IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 4, July 2026
Impact Factor: 8.2

Comparative Evaluation of Machine Learning and Deep Learning Models for Targeted Metaphor Detection

Muhammad Hassam Aslam Khan[1] and Ninad Deshpande[2]
University of Arkansas at Little Rock, Arkansas, USA [1]
Indiana University Indianapolis, Indianapolis, Indiana, USA. [2]
mkhan5@ualr.edu, ninad9021@gmail.com

Abstract: *Metaphor detection is an important natural language processing task because figurative language often conveys meanings that differ from literal word usage. This paper evaluates multiple machine learning and deep learning approaches for targeted metaphor detection, where the goal is to classify whether a predefined candidate word is used metaphorically or literally within a given textual context. Using a labeled dataset of 1,870 training samples and 800 test samples covering seven target words—road, candle, light, spice, ride, train, and boat—we compare TF-IDF-based logistic regression, LDA combined with Sentence-BERT embeddings and random forests, LSTM-based neural models, BERT-enhanced LSTM representations, XGBoost with Word2Vec embeddings, and a majority-vote consensus model. The LDA-SBERT-Random Forest model achieved the strongest overall performance, with 84% accuracy and a weighted F1-score of 0.82. The consensus model achieved 82% accuracy and a weighted F1-score of 0.80, suggesting that ensemble voting provided balanced predictions but did not outperform the best individual model. These results indicate that hybrid feature representations combining contextual sentence embeddings with topic-level information can be effective for metaphor detection in small, labeled datasets.*

Keywords: Metaphor Detection, Natural Language Processing, Machine Learning, Deep Learning, Sentence-BERT, Ensemble Learning

I. INTRODUCTION

Metaphor detection is an important task in natural language processing because metaphorical expressions often convey meanings that differ from their literal interpretation. A word such as “road,” “light,” or “train” may be used literally in one context and metaphorically in another. This makes metaphor detection a context-dependent classification problem that requires models to capture semantic and linguistic cues beyond simple keyword matching.

This study evaluates machine learning and deep learning models for targeted metaphor detection. The task is to determine whether the first occurrence of a predefined candidate word is used metaphorically or literally within a given text. The dataset contains 1,870 labeled training samples and 800 test samples covering seven target words: “road,” “candle,” “light,” “spice,” “ride,” “train,” and “boat.”

Metaphor detection is challenging because figurative language is nuanced, context-sensitive, and often shaped by cultural and discourse-level meaning. Traditional manual annotations are time-consuming, subjective, and difficult to scale to large datasets. Automated metaphor detection can therefore support broader NLP applications, including literary analysis, sentiment interpretation, dialogue systems, and semantic understanding.

To address this task, we compare several modeling approaches, including TF-IDF-based logistic regression, LDA combined with Sentence-BERT embeddings and Random Forest classification, LSTM-based neural models, LSTM with BERT embeddings, XGBoost with Word2Vec, and a majority-vote consensus model. Prior work emphasizes that

DOI: 10.48175/IJARSCT-37941

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 4, July 2026
Impact Factor: 8.2

metaphorical language requires attention to contextual meaning and semantic representation [4]. Our results show that the LDA-SBERT-Random Forest model achieves the strongest overall performance, suggesting that hybrid feature representations are useful for metaphor detection in small labeled datasets.

II. RESEARCH CONTRIBUTIONS

This study makes three main contributions. First, it provides a comparative evaluation of traditional machine learning, deep learning, contextual embedding, and ensemble approaches for targeted metaphor detection. Second, it examines how lexical, topic-level, sequential, and contextual representations differ in their ability to distinguish literal and metaphorical usage across seven target words. Third, the findings show that combining LDA topic features with SBERT contextual embeddings and Random Forest classification produces the strongest overall accuracy, while simple majority voting provides balanced predictions but does not necessarily outperform the best individual model

III. RELATED WORK

This section introduces some of the previous works that either inspire or contribute to the general mentality and methodologies of our project.

