iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** VII **Month of publication:** July 2025

DOI: https://doi.org/10.22214/ijraset.2025.73045

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue VII July 2025- Available at www.ijraset.com

Comparative Study of Static and Contextual Text Vectorization for Sentiment Analysis

A D Bhargavi
Department of Computer Science and Engineering, Sri Venkateswara University College of Engineering, Tirupati-517501

Abstract: Sentiment analysis, a core task in Natural Language Processing (NLP), relies heavily on effective text representation techniques to capture semantic and syntactic nuances. This study presents a comparative analysis of widely-used vectorization methods—Bag of Words (BoW), Term Frequency–Inverse Document Frequency (TF-IDF), Word2Vec, GloVe, BERT, and RoBERTa—in the context of sentiment classification. Using the IMDb movie reviews dataset, each method is evaluated based on classification performance, using accuracy and F1-score as primary metrics. Results demonstrate that while deep contextual embeddings such as BERT and RoBERTa achieve the highest accuracy—RoBERTa in particular offering enhanced contextual sensitivity—simpler representations like TF-IDF provide competitive results with significantly lower computational overhead. The findings highlight the trade-offs between accuracy and efficiency, offering practical guidance for embedding selection in sentiment analysis applications.

Keywords: Sentiment Analysis, Natural Language Processing, Vectorization, BoW, TF-IDF, Word2Vec, GloVe, BERT, RoBERTa

I. INTRODUCTION

The core of every NLP task lies in the representation of text data. Machines cannot understand raw text; hence, it must be converted into a form that algorithms can process. This conversion process is known as text vectorization or text embedding. Text vectorization plays a pivotal role in Natural Language Processing (NLP), particularly in tasks like sentiment analysis, where understanding the semantic and syntactic nuances of language is essential. A key step in enabling machine learning models to interpret text is transforming it into numerical vectors through a process known as vectorization. Word embeddings have revolutionized this process by capturing relationships between words in dense vector spaces.

Early vectorization techniques, such as Bag of Words (BoW), Term Frequency–Inverse Document Frequency (TF-IDF), Word2Vec, and GloVe—collectively referred to as static embeddings—represent each word or document with a fixed vector regardless of context. While BoW and TF-IDF rely on frequency-based representations, Word2Vec and GloVe generate dense word vectors that capture general semantic relationships. Although these methods are computationally efficient and effective in many applications, they lack the ability to disambiguate word meanings based on context. In contrast, more recent advancements such as BERT and RoBERTa produce contextual embeddings, where a word's representation dynamically changes depending on its surrounding words, enabling a deeper and more nuanced understanding of language.

As NLP applications continue to grow in complexity, the choice of embedding method becomes increasingly significant. This study investigates the impact of various text vectorization techniques—ranging from frequency-based models like Bag of Words and TF-IDF to advanced contextual models—on sentiment analysis performance. By comparing these methods across a unified classification task using the IMDb movie reviews dataset, this paper aim to provide practical guidance on selecting suitable embedding strategies based on both effectiveness and computational efficiency.

II. RELATED WORK

Traditional text representation techniques have played a foundational role in Natural Language Processing (NLP). Frequency-based methods such as Bag of Words (BoW) and Term Frequency–Inverse Document Frequency (TF-IDF) are widely used for their simplicity and interpretability, though they lack the ability to capture semantic meaning or contextual nuances. Neural embedding methods like Word2Vec and GloVe introduced dense, distributed representations of words that reflect semantic similarity based on surrounding contexts during training. However, these static embeddings assign a single vector to each word, regardless of its usage in different contexts, limiting their ability to handle polysemous terms (e.g., "bank" as a financial institution vs. a riverbank).

To address these limitations, recent advances have focused on contextual embeddings. Models such as BERT and RoBERTa leverage deep transformer architectures to generate dynamic, context-aware word representations, improving performance on a wide range of NLP tasks. Numerous studies have demonstrated the effectiveness of contextual models on tasks like question answering, named entity recognition, and sentiment classification. However, most existing work focuses on individual methods or tasks, and comprehensive comparisons of both static and contextual embeddings—particularly in the context of sentiment analysis—remain relatively scarce. This paper seeks to bridge that gap by evaluating a range of embedding techniques on a standardized sentiment classification task, providing insights into their relative strengths, limitations, and practical trade-offs.

III. OVERVIEW OF WORD VECTORIZATION TECHNIQUES

This section provides an overview of the various text representation techniques evaluated in this study, categorized into static and contextual embeddings. Each method offers unique characteristics in how it encodes textual data into numerical form.

