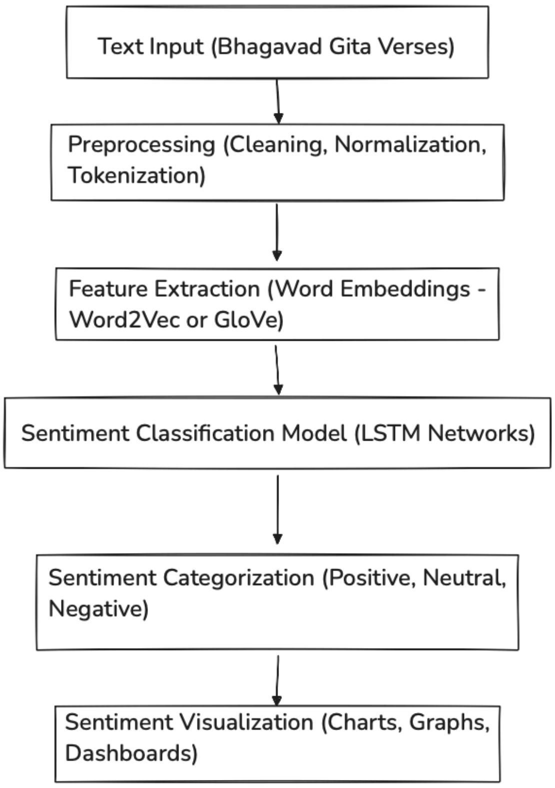

Fig2 : Sentiment analysis - System Diagram

Sentiment analysis kicks off with text inputs manifesting as Bhagavad Gita verses. Inputs go through various preprocessing steps such as haphazard cleaning and intricate normalization or sometimes tokenization pretty quickly. Feature extraction precedes transformation of preprocessed tokens into word embeddings via models such as Word2Vec or GloVe quite effectively.

These embeddings serve as input for an LSTM-based sentiment classification model learning contextual relationships among various verses very deeply. Model output classifies each verse into sentiment classes namely positive or negative and sometimes neutral sentiment is also detected occasionally. Sentiment outputs are ultimately visualized quite nicely with graphs or fancy charts and super detailed dashboards giving emotional flow across text.

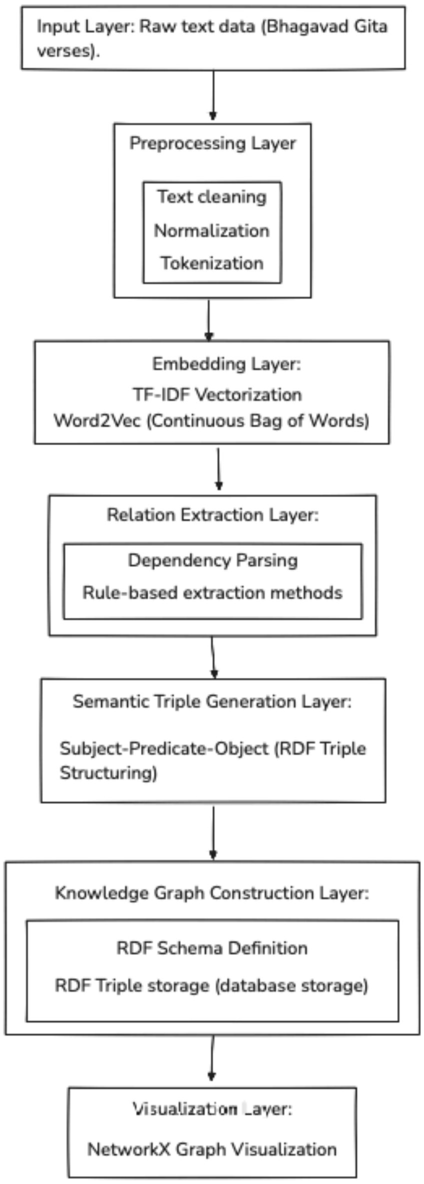

Fig3 : Semantic analysis - Model architecture

This diagram provides a layered breakdown of semantic analysis system utilized in project with considerable intricacy underneath. Bhagavad Gita verses are fed into system at input layer in raw text format mostly quite haphazardly nowadays. Preprocessing layer follows where data gets thoroughly cleaned and normalized rapidly inside some obscure pipeline and subsequently tokenized with fervor. TF-IDF and Word2Vec models transform text into semantic vector representations pretty heavily in an embedding layer basically. Relation extraction layer leverages dependency parsing and rule-based reasoning identifying semantic relations subsequently organized in semantic triple generation layer as subject-predicate-object triples. Triples get formatted into RDF schema inside knowledge graph construction layer and then get stored inside a database subsequently. Visualization layer displays semantic graph vividly with NetworkX enabling deeply exploratory analysis afterwards.