"Multi-co-training for document classification using various document representations: TF–IDF, LDA, and Doc2Vec" (Kim, 2019) provides a brief overview of the research area of document classification. The authors proposed a multi-co-training framework and compare and contrast three different document representation methods: term frequency–inverse document frequency (TF–IDF) based on the bag-of-words scheme, topic distribution based on latent Dirichlet allocation (LDA), and neural-network-based document embedding known as document to vector (Doc2Vec) [12].

Earlier metaphor detection research moved from rule-based and lexical approaches toward supervised and embedding-based systems. Shutova [17] reviews computational approaches to metaphor processing, while Tsvetkov et al. [18] show that semantic features can support metaphor classification. More recent shared tasks on VUA and TOEFL metaphor detection show that metaphor identification remains a challenging context-dependent classification problem across datasets and domains [13].

"Unsupervised sentence representations as word information series: Revisiting TF–IDF" (Arroyo-Fernández, 2019) provides a review of the TF-IDF based approach, relative to unsupervised embedding methods that aim at learning unsupervised sentence representations from unlabeled text [1].

"Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet allocation" (Guo, 2017) discusses latent Dirichlet allocation as a data mining approach and applies it to online ratings and reviews, based on hotel-visitor interaction. It uncovers up to 19 controllable dimensions that are key for hotels to manage their interactions [6].

"Public discourse and sentiment during the COVID 19 pandemic: Using Latent Dirichlet Allocation for topic modeling on Twitter" (Xue, 2020) aims to understand Twitter users' discourse and psychological reactions to COVID-19, using a Latent Dirichlet Allocation based machine learning approach. Eleven topics are identified and organized. It was found that fear for the unknown nature of the virus dominates all topics [19].

"Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks" (Reimers, 2018) introduces SBERT: a sentence embedding technique that uses siamese and triplet network structure to generate vector embeddings for meaningful sentences. It excels at multiple NLP tasks such as semantic similarity search and clustering. Under the SentEval framework, SBERT achieved around 0.807 accuracy [16].

"Long Short-Term Memory" (Hochreiter, 1997) introduces a neural network structure called LSTM, which can effectively solve the problem of gradient disappearance during back propagation in traditional RNN, so that it can learn information at long intervals. The article introduces the structure and principles of LSTM in detail, and compares its performance with other RNN in solving various tasks through experiments. The LSTM structure achieves average accuracy of 0.98 in a variety of RNN oriented evaluation tasks [8].

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 4, July 2026
Impact Factor: 8.2

"Hybrid Attention Based Neural Architecture for Text Semantics Similarity Measurement" (Liu, 2019) proposes a neural network structure with hybrid attention to highlight important text fragments. A Bi-directional LSTM network, together with a hybrid attention mechanism on top of it, was applied to encode each sentence. Self-attention is a crucial component to generate sentence level representations [14].

"Sentiment and Context-Aware Hybrid DNN With Attention for Text Sentiment Classification" (Kha) proposes a neural network model for sentimental and contextual perception with an attention mechanism, that can learn and highlight the significant features of the relevant material in the text. Researchers firstly use a wide-coverage emotional dictionary to identify the sentiment features, and then use a bi-directional encoder representation to generate word embeddings for text semantic extraction. The BiLSTM method is used to capture the semantic information of word context and the long dependency in word sequence. The results of SCA-HDNN are 93.5%, 94.7%, 91.9%, 94.7%, 84.5% for MR, RT-2k, SST-2, CR, SemEval 2013, and SemEval 2014 datasets [9].

"BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding" (Devlin, 2018) introduces a new language representation model BERT, which is a pre-trained model of a deep bi-directional transformer. It can create the most advanced model by pre-training in all layers in contexts without a large number of task-specific architecture modifications, such as questions and answers and language reasoning. Results on eleven NLP tasks are: improving the GLUE score to 80.5; MultiNLI accuracy to 86.7%; SQuAD v1.1 question answering Test F1 to 93.2% and SQuAD v2.0 Test F1 to 83.1% [5].