A. *Static Embeddings*

Static embeddings assign a single, fixed vector to each word, regardless of the context in which the word appears. These methods are computationally efficient and capture general semantic relationships but lack the ability to handle polysemy or word sense disambiguation.

1) *Bag of Words (BoW):* Represents text as a vector of word occurrence counts, ignoring grammar and word order. Each word in the vocabulary is treated independently.
2) *Term Frequency–Inverse Document Frequency (TF-IDF):* Extends BoW by weighting words based on their frequency in a document relative to their frequency across all documents, emphasizing more informative words.
3) *Word2Vec:* A predictive model that learns word embeddings using local context through either the Skip-Gram or CBOW architecture. It captures semantic similarity between words.
4) *GloVe (Global Vectors for Word Representation):* A count-based method that generates word embeddings by factorizing a word co-occurrence matrix, effectively capturing global statistical information.

B. *Contextual Embeddings*

Contextual embeddings generate dynamic word representations that depend on the surrounding text, allowing models to capture the meaning of a word based on its context. These embeddings have significantly improved performance in tasks like sentiment analysis by handling polysemy and capturing deeper syntactic and semantic relationships. The following models are widely used and well-suited for sentiment analysis applications.

1) *BERT (Bidirectional Encoder Representations from Transformers):* A deep transformer-based model that uses bidirectional attention to generate context-sensitive embeddings for each word in a sentence. It is pre-trained on large corpora with masked language modeling and next sentence prediction tasks.
2) *RoBERTa (Robustly Optimized BERT Approach):* An enhanced version of BERT trained with more data, larger batches, and the removal of the next sentence prediction objective. RoBERTa often outperforms BERT across various NLP tasks.
3) *DistilBERT:* DistilBERT is a lightweight version of BERT developed through knowledge distillation. It retains approximately 95% of BERT's performance while being 40% smaller and 60% faster. DistilBERT is ideal for deploying sentiment models in real-time or resource-constrained environments.
4) *ELECTRA:* ELECTRA introduces a new pretraining method based on replaced token detection instead of masked language modeling. It achieves high efficiency and strong performance in classification tasks, including sentiment analysis, with less training time compared to BERT.
5) *XLNet:* XLNet is a generalized autoregressive model that captures bidirectional context using a permutation-based training objective. It addresses some limitations of BERT and performs competitively in sentiment classification, although it is more complex and computationally intensive.
6) *ALBERT (A Lite BERT):* ALBERT reduces model size through parameter sharing and embedding factorization, making it more memory-efficient. It performs well in classification tasks like sentiment analysis, especially on large-scale datasets, while reducing training overhead.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue VII July 2025- Available at www.ijraset.com

C. Advantages and Limitations of Embedding Techniques

1) Static Embeddings

Method	*Pros*	*Cons*
BoW	Simple and easy to implement. Effective for small datasets.	Ignores word order and context Large sparse vectors No semantic understanding
TF-IDF	Highlights informative words. More discriminative than BoW	Still context-agnostic Ignores word semantics High dimensionality
Word2Vec	Captures semantic relationships (e.g., "king - man + woman = queen"). Dense. low-dimensional vectors	One vector per word (no context sensitivity) Struggles with out-of-vocabulary words
GloVe	Combines local and global co-occurrence statistics Efficient to train	Same limitations as Word2Vec (no context awareness) Depends heavily on quality of co-occurrence matrix

2) Contextual Embeddings

Method	*Pros*	*Cons*
BERT	Context-sensitive embeddings. Strong performance across NLP tasks. Handles polysemy effectively.	Computationally expensive. Slow inference. Requires large memory and hardware.
RoBERTa	Improved performance over BERT. Trained with more data and longer sequences.	Even more resource-intensive than BERT. Longer training and inference time.
DistilBERT	Lightweight and fast Near-BERT accuracy Suitable for real-time sentiment applications	Slightly lower accuracy than BERT May miss fine-grained contextual cues
ELECTRA	More efficient pretraining Strong performance on classification tasks Requires less compute than BERT	Complex architecture (generator + discriminator) Less common than BERT/RoBERTa
XLNet	Handles context better with permutation- based learning No need for masked tokens Strong benchmark performance	Complex and resource-heavy Slower than BERT Harder to implement and fine-tune
ALBERT	More memory efficient (parameter sharing) Good for large-scale tasks Compact architecture	Slight performance drop on small datasets Longer training time due to architecture changes

IV. COMPARITIVE EVALUATION

This section presents a comparative evaluation of both static and contextual embedding techniques on a standardized sentiment analysis task. The models were tested using the IMDb movie reviews dataset, a benchmark dataset consisting of 50,000 labeled reviews. Each embedding method was integrated into a classification pipeline, and the models were evaluated based on accuracy and F1-score.