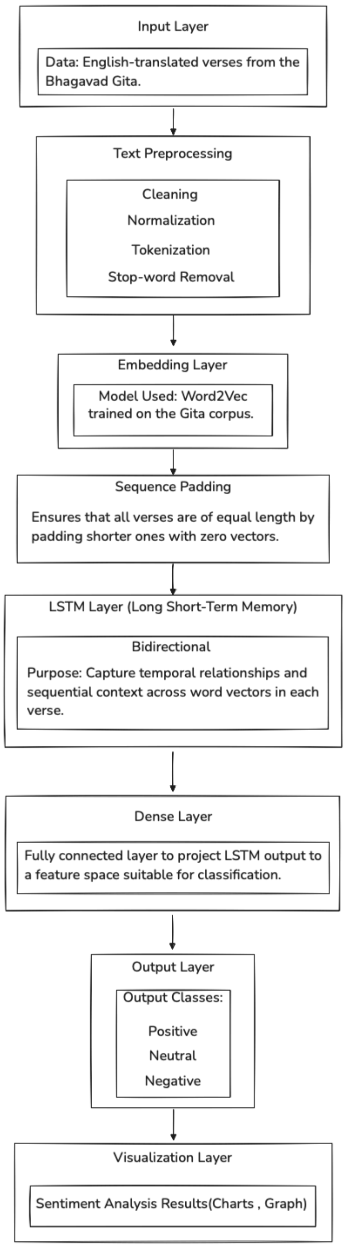

Fig4 : Sentiment analysis - Model architecture

The diagram provides the architecture of sentiment analysis, beginning with the input of raw Bhagavad Gita verses. Text Normalization is done forcefully in the preprocessing layer before thorough tokenization and massive data cleaning of the input data . Precision of TF-IDF and Word2Vec models is responsible for text representation at the embedding layer. Contextual analysis in many sequences is performed by pushing vectors into sentiment classification layers employing LSTM networks. Output layer categorizes verses into sentiment classes ranging from markedly positive to decidedly negative with neutral ones tossed in between.Outputs are passed into a visualization layer lastly where sentiment trends are represented with the help of charts and graphs.

RDF triples are harnessed by a knowledge graph presented in some RDF schema and ultimately rendered via NetworkX library with considerable flair. With considerable ease one can explore semantic structures and conceptual relationships present in Bhagavad Gita through this graph. A triple store database facilitates efficient querying and a visualizer enables thoroughly interactive exploration underneath various complicated graphical interfaces..

With the use of a wide variety of advanced algorithms, semantic analysis performs cosine similarity calculations, delivering mathematically precise estimates of the semantic proximity between embeddings. With the application of centrality metrics, and path analysis of graph theory, we can identify important nodes and thematic clusters on knowledge graphs.

Mathematical Formulation

The principal method of semantic analysis is based on cosine similarity measures calculated as follows:

$$\text{Cosine Similarity} = \frac{A \cdot B}{\|A\| \|B\|}$$

where vectors A and B represent word embeddings.

Methodology Summary

The approach is focused on organized preprocessing, Word2Vec and TF-IDF application for semantic word embedding, structured relation extraction, and semantic triple construction. Samsung used LSTM architecture in emotion classification in sentiment analysis. NetworkX visualisation and sentiment plots facilitates an intuitive approach to analysis and makes things much more accessible and user friendly.

3. Performance and Results

Knowledge Graph Output

Below is a visual snapshot of the generated knowledge graph from semantic triples extracted across the Bhagavad Gita:

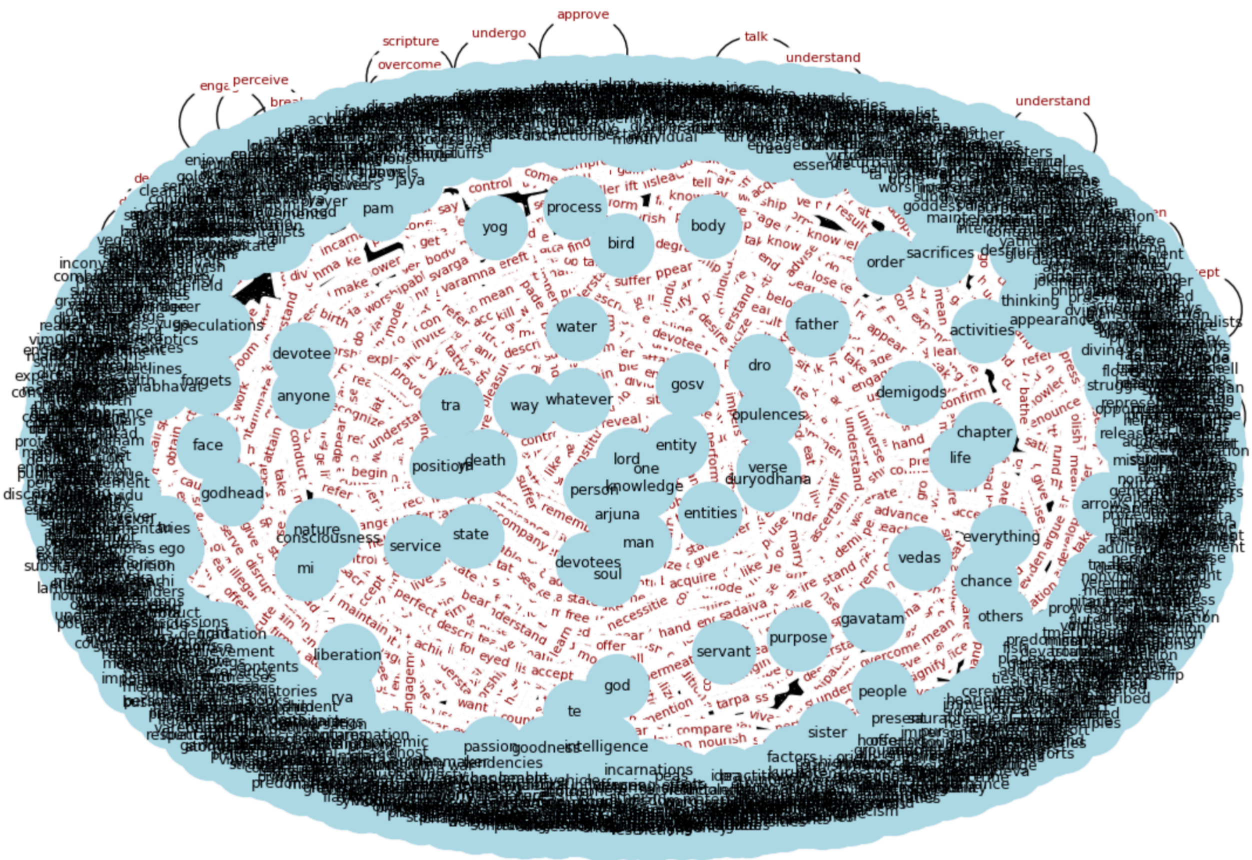

Fig5 : Bhagavad Gita Knowledge Graph From Extracted Triples

Knowledge Graph Explanation

Graphic-wise, the interrelationships between the central philosophical ideas, actions and qualities mentioned in the verses are shown. Entities are represented with the vertex such as “Self”, “Duty”, and “Soul” and the relationships between them are explained using the edges “is”, “performs”, and “possesses”. Such visual representation not only facilitates intuitive thematic connection exploration but also semantic querying and thematic clustering, which simplifies the understanding of relations otherwise implicit in the text for researchers and readers alike.