"XGBoost: A Scalable Tree Boosting System" (Chen, 2016) introduces XGBoost, a scalable tree boosting system, which is an efficient and widely used machine learning method. The main contributions of XGBoost include: proposing a sparse perception algorithm for sparse data; providing insights into access pattern and data compression to build a scalable tree enhancement system. This article also discusses the most advanced results achieved by XGBoost in various machine learning challenges [2].

"Comparison of Naïve Bayes Algorithm and XGBoost on Local Product Review Text Classification" (Hendrawan, 2022) aims to tackle the challenge of product review based text classification, by comparing Naïve Bayes against XGBoost on top of vector embeddings of word2vec and TF-IDF. Word2vec with XGBoost scores the highest F1 score of 0.941 [7].

Recent applied NLP studies also show how transformer and LLM-based methods can be combined with interpretable scoring frameworks for domain-specific text analysis, including trust-based review ranking and empathy measurement in online counseling [11, 10].

IV. DATASET

The dataset used in this project consists of 1,870 training samples and 800 testing samples, each designed to identify metaphorical usage of specific words within textual contexts. Each sample includes a metaphorID that corresponds to a candidate metaphor word, a binary label indicating whether the word is used metaphorically (TRUE) or literally (FALSE), and the textual context in which the word appears.

The following table provides the mapping of metaphorID to specific metaphor words. These identifiers are crucial for analyzing the context in which these words are used, helping to determine whether their usage is metaphorical or literal.

Table I: Mapping of metaphorID to Metaphor Words Used in the Dataset

metaphorID	Metaphor Word
0	road
1	candle
2	light
3	spice
4	ride
5	train

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 4, July 2026
Impact Factor: 8.2

6	boat

To further understand the dataset, we analyzed the distribution of metaphorical and literal labels. The bar chart in Fig. 1 shows the counts for each label, indicating the balance between metaphorical and literal samples in the dataset.

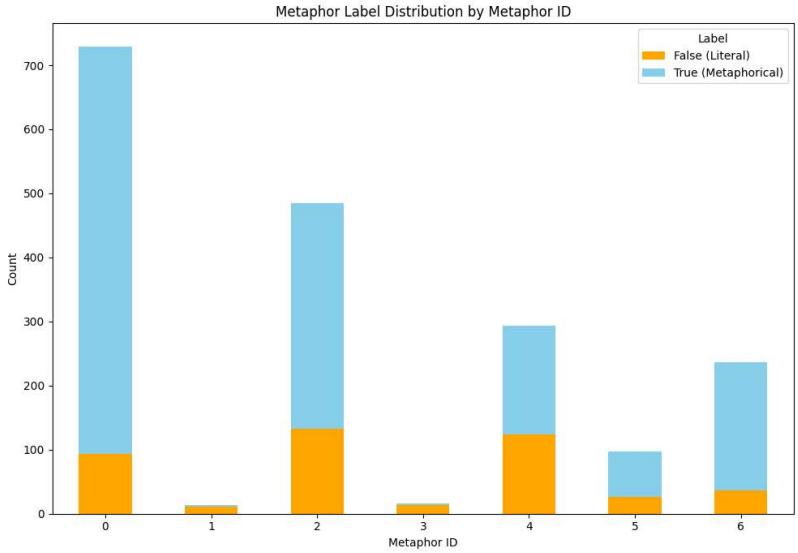

Fig. 1 Distribution of labels in the dataset (TRUE: Metaphorical, FALSE: Literal)

V. METHODOLOGY

This section describes the machine learning and deep learning methods used for targeted metaphor detection, as shown in Fig. 2. Each subsection explains the model architecture, feature representation, and role of the method in classifying target words as metaphorical or literal. The section concludes with the consensus method, which combines predictions from the individual models using majority voting.