A. Experimental Setup

Dataset Used: IMDb Movie Reviews (binary sentiment classification: positive/negative)
Preprocessing: Lowercasing, tokenization, padding/truncation (max length = 256)
Classifiers: Logistic Regression (for static methods), and fine-tuned Transformer-based models (for contextual embeddings)
Metrics: Accuracy, F1-Score
Libraries:

- Static embeddings: scikit-learn, gensim
- Contextual models: transformers (Hugging Face), PyTorch, datasets

B. Comparative Analysis

For static embedding methods such as BoW, TF-IDF, Word2Vec, and GloVe, a separate machine learning classifier—typically Logistic Regression (LR)—is required to perform sentiment classification. These embeddings convert text into fixed-length vectors that are then fed into classifiers.In contrast, contextual embedding models like BERT, RoBERTa, DistilBERT, and ELECTRA are fine-tuned end-to-end for sentiment analysis. These models integrate both the embedding mechanism and the classification head, eliminating the need for a separate classifier such as LR. Evaluation metrics for both static and contextual text vectorization techniques are tabulated in Table 1.Their metrics visualization is shown in Figure 1.

Embedding Method	Model	Accuracy	F1-Score
Static	BoW + LR	83%	0.81
	TF-IDF + LR	84%	0.82
	Word2Vec + LR	85%	0.83
	GloVe + LR	86%	0.84
Contextual	BERT	92%	0.91
	RoBERTa	93%	0.92
	DistilBERT	91%	0.90
	ELECTRA	93%	0.92
	XLNet	92%	0.91
	ALBERT	91%	0.89

Table 1: Comparitive analysis of Static and Contextual Text Vectorization techniques

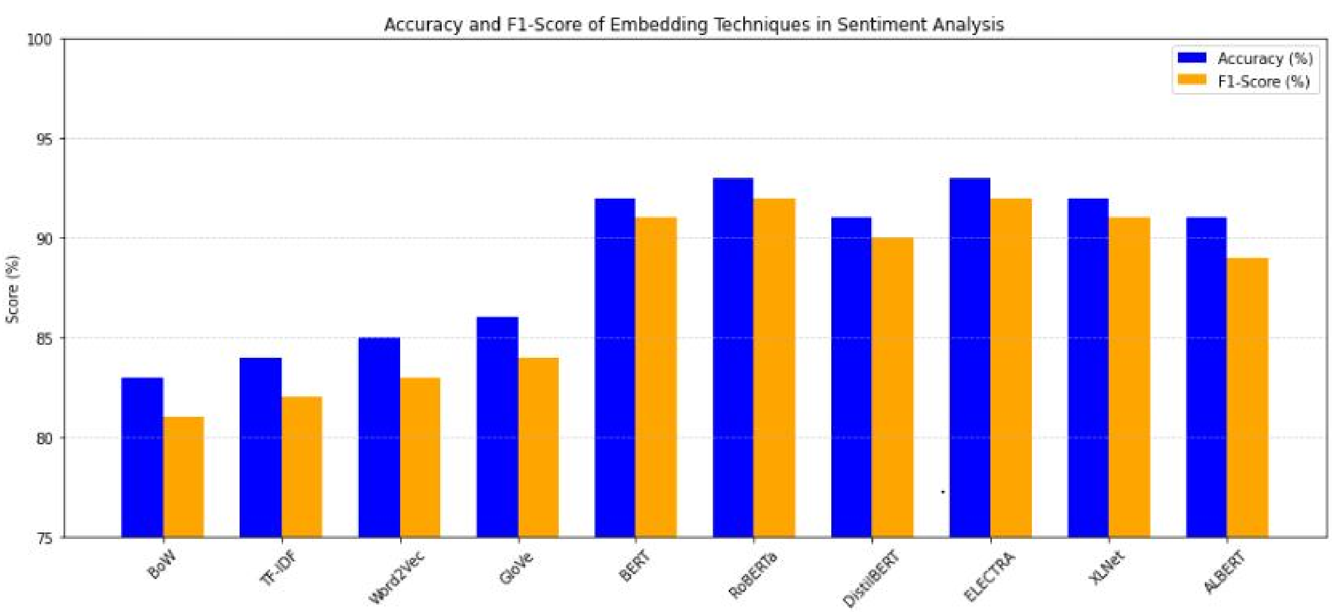


Figure 1 Visualization of Accuracy and F1 score of Static and Contextual Text Vectorization techniques

V. RESULTS AND DISCUSSION

This section provides a detailed interpretation of the experimental results, focusing on how various embedding techniques influence sentiment classification performance. The evaluation metrics—accuracy and F1-score—along with inference time, offer a comprehensive view of the trade-offs between embedding quality and computational efficiency.