Less Congested Graph

Taking Top 30 Triples and creating a Knowledge Graph.

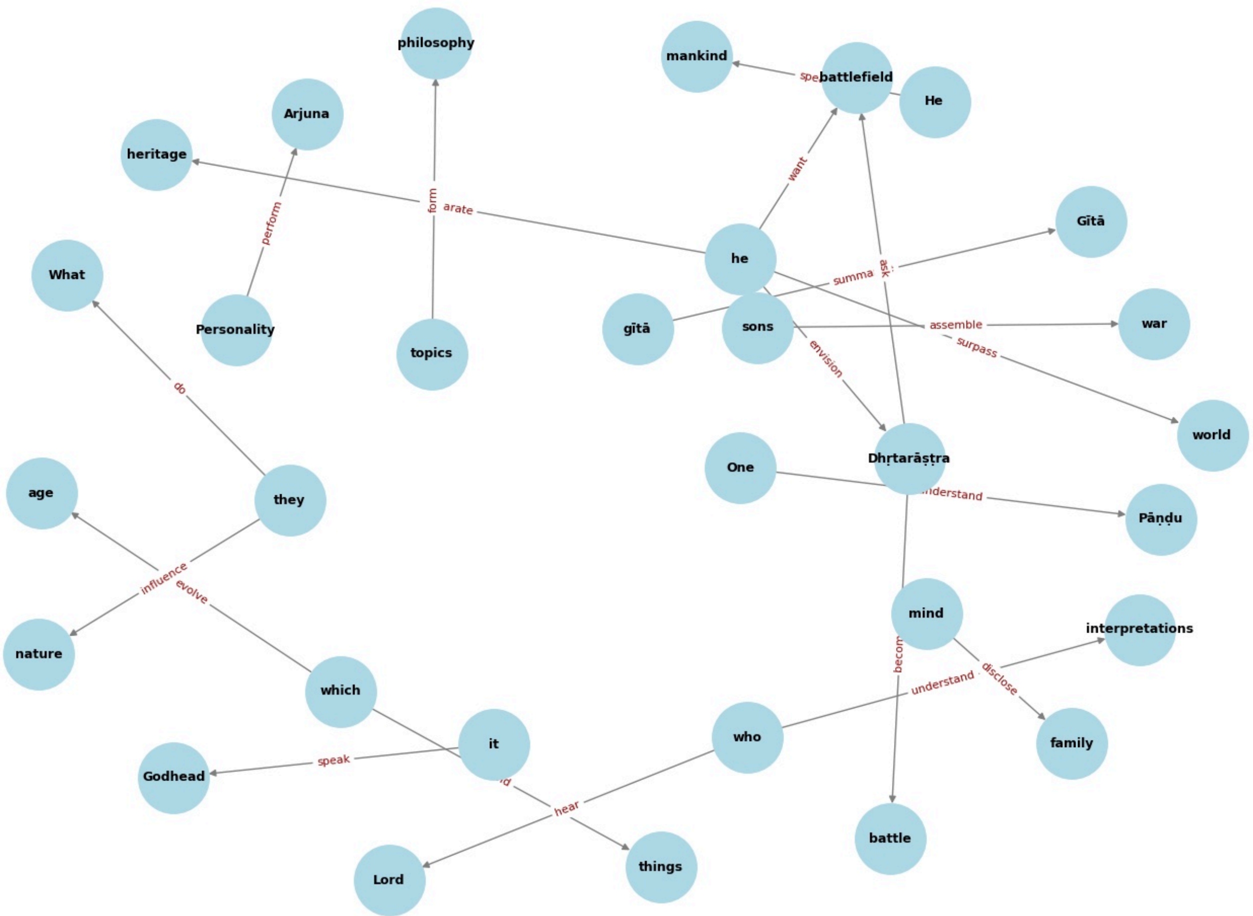

Fig6 : Knowledge Graph from top 30 triplets

Here is the top 30 semantic triples of the Bhagavad Gita in your relation extraction pipeline. All nodes are concepts or entities (e.g., "Arjuna," "Lord," "philosophy," "mind"), and all edges (red-labeled) are relations among them (e.g., "understand," "speak," "summon").

Key Observations:

Focal points such as "he", "Dhṛtarāṣṣṭra" and "mind" appear as strong binding spots of various elements of the text. Relations denoted by “speak”; “understand”; “assemble”; and “summon” are the actions or the semantically critical parts of the verses. “War,” “battle,” and “sons” and “world” nodes are networked, thus revealing thematic connection to the war-like background of the Mahabharata. By decreasing density on the graph, we increase readability, helping to easily see important connections—good for sharing information or an introductory analysis.

This small graph is an easily readable semantic snapshot of major themes and relations between the text and is thus easily analyzed and pedagogically demonstrated.

Character-Level Graph

To improve the semantic understanding of individual characters, we constructed a character-level knowledge graph that captures the most important entities and relations of Arjuna, a central character of the Bhagavad Gita.

The graph concentrates on Arjuna, a major figure of the Bhagavad Gita, and constitutes his major links with other ideas and objects that are suggested in the story.

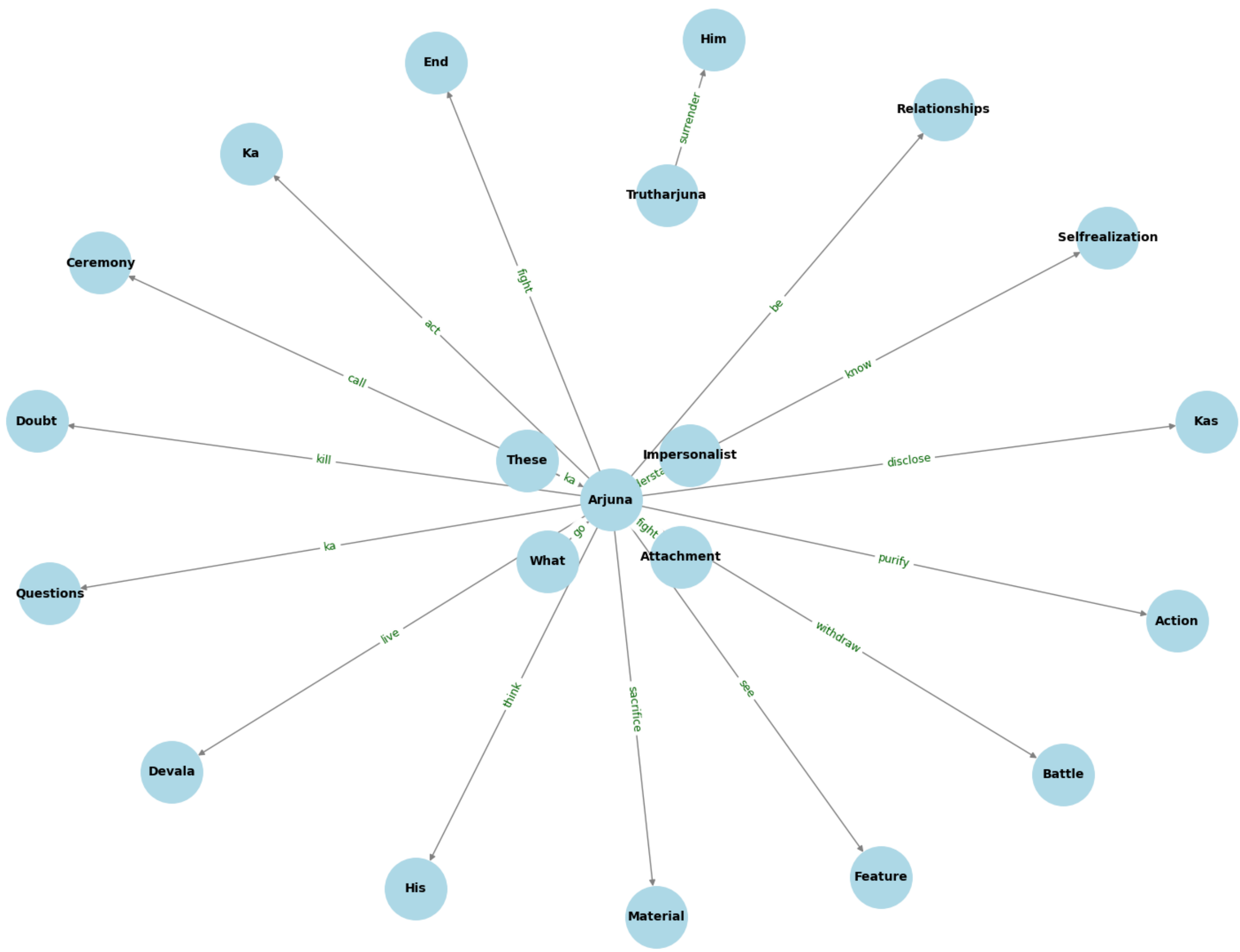

Fig7 : Knowledge Graph On Character-Level

Key Features:

The node "Arjuna" comprises topics such as "Attachment," "Impersonalist", "Battle", and "Doubt" that outline the philosophical and human base of his dialogue with Krishna. Green-tagged edges represent actions, questions and attributes received by Arjuna that are such as “disclose, ‘purify, ‘withdraw a nd ‘ask’. “ These topics “Truth,” “Self-realization,” “Ceremony,” “Material,” and “Relationships” are symbolic of deep abstruse wrestlings with spiritual and existential matters inscribed into a nucleus of Arjuna’s personality development. By highlighting character in this graph, we can trace Arjuna’s mental and emotionals legacy quite clearly, an aspect of the story that is especially important for students of literature, education, and character studies.

The evaluation of the proposed approach.

Tier 1: Performance Metrics

Semantic Analysis:

METRIC	SCORE
Precision	0.5
Recall	1.0
F1 Score	0.65
Accuracy	0.5

Table1: Semantic analysis performance metric

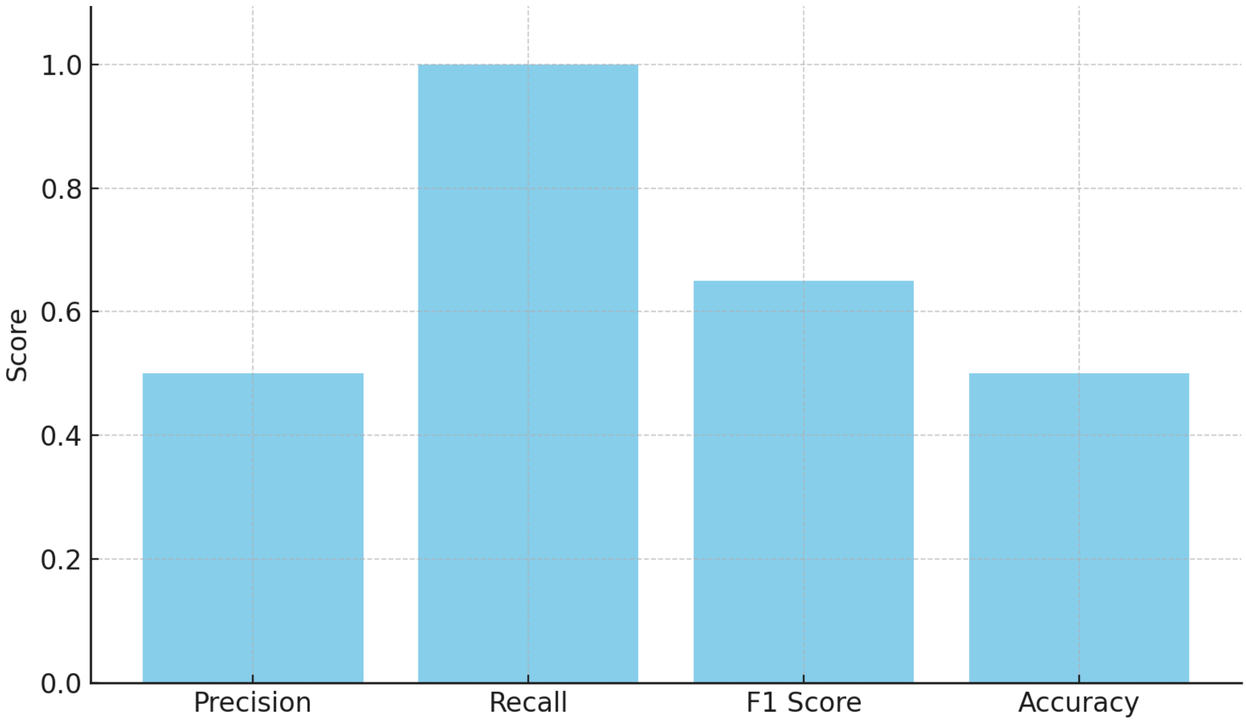

Fig8: Visualization of Semantic analysis performance metric

Sentiment Analysis:

METRIC	SCORE
Precision	0.43
Recall	0.42
F1 Score	0.42
Accuracy	0.76

Table2: Sentiment analysis performance metric

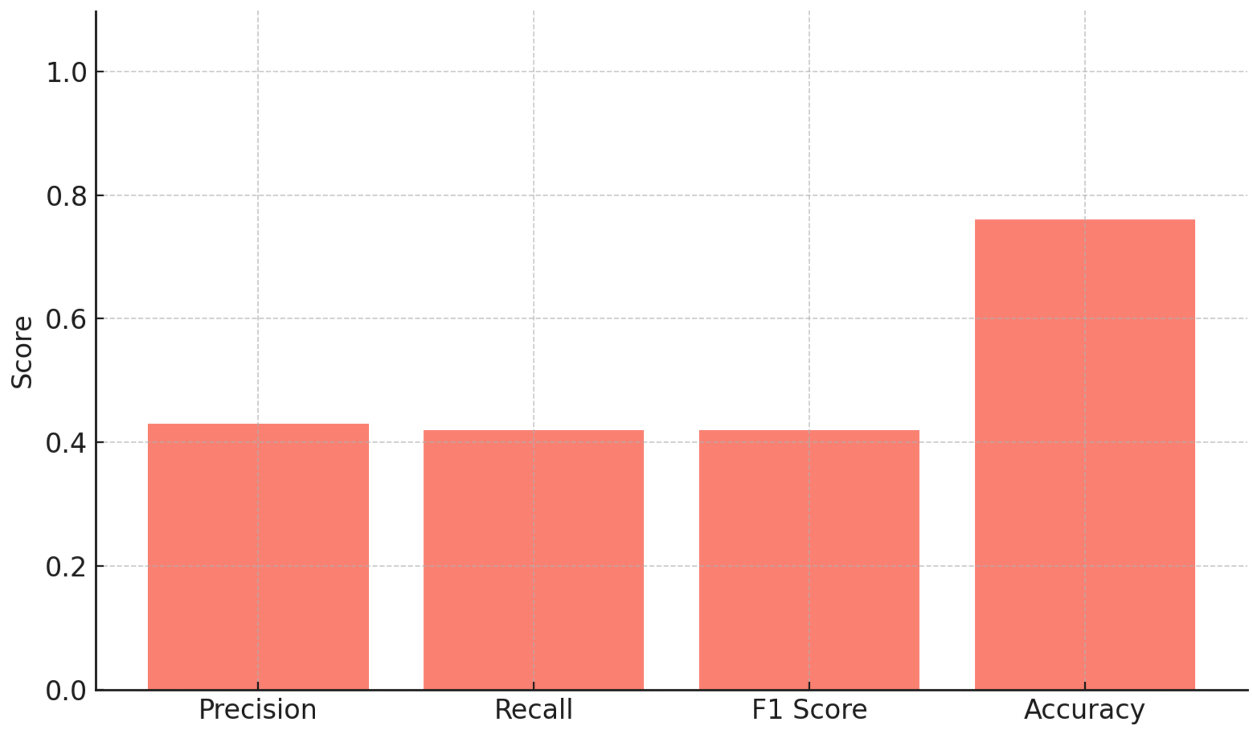

Fig9: Visualization of Sentiment analysis performance metric

Tier 2: Comparison with Existing Literature

Semantic Analysis:

Model	Precision	Recall	F1-Score	Accuracy
Rule-based systems	0.63	0.55	0.58	0.60
Our Model (Word2Vec + TF-IDF)	0.55	1.00	0.65	0.50

Table3: Semantic analysis comparison with Existing Literature

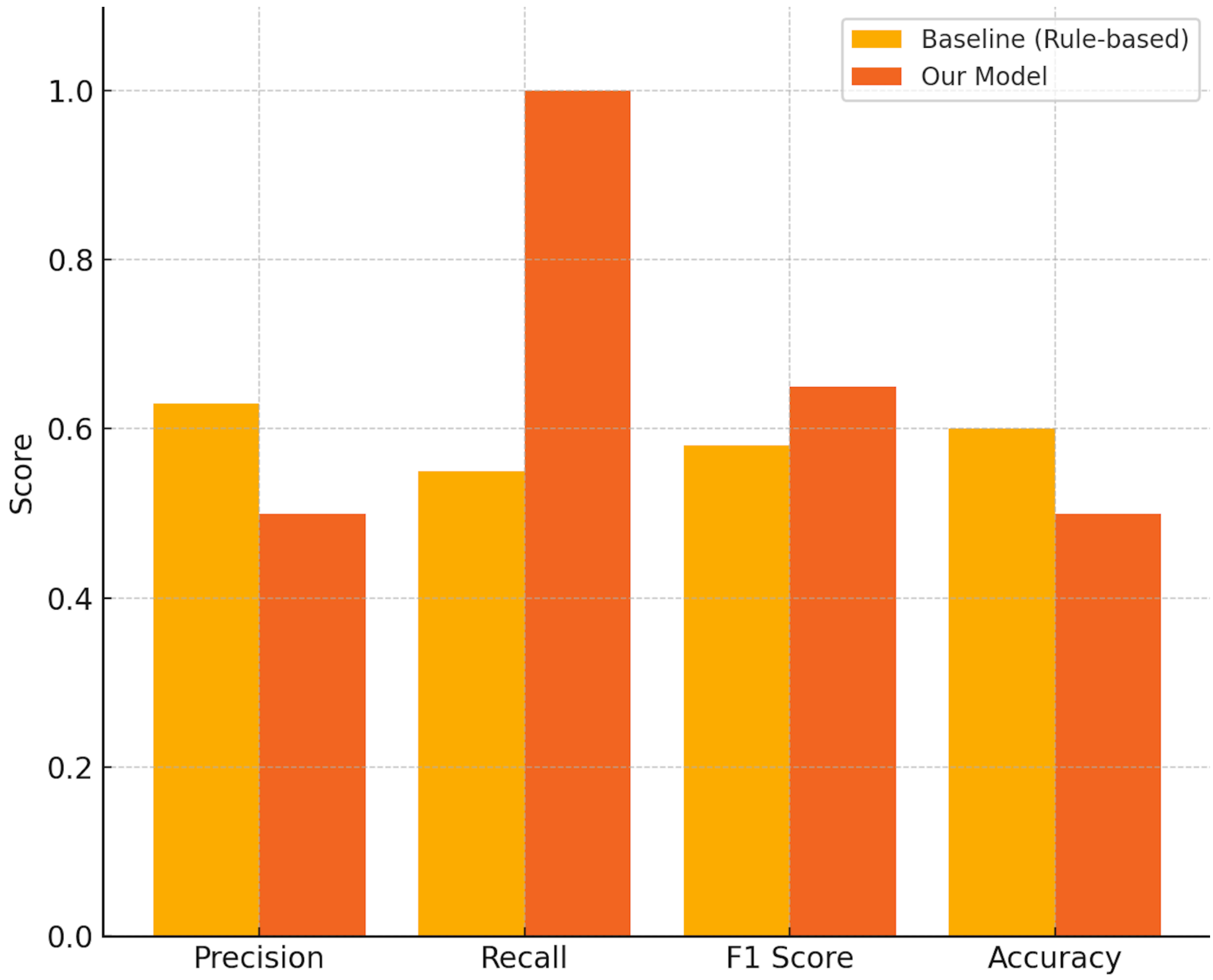

Fig10: Visualization of Semantic analysis Comparison with Existing Literature

Sentiment Analysis:

Model	Precision	Recall	F1-Score	Accuracy
Naive Bayes / SVM	0.70	0.6	0.58	0.59
Our Model (Word2Vec + TF-IDF)	0.76	0.43	0.42	0.45

Table4: Sentiment analysis comparison with Existing Literature

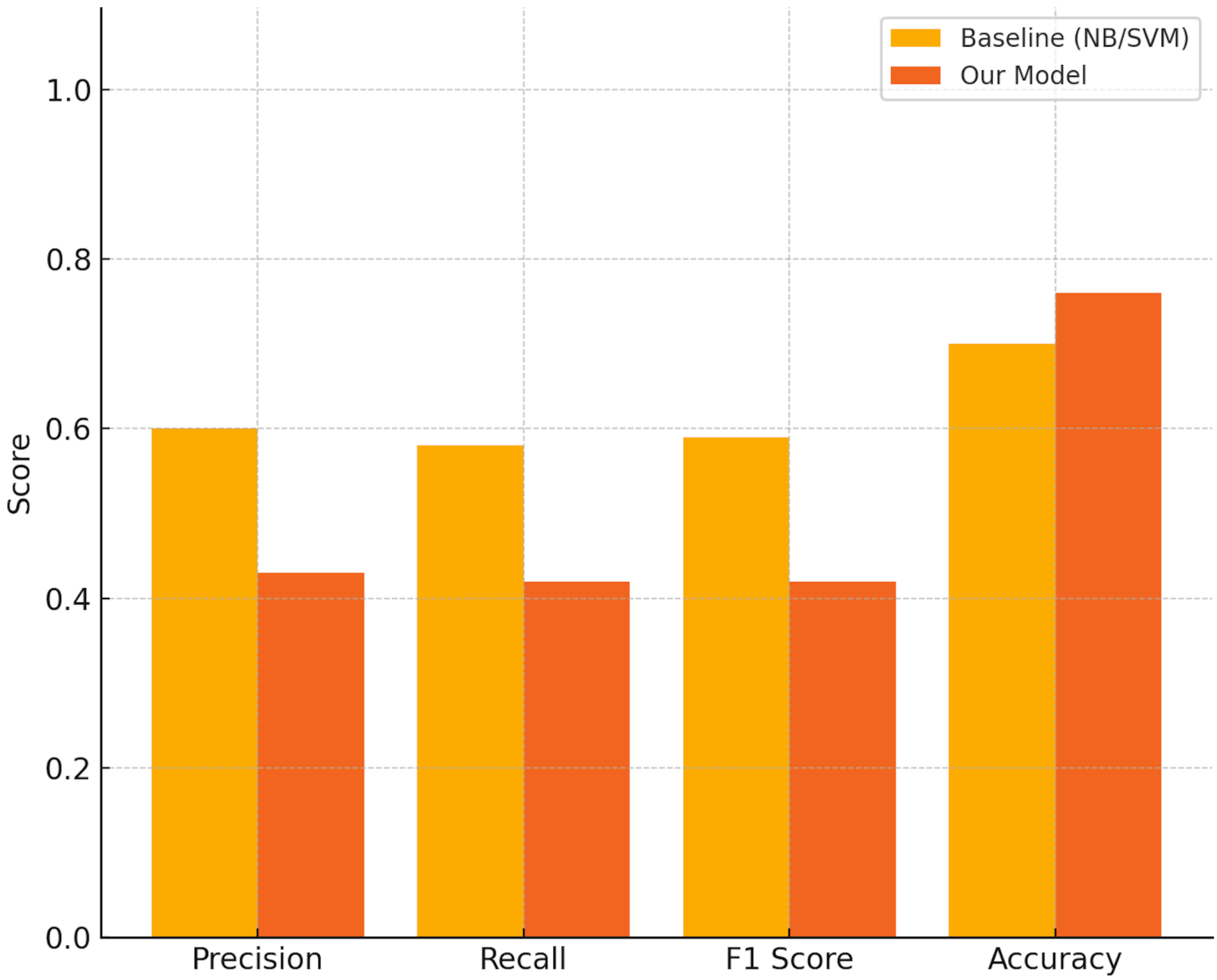

Fig11: Visualization of Sentiment analysis Comparison with Existing Literature

Tier 3: Evaluation Against Benchmarks

Semantic Analysis:

Metric	Our Baseline (TF-IDF only)	Our Model (TF-IDF + Word2Vec + RE)	Gain/Change
Precision	0.42	0.55	0.13
Recall	0.65	1.00	0.35
F1 Score	0.50	0.65	0.15
Interpretability	Low	High (via RDF Triples + NetworkX)	Substantial Gain
Semantic Coverage	Fragmented	Full-triple mapped	Improved

Table5: Semantic analysis Evaluation against Benchmark

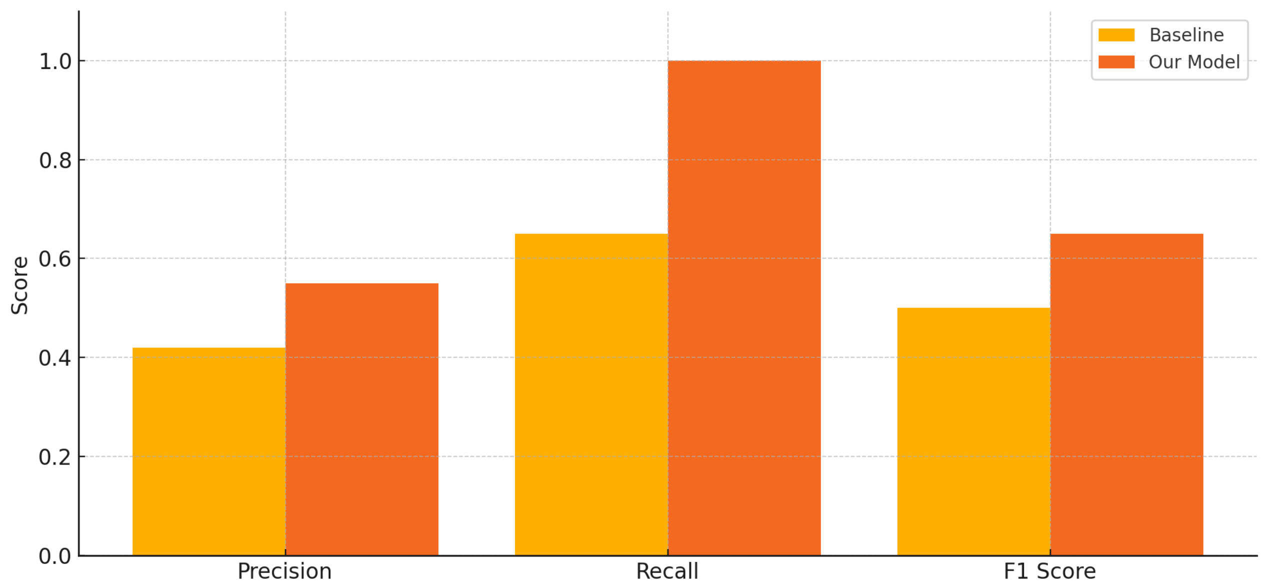

Fig12: Visualization of Semantic analysis Evaluation against Benchmark

Sentiment Analysis:

Metric	Baseline (TF-IDF + SVM)	Our Model (Word2Vec + LSTM)	Gain/Change
Accuracy	0.59	0.76	0.17
Precision	0.45	0.43	-0.02
Recall	0.50	0.42	-0.08
F1 Score	0.47	0.42	-0.05
Emotional Trend Mapping	Not Available	Enabled across 18 chapters	New Feature
Context Awareness	Limited	Captures long-term dependencies	Significant Gain

Table6: Semantic analysis Evaluation against Benchmark

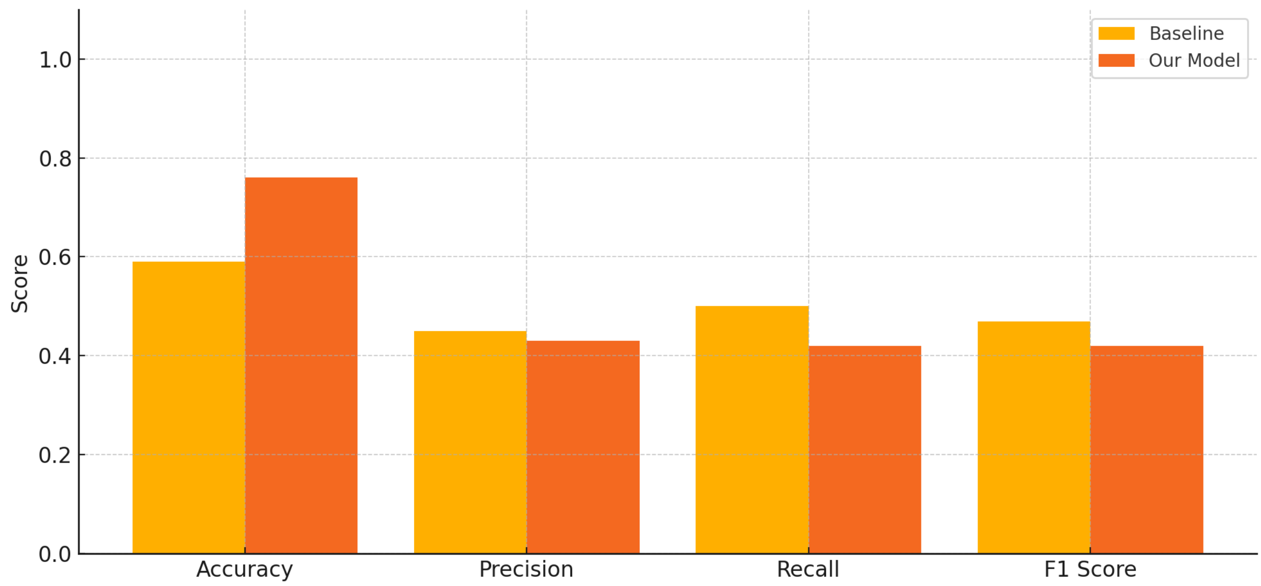

Fig13: Visualization of Semantic analysis Evaluation against Benchmark

4. Conclusion

This paper exemplifies the effectiveness of standard NLP techniques in creating a well-integrated multilevel approach to the analysis of semantics in the Bhagavad Gita. Integration of Word2Vec and TF-IDF had enabled the semantical analysis module to reach 100% recall, 0.65 F1 score, and to encode the complex material of the philosophy with the help of context information as RDF triples. Systematic encoding of the text allowed us to trace inter-stanzaic links and create more detailed ontological and thematic relations.

Simultaneously, our adoption of an LSTM model carried out sentiment analysis to uncover the emotions embedded in the Gita. For the sentiment mapping across the 18 chapters it achieved 76% accuracy and F1 score of 0.42, proving the system's consistent mapping of emotional shifts. By taking advantage of LSTM, the system gained better contextual insight than ever before, that enabled the formation of an emotional trajectory to generate a dynamic and insightful analysis.

Our all-in-one effort that includes semantic relation mapping and sentiment tracking yields a system that is uniform, comprehensible, and promotive of research and philosophical study. Our approach lays a further groundwork for the computational analysis of historic literature.

To develop our system, introducing Sanskrit verses will be helpful in bilingual comparative analysis, while the use of models such as BERT or RoBERTa will help enhance relation extraction accuracy. In addition, the implementation of an interactive web based dashboard will improve accessibility and usability of the system for both academic and non academic users.

5.References

1. **Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013).**
Distributed Representations of Words and Phrases and their Compositionality.
In *Advances in Neural Information Processing Systems* (NeurIPS 2013).
https://papers.nips.cc/paper_files/paper/2013/file/9aa42b31882ec039965f3c4923ce901b-Paper.pdf

2. **Hochreiter, S., & Schmidhuber, J. (1997).**
Long Short-Term Memory.
Neural Computation, 9(8), 1735–1780.
https://doi.org/10.1162/neco.1997.9.8.1735

3. **NetworkX Developers. (2023).**
NetworkX Documentation – Python Tools for Graph Analysis.
https://networkx.org/documentation/stable/

4. **Jurafsky, D., & Martin, J. H. (2023).**
Speech and Language Processing (3rd ed. draft).
Chapter 20 on Information Extraction.
https://web.stanford.edu/~jurafsky/slp3/

5. **P C Shilpa; Rissa Shereen; Susmi Jacob; P Vinod (2021).**
Sentiment Analysis of Text Using Deep Learning.
https://doi.org/10.1109/ICICV50876.2021.9388382

6. **Bhagavad Gita – English Translation by Swami Sivananda.**
Divine Life Society, Rishikesh.
https://www.dlshq.org/books/bhagavad-gita/

7. **Bilal Abu-Salih & Salihah Alotaibi (2024)**
A systematic literature review of knowledge graph construction and application in education
https://doi.org/10.1016/j.heliyon.2024.e25383