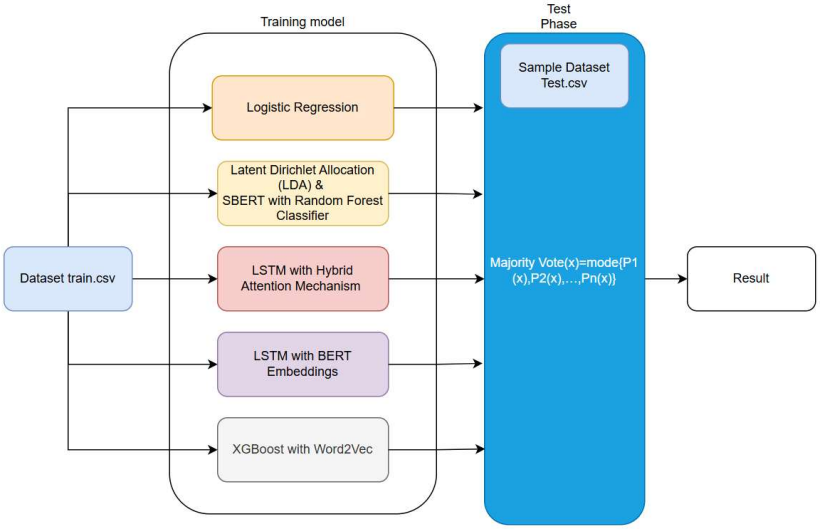

Fig. 2 High-level architecture of the metaphor detection pipeline

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 4, July 2026
Impact Factor: 8.2

A. Training Setup

The training process begins with text preprocessing, including cleaning, tokenization, and feature extraction. Each model then applies a different representation strategy to capture lexical, semantic, or contextual information relevant to metaphor detection.

B. Logistic Regression

The Logistic Regression model identifies metaphorical usage of specific words (e.g., "road," "candle," "light," "spice," "ride," "train," "boat") by leveraging TF-IDF vectorization and logistic regression. This model preprocesses text through tokenization and lemmatization, aligning candidate words to metaphor IDs to capture contextual nuances. Features are extracted using a TF-IDF Vectorizer, emphasizing both unigrams and bigrams, with a limit of 5000 features to optimize both performance and computational efficiency. According to Razali et al., logistic regression can effectively discern metaphorical language when complemented by contextual feature engineering, thus enhancing the model's accuracy in identifying metaphors across various contexts [15]. The logistic regression is integrated into a pipeline that adjusts for class imbalances, ensuring that each class is weighted appropriately during training. Post-training, the model's effectiveness is evaluated using precision, recall, and F1-scores, and the trained model is saved for deployment and further validation.

C. Latent Dirichlet Allocation (LDA) & SBERT with Random Forest Classifier

This hybrid model integrates LDA, SBERT embeddings, and a Random Forest Classifier to detect metaphors involving specific keywords. Initially, the model preprocesses text by mapping these keywords to corresponding metaphor IDs and extracting a context window around each keyword occurrence. LDA is then applied to this processed text to identify underlying topics, providing a thematic structure to the embeddings. Concurrently, SBERT generates dense embeddings that capture deep semantic meanings, which are combined with the LDA features. A Random Forest Classifier, trained on these combined features, classifies text into metaphorical or literal categories. The model's performance is quantified through a classification report, assessing its precision, recall, and F1-scores.

D. LSTM with Hybrid Attention Mechanism

This model employs a Long Short-Term Memory (LSTM) network augmented with a novel Hybrid Attention mechanism to enhance metaphor detection in texts containing predefined keywords. The LSTM processes sequences of word embeddings, providing a dynamic memory model that captures temporal dependencies within text data. The innovative Hybrid Attention layer combines additive and dot-product attention mechanisms, allowing the model to focus on specific elements of the input sequence that are most relevant for identifying metaphors. Chen et al. discuss how integrating keyword similarity computations with LSTM models can significantly improve the identification of metaphoric content by leveraging both local feature analysis and broader contextual understanding, which aligns with our approach to enhancing the LSTM's sensitivity to subtle metaphorical nuances [3]. Text data is preprocessed through tokenization and lemmatization to ensure uniformity before being fed into the model. The attention mechanism prioritizes important tokens dynamically, improving the model's ability to discern subtle metaphorical nuances. Training involves adjusting attention weights and LSTM parameters to optimize performance, with results evaluated based on accuracy and loss metrics.