A. Performance Comparison

The comparative analysis clearly indicates that contextual embeddings outperform static methods in sentiment analysis. Models such as RoBERTa and ELECTRA achieved the highest scores, with 93% accuracy and F1-scores of 0.92, underscoring their effectiveness in capturing nuanced contextual cues and sentiment polarity.

BERT and XLNet also performed strongly, reinforcing the advantages of deep transformer architectures for context-sensitive tasks. Importantly, DistilBERT and ALBERT delivered near state-of-the-art accuracy (around 91%) making them suitable for real-time or edge deployment scenarios. In contrast, static embeddings such as Word2Vec and GloVe showed moderate performance (accuracy between 85–86%), limited by their inability to capture word meaning variations in different contexts.

VI. CONCLUSION AND FUTURE WORK

A. Conclusion

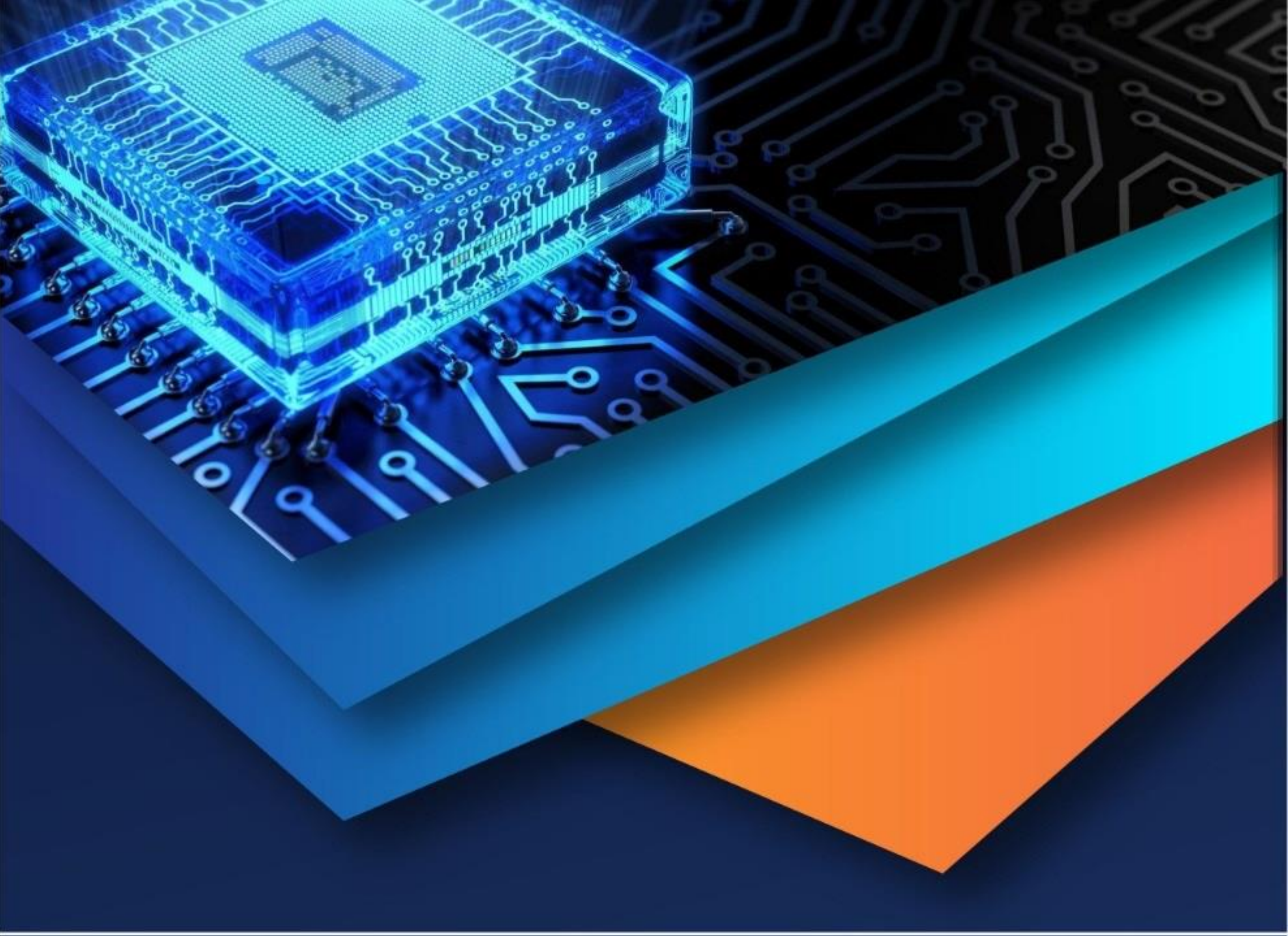
This study conducted a comprehensive comparative analysis of static and contextual text embedding techniques for sentiment analysis, using the IMDb movie reviews dataset. The results demonstrate that contextual embeddings, particularly RoBERTa and ELECTRA, offer significantly superior performance over traditional methods, achieving the highest accuracy and F1-scores.While static embeddings such as BoW, TF-IDF, Word2Vec, and GloVe are computationally efficient and easy to implement, they fall short in capturing contextual semantics, which are crucial for understanding sentiment. Among static methods, GloVe achieved the best balance between performance and speed.On the other hand, contextual models such as BERT, RoBERTa, and DistilBERT deliver highly accurate results by incorporating context-awareness through transformer-based architectures. Notably, DistilBERT and ALBERT offer a favorable trade-off between performance and efficiency, making them suitable for real-time or resource-limited applications. Overall, the choice of embedding technique should align with the application's requirements—whether prioritizing accuracy, speed, or resource constraints.