E. LSTM with BERT Embeddings

This model combines LSTM networks with BERT embeddings to leverage deep contextual representations for enhanced metaphor detection in texts. The BERT model preprocesses text to produce embeddings that encapsulate a rich understanding of language nuances, particularly beneficial for detecting metaphors involving predefined keywords.
The LSTM layer processes these embeddings, capturing both the short-term and long-term dependencies within the text. This integration allows the model to maintain context over longer sequences, improving its ability to recognize complex

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 4, July 2026
Impact Factor: 8.2

metaphorical expressions. Training the model involves tuning the LSTM parameters to optimize the handling of BERT's high-dimensional embeddings, with a focus on balancing the learning rate and avoiding overfitting.
Model performance is evaluated using standard metrics such as accuracy and recall, providing insights into its effectiveness in metaphor detection tasks.

F. XGBoost with Word2Vec

The XGBoost model integrated with Word2Vec embeddings utilizes advanced machine learning techniques for detecting metaphors in text. This approach harnesses the power of Word2Vec to generate dense word embeddings, capturing semantic and syntactic word relationships, which are particularly useful for identifying metaphors involving specific keywords. The XGBoost algorithm, known for its efficiency and effectiveness in classification tasks, processes these embeddings to predict whether a text contains metaphorical expressions. The model is trained using a gradient boosting framework that optimizes decision trees sequentially, thus reducing errors and enhancing predictive accuracy.
Performance is assessed through a classification report that evaluates precision, recall, and F1-score, offering a comprehensive view of the model's capability to differentiate between literal and metaphorical usage.

VI. RESULTS AND DISCUSSION

A. Testing Strategy

The dataset was split into training and testing subsets using an 80/20 ratio, as implemented in split.py. The training data (temp_train.csv) was used to train individual machine learning models, while the testing data (temp_test.csv) was reserved for evaluating model performance. This ensures that the models are validated on unseen data, thereby providing an unbiased measure of their effectiveness. Each model generates predictions on the test data, and their results are summarized in classification reports.

B. Individual Results

The performance of each model is evaluated using standard classification metrics, including precision, recall, weighted F1-score, macro F1-score, and accuracy. Precision and recall are reported in Table II. Weighted F1-scores are reported in Table III, macro F1-scores in Table IV, and accuracy scores in Table V. A summary of the results for each model is provided below.
Logistic Regression achieved a weighted F1-score of 0.81 and an accuracy of 80%, with strong recall for metaphorical labels but slightly lower precision for literal labels.
LDA-SBERT-RF performed well with a weighted F1-score of 0.82 and an accuracy of 84%, demonstrating high precision for literal labels and excellent recall for metaphorical labels.
LSTM with Custom Layer struggled to balance recall and precision, resulting in a weighted F1-score of 0.52 and an accuracy of 49%.
LSTM with BERT Embeddings achieved a weighted F1-score of 0.72 and an accuracy of 73%, effectively capturing metaphorical language but with lower performance for literal labels.
XGBoost with Word2Vec scored a weighted F1-score of 0.64 and an accuracy of 74%, excelling in recall for metaphorical labels but significantly underperforming on literal labels.

C. Consensus Results

To evaluate whether ensemble voting improved prediction stability, we developed a consensus model that combines the outputs of all individual models. The consensus method selects the mode of the predictions generated by the individual classifiers. This approach achieved a weighted F1-score of 0.80 and an accuracy of 82%, with balanced precision and recall across the two classes.
However, the consensus model did not outperform the strongest individual model. The LDA-SBERT-RF model achieved the best overall accuracy at 84% and the highest weighted F1-score at 0.82. This suggests that simple majority voting can

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 4, July 2026
Impact Factor: 8.2

improve balance across predictions, but it may not improve overall performance when some individual models are substantially weaker than others.

Table II: Precision and Recall for Individual and Consensus Metaphor Detection Models

Model	Precision	Recall
Logistic Regression	0.82	0.80
LDA-SBERT-RF	0.85	0.84
LSTM (Custom Layer)	0.64	0.49
LSTM (BERT)	0.71	0.73
XGBoost Word2Vec	0.81	0.74
Consensus Model	0.82	0.82

Table III: Weighted F1-Scores for Individual and Consensus Metaphor Detection Models

Model	F1 (Weighted)
Logistic Regression	0.81
LDA-SBERT-RF	0.82
LSTM (Custom Layer)	0.52
LSTM (BERT)	0.72
XGBoost Word2Vec	0.64
Consensus Model	0.80

Table IV: Macro F1-Scores for Individual and Consensus Metaphor Detection Models

Model	F1 (Macro)
Logistic Regression	0.76
LDA-SBERT-RF	0.74
LSTM (Custom Layer)	0.48
LSTM (BERT)	0.62
XGBoost Word2Vec	0.44
Consensus Model	0.71

Table V: Accuracy Scores for Individual and Consensus Metaphor Detection Models

Model	Accuracy
Logistic Regression	80%
LDA-SBERT-RF	84%
LSTM (Custom Layer)	49%
LSTM (BERT)	73%
XGBoost Word2Vec	74%

DOI: 10.48175/IJARSCT-37941

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 4, July 2026
Impact Factor: 8.2

Consensus Model	82%

VII. PRACTICAL IMPLICATIONS

The results suggest that hybrid contextual and topic-based representations are suitable for targeted metaphor detection. However, the competitive performance of logistic regression indicates that simpler models may remain useful when computational resources are limited. The lower performance of the consensus model compared with LDA-SBERT-RF also shows that ensemble effectiveness depends on the quality and complementarity of the included models.

VIII. LIMITATION

This study focuses on seven target words, allowing a controlled comparison of how different models distinguish literal and metaphorical meanings across varied contexts. However, performance may differ in other words, domains, or linguistic styles not represented in this evaluation. Future work can test the framework on additional target words and datasets to assess broader generalizability.

IX. CONCLUSION

This study evaluated several machine learning and deep learning models for targeted metaphor detection. The task was to classify whether predefined candidate words were used metaphorically or literally within a given textual context. We compared TF-IDF-based logistic regression, LDA-SBERT-Random Forest, LSTM with a custom attention layer, LSTM with BERT embeddings, XGBoost with Word2Vec, and a majority-vote consensus model.

The LDA-SBERT-Random Forest model achieved the strongest overall performance, with a weighted F1-score of 0.82 and an accuracy of 84%. Logistic regression also performed competitively, achieving a weighted F1-score of 0.81 and an accuracy of 80%. The consensus model achieved a weighted F1-score of 0.80 and an accuracy of 82%, producing balanced predictions but not outperforming the best individual model. This suggests that simple majority voting can improve prediction stability, but its effectiveness depends on the strength and complementarity of the models included in the ensemble.

Overall, the results indicate that hybrid representations combining contextual sentence embeddings with topic-level information can be effective for metaphor detection, particularly in small-data settings. Future work can extend this study by using larger and more diverse metaphor datasets, adding detailed error analysis by target word, and testing more advanced transformer-based or confidence-weighted ensemble methods.

DECLARATIONS

Competing Interests: The authors declare that they have no competing interests.

Funding: No external funding was received for this work.

Data Availability: The dataset used in this study is available from the authors upon reasonable request, subject to any restrictions associated with the original data source.

ACKNOWLEDGMENT

The authors thank the instructor and peers who provided feedback during the early development of this study.

REFERENCES

[1] Ignacio Arroyo-Fernández, Carlos-Francisco Méndez-Cruz, Gerardo Sierra, Juan-Manuel Torres-Moreno, and Grigori Sidorov, “Unsupervised sentence representations as word information series: Revisiting TF–IDF,” Computer Speech & Language, pp. 107–129, Jun. 2019.

[2] Tianqi Chen and Carlos Guestrin, “XGBoost: A scalable tree boosting system,” in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Aug. 2016.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 4, July 2026
Impact Factor: 8.2

[3] Zhiheng Chen, Lijun Fu, Hongjun Wang, and Yu-Jiang Liu, "A metaphor recognition model based on LSTM and keyword similarity computation," in 2021 33rd Chinese Control and Decision Conference (CCDC), pp. 2347–2352, 2021.
[4] Bogdan-Ionut Cirstea and Costin-Gabriel Chiru, "Metaphor detection," in 2013 19th International Conference on Control Systems and Computer Science, pp. 210–217, 2013.
[5] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in Proceedings of the 2019 Conference of the North, Dec. 2018.
[6] Yue Guo, Stuart J. Barnes, and Qiong Jia, "Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet allocation," Tourism Management, pp. 467–483, Mar. 2017.
[7] Ivan Rifky Hendrawan, Ema Utami, and Anggit Dwi Hartanto, "Comparison of Naïve Bayes algorithm and XGBoost on local product review text classification," Edumatic: Jurnal Pendidikan Informatika, pp. 143–149, Oct. 2022.
[8] Sepp Hochreiter and Jürgen Schmidhuber, "Long short-term memory," Neural Computation, pp. 1735–1780, Oct. 1997.
[9] Jawad Kha, Niaz Ahmad, Farman Ali, and Youngmoon Lee, "Sentiment and context-aware hybrid DNN with attention for text sentiment classification."
[10] Muhammad Hassam Aslam Khan, "A hybrid NLP framework for measuring therapist empathy in online counseling responses," Research Square preprint, 2026. DOI: https://doi.org/10.21203/rs.3.rs-9667616/v1.
[11] Muhammad Hassam Aslam Khan and Rajeev R. Raje, "Enhancing mobile app ranking through trust-based sentiment analysis using large language models," Research Square preprint, 2026. DOI: https://doi.org/10.21203/rs.3.rs-8943112/v1.
[12] Donghwa Kim, Deokseong Seo, Suhyoun Cho, and Pilsung Kang, "Multi-co-training for document classification using various document representations: TF–IDF, LDA, and Doc2Vec," Information Sciences, pp. 15–29, Feb. 2019.
[13] Chee Wee Leong, Beata Beigman Klebanov, Chris Hamill, Egon Stemle, Rutuja Ubale, and Xianyang Chen, "A report on the 2020 VUA and TOEFL metaphor detection shared task," in Proceedings of the Second Workshop on Figurative Language Processing, pp. 18–29, Association for Computational Linguistics, 2020.
[14] Kaixin Liu, Yong Zhang, and Chunxiao Xing, "Hybrid attention based neural architecture for text semantics similarity measurement," pp. 90–106, Dec. 2019.
[15] Md Saifullah Razali, Alfian Abdul Halin, Yang-Wai Chow, Noris Mohd Norowi, and Shyamala Doraisamy, "Deep learning metaphor detection with emotion-cognition association," in 2022 International Conference on Digital Transformation and Intelligence (ICDI), pp. 8–14, 2022.
[16] Nils Reimers and Iryna Gurevych, "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), Dec. 2018.
[17] Ekaterina Shutova, "Models of metaphor in NLP," in Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, pp. 688–697, Association for Computational Linguistics, 2010.
[18] Yulia Tsvetkov, Leonid Boytsov, Anatole Gershman, Eric Nyberg, and Chris Dyer, "Metaphor detection with cross-lingual model transfer," in Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, pp. 248–258, Association for Computational Linguistics, 2014.
[19] Jia Xue, Junxiang Chen, Chen Chen, Chengda Zheng, Sijia Li, and Tingshao Zhu, "Public discourse and sentiment during the COVID 19 pandemic: Using latent dirichlet allocation for topic modeling on Twitter," PLOS ONE, p. e0239441, Sep. 2020.