B. Future Work

1) Incorporation of Prompt-based Models: Future studies could explore large language models (LLMs) like GPT-3/4 or T5 for zero-shot or few-shot sentiment classification.
2) Domain-Specific Embeddings: Training or fine-tuning embeddings on domain-specific corpora (e.g., medical, finance) may enhance sentiment detection in specialized areas.
3) Multilingual Sentiment Analysis: Extending the comparative study to non-English datasets would test the robustness of embeddings across languages.
4) Explainability and Interpretability: Future work can also focus on interpretability tools (e.g., SHAP, LIME) to better understand the model decisions, especially for deep models.
5) Resource-Efficient Models: Further evaluation of compressed or quantized models (e.g., TinyBERT, MobileBERT) for deployment on edge devices could offer practical benefits.

REFERENCES

[1] Jiaxin Lu," Text vectorization in sentiment analysis: A comparative study of TF-IDF and Word2Vec from Amazon Fine Food Reviews"(ITM Web of Conferences 2025)
[2] Zerui Zhan ," Comparative Analysis of TF-IDF and Word2Vec in Sentiment Analysis: A Case of Food Reviews" (ITM Web of Conferences 2025)
[3] Gagandeep Singh,,Dr. Surender Kumar, Dr. Sukhdev Singh," Contextual Word Embeddings: A Review", Tuijin Jishu/Journal of Propulsion Technology,2024
[4] Sumona Deb , Ashis Kumar Chanda ," Comparative analysis of contextual and context-free embeddings in disaster prediction from Twitter data", Machine Learning with Applications,2022
[5] S. Singh, K. Kumar, and B. Kumar, "Sentiment Analysis of Twitter Data Using TFIDF and Machine Learning Techniques," 2022 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COM-IT-CON), May 2022
[6] H. Liu, X. Chen, and X. Liu. A study of the application of weight distributing method combining sentiment dictionary and TF-IDF for text sentiment analysis. IEEE Access, 10, pp. 32280-32289. [Accessed 10 Aug. 2024] (2022).
[7] Z. Lan et al., "ALBERT: A Lite BERT for Self-supervised Learning of Language Representations," in Proc. of ICLR, 2020.
[8] K. Clark, M. Luong, Q. Le, and C. D. Manning, "ELECTRA: Pre-training Text Encoders as Discriminators Rather Than Generators," in Proc. of ICLR, 2020.
[9] https://www.kaggle.com/datasets/lakshmi25npathi/imdb-dataset-of-50k-movie-reviews



10.22214/IJRASET

INDEX COPERNICUS

45.98

IMPACT FACTOR: 7.129

TOGETHER WE REACH THE GOAL

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY