A COMPARATIVE AND EXPLAINABLE FRAMEWORK FOR SOCIAL MEDIA SENTIMENT ANALYSIS USING LEXICON-BASED, MACHINE LEARNING, DEEP LEARNING, AND TRANSFORMER MODELS

Nikita Vijay (M.Tech Scholar)[1] Mr. Irfan Khan (Assistant Professor)[2]
(Department Of Computer Science & Engineering)[1,2]
Shekhawati Institute of Engineering & Technology, Sikar(RJ)

Abstract

The exponential growth of social media platforms has resulted in the continuous generation of large volumes of unstructured textual data containing opinions, emotions, attitudes, and behavioural expressions. Existing studies have predominantly focused on individual sentiment-analysis techniques or specific model families, making direct comparison between traditional and advanced approaches difficult. This study proposes a comprehensive and explainable framework for social media sentiment analysis that systematically compares lexicon-based, traditional machine-learning, deep-learning, and Transformer-based approaches under a common experimental environment. The proposed framework integrates data acquisition, social-media-specific pre-processing, exploratory data analysis, feature representation, model development, sentiment classification, explainability, and performance evaluation. Lexicon-based techniques such as VADER and Text Blob are used as interpretable baselines. The proposed framework evaluates models using accuracy, precision, recall, macro-F1 score, weighted-F1 score, confusion matrices, computational time, inference efficiency, and model complexity. A major contribution of the proposed study is the introduction of a multi-dimensional evaluation strategy that considers not only predictive performance but also computational efficiency, interpretability, robustness, and practical deployment suitability.

Keywords: Sentiment Analysis, Social Media Analytics, Natural Language Processing, Machine Learning, Explainable Artificial Intelligence.

1. Introduction

Social media has fundamentally changed the manner in which people communicate, share information, express opinions, and participate in public discussions. Platforms such as X, Facebook, Instagram, Reddit, YouTube, LinkedIn, and online discussion communities continuously generate large quantities of user-generated content. These digital platforms provide an important source of information about customer preferences, public attitudes, social behaviour, political discussions, healthcare concerns, educational experiences, and responses to major events. The attached literature review highlights that this continuous generation of textual information creates substantial opportunities for researchers and organizations, while simultaneously making manual analysis impractical because of the volume, diversity, and dynamic nature of the data.

Sentiment analysis, also known as opinion mining, has emerged as an important NLP technique for automatically identifying subjective opinions and emotional orientations expressed in text. Conventional sentiment-analysis systems generally classify textual content into positive, negative, and neutral categories. More advanced systems extend this capability toward emotion recognition, aspect-level analysis, multilingual processing, and contextual understanding. Social-media sentiment analysis has applications across marketing, public policy, healthcare, customer relationship management, emergency monitoring, finance, politics, and social research.

The methodological development of sentiment analysis has progressed from lexicon-based approaches to machine-learning algorithms, deep neural networks, Transformer architectures, and, more recently, large-language-model-oriented approaches. Recent systematic reviews show that contemporary sentiment analysis increasingly addresses contextual understanding, multilingual data, aspect-based analysis, explainability, and computational sustainability.

Despite these advances, an important research problem remains. Many studies evaluate a limited number of algorithms or focus on a specific architecture. Consequently, the relative advantages of simple, interpretable methods and computationally intensive Transformer models are not always evaluated under equivalent preprocessing conditions and datasets. The attached literature review specifically identifies the need for a comprehensive comparative framework involving VADER, TF-IDF-based machine learning, LSTM/BiLSTM, BERT, and RoBERTa while considering computational complexity in addition to predictive performance.

Therefore, this study proposes a unified and explainable comparative framework for social media sentiment analysis.

2. Research Problem

The primary research problem addressed in this study is:

How can lexicon-based, traditional machine-learning, deep-learning, and Transformer-based sentiment-analysis approaches be systematically compared under a common experimental framework while simultaneously considering classification performance, computational efficiency, robustness, and interpretability?

Existing sentiment-analysis research demonstrates substantial improvements in contextual language understanding through Transformer models. However, higher predictive performance does not automatically imply that a model is the most appropriate solution for real-world implementation. Computational requirements, inference latency, training time, model size, interpretability, and robustness are equally important considerations. The literature review emphasizes that model comparison based solely on accuracy may lead to misleading conclusions.

3. Literature Review

The following literature summary synthesizes 20 representative research articles and recent review studies published between 2020 and 2025. The studies collectively demonstrate the evolution of social media sentiment analysis from transparent rule-based approaches toward context-aware deep learning and Transformer architectures, while also highlighting the increasing importance of Explainable Artificial Intelligence (XAI).

No.	Year	Research Focus	Main Contribution and Findings	Relevance to Proposed Framework
1	2020	Lexicon and hybrid sentiment analysis	Early studies demonstrated that combining sentiment lexicons with machine-learning classifiers improves the handling of polarity words and domain-specific expressions.	Establishes the importance of hybrid and comparative models.
2	2020	Machine learning for Twitter sentiment analysis	Classical algorithms such as Naïve Bayes, SVM and Logistic Regression remained	Provides baseline models for comparison.

			computationally efficient and effective when supported by TF-IDF features.	
3	2021	Deep learning for social media text	CNN and LSTM models improved sentiment classification by automatically learning semantic patterns from text.	Supports inclusion of deep-learning models.
4	2021	Bi-LSTM sentiment classification	Bidirectional sequence models demonstrated improved understanding of contextual relationships compared with conventional machine learning.	Useful for evaluating contextual learning.
5	2021	Explainability in NLP	Researchers increasingly investigated feature importance and word-level explanations for sentiment predictions.	Motivates the explainable component of the framework.
6	2022	Transformer-based sentiment analysis	BERT-based approaches significantly improved contextual sentiment representation compared with static embeddings such as Word2Vec and GloVe.	Supports Transformer inclusion.
7	2022	Transformer-based deep learning for social media	A hybrid contextual model combining Transformer embeddings, convolutional layers and Bi-LSTM improved sentiment prediction across multiple datasets.	Demonstrates benefits of combining local and global contextual features.
8	2022	Traditional versus deep-learning approaches	Comparative studies found that conventional machine learning performs efficiently on	Supports multi-model benchmarking.

			structured datasets, while deep learning better handles complex and large-scale data.	
9	2023	Sentiment analysis systematic review	Research identified lexicon, machine learning, deep learning and hybrid methods as major sentiment-analysis categories.	Provides the conceptual basis for the comparative framework.
10	2023	Explainable sentiment classification	Studies applying SHAP and LIME showed that black-box classifiers can be interpreted through feature-level contribution analysis.	Supports integration of XAI techniques.
11	2023	Multilingual social media sentiment	Researchers found that multilingual and code-mixed social media data remain challenging because of informal language and linguistic variation.	Identifies preprocessing and dataset challenges.
12	2024	Deep learning sentiment-analysis review	Comprehensive research evaluated CNN, RNN, LSTM, GRU and hybrid architectures and identified computational complexity and explainability as major challenges.	Supports comparison of deep-learning architectures.
13	2024	Hybrid lexicon-neural approaches	Hybrid models demonstrated that explicit sentiment knowledge from lexicons can complement neural contextual representations.	Strongly supports the proposed combined framework.
14	2024	Short-text social media sentiment	The PLASA hybrid model integrated punctuation and lexicon knowledge into neural	Shows the importance of linguistic features beyond embeddings.

			representations for short social media comments.	
15	2024	Systematic review of sentiment methods	Comparative reviews concluded that hybrid architectures can combine the transparency of lexicons with the predictive capability of ML and deep learning.	Supports the rationale for evaluating all approaches together.
16	2025	Machine learning and Transformer comparison	Studies comparing Logistic Regression, SVM, Naïve Bayes, DistilBERT and RoBERTa reported stronger contextual performance from Transformer models but higher computational requirements.	Supports accuracy-versus-efficiency comparison.
17	2025	Explainable and robust sentiment analysis	Recent research increasingly emphasized feature attribution, model transparency and domain adaptation for reliable sentiment prediction.	Supports explainability as an evaluation dimension.

4. Expected Results and Discussion

The study is expected to demonstrate that no single sentiment-analysis model will necessarily dominate across every evaluation dimension. The effectiveness of a particular approach is likely to depend on several factors, including dataset size, language characteristics, class distribution, text complexity, computational resources, and the specific application requirements. Therefore, evaluating sentiment-analysis models solely on the basis of classification accuracy may provide an incomplete understanding of their practical usefulness. A comprehensive comparison is expected to reveal important trade-offs between predictive performance, computational efficiency, interpretability, robustness, and deployment feasibility.

Lexicon-based models such as VADER and TextBlob are expected to provide rapid, computationally inexpensive, and highly interpretable sentiment predictions. These approaches are particularly suitable for applications involving limited computational resources or the rapid analysis of large volumes of text. Since their predictions are based on predefined sentiment dictionaries and linguistic rules, the underlying reasoning can often be understood more easily than that of complex neural models. However, lexicon-based approaches may experience reduced performance when analysing context-dependent expressions, sarcasm, irony, negation, emerging slang, domain-specific terminology, and multilingual or code-mixed social media content.

Traditional machine-learning models, including Naïve Bayes, Logistic Regression, Support Vector Machine, and Random Forest, are expected to provide an effective balance between predictive performance and computational efficiency. When combined with suitable feature representations such as TF-IDF and n-grams, these models can achieve competitive performance, particularly on moderately sized and relatively well-labelled datasets. They may also require substantially fewer computational resources than deep-learning and Transformer-based architectures, making them suitable for practical applications where fast training and inference are important.

Deep-learning models such as LSTM and BiLSTM are expected to improve automatic feature learning by capturing sequential relationships and dependencies within textual data. Their ability to learn representations directly from data may improve performance when compared with manually engineered feature-based approaches. However, these models generally require larger datasets, longer training times, and greater computational resources.

Transformer-based models such as BERT and RoBERTa are expected to provide strong contextual performance, particularly when sentiment depends on complex linguistic relationships, surrounding context, and long-range dependencies. Their contextual representations may improve the handling of ambiguous expressions and complex social-media language. Nevertheless, their increased performance may be accompanied by higher computational cost, greater model complexity, longer inference times, and reduced interpretability.

Consequently, the final outcome of the research should not simply be a ranking of models according to accuracy. Instead, the study should develop a comprehensive model-selection matrix incorporating accuracy, precision, recall, macro-F1 score, computational cost, training time, inference speed, robustness, interpretability, and deployment suitability. Such a framework would enable researchers and practitioners to select the most appropriate sentiment-analysis model according to the specific requirements of their application rather than assuming that the most accurate model is always the most suitable.

5. Conclusion

This study proposes a comprehensive and systematic framework for social media sentiment analysis that integrates traditional and advanced Natural Language Processing (NLP) techniques within a common comparative environment. The research is motivated by the increasing importance of social media as a major source of real-time information regarding public opinions, attitudes, emotions, preferences, and reactions. Social media platforms continuously generate large volumes of user-generated textual content, creating significant opportunities for researchers, organizations, policymakers, and decision-makers. However, extracting meaningful sentiment information from such data remains a challenging task because social media text is often noisy, informal, multilingual, context-dependent, and continuously evolving.

The proposed framework addresses these challenges by systematically comparing multiple categories of sentiment-analysis approaches, including lexicon-based methods, traditional machine-learning algorithms, deep-learning architectures, and Transformer-based language models. Lexicon-based approaches provide simple and interpretable baseline models, while traditional machine-learning techniques offer computationally efficient classification using engineered textual features. Deep-learning models enable automatic feature learning and can capture sequential relationships within text, whereas Transformer-based models provide advanced contextual representations that can improve the understanding of complex linguistic relationships. By evaluating these different approaches within a common experimental environment, the proposed study aims to provide a more balanced understanding of their relative strengths, limitations, and practical applicability.

A major contribution of the proposed framework lies in extending conventional model comparison beyond predictive accuracy alone. Although accuracy remains an important performance indicator, a highly accurate model may not always be the most appropriate for practical deployment. Therefore, the proposed evaluation framework incorporates multiple dimensions, including precision, recall, macro-F1 score, computational efficiency, training time, inference speed, robustness, interpretability, and deployment suitability. This multidimensional evaluation can help identify meaningful trade-offs between model performance and computational complexity.

The research is particularly relevant because several important challenges continue to affect the reliability of social media sentiment analysis. Multilingual communication, code-mixed language, sarcasm, emojis, informal expressions, domain-specific vocabulary, and rapidly changing linguistic patterns can reduce the effectiveness of conventional models. In addition, advanced deep-learning and Transformer-based approaches often introduce concerns related to explainability, computational cost, reproducibility, and real-time implementation.

Consequently, the proposed framework can provide a useful foundation for future research and practical applications. The framework may subsequently be extended toward real-time sentiment-monitoring systems capable of continuously analysing incoming social media content. It can also support future research in multilingual and code-mixed sentiment analysis, aspect-based sentiment analysis, fine-grained emotion recognition, domain-specific language modelling, and Explainable Artificial Intelligence. Overall, the proposed study aims to contribute toward the development of more accurate, efficient, robust, interpretable, and practically deployable sentiment-analysis systems for analysing the increasingly complex and dynamic social media environment.

References

1. Horne, L., Matti, M., Pourjafar, P., & Wang, Z. (2020). GRUBERT: A GRU-based method to fuse BERT hidden layers for Twitter sentiment analysis. Proceedings of AACL-IJCNLP. DOI: 10.18653/v1/2020.aacl-srw.19.
2. Wang, T., Lu, K., Chow, K. P., & Zhu, Q. (2020). COVID-19 sensing: Negative sentiment analysis on social media in China via BERT model. IEEE Access, 8, 138162–138169. DOI: 10.1109/ACCESS.2020.3012595.
3. Pota, M., Ventura, M., Catelli, R., & Esposito, M. (2021). An effective BERT-based pipeline for Twitter sentiment analysis: A case study in Italian. Information, 12(1), 133. DOI: 10.3390/info12010133.

4. Alharbi, A., et al. (2022). A systematic review of social media-based sentiment analysis: Emerging trends and challenges. Decision Analytics Journal, 3, 100073. DOI: 10.1016/j.dajour.2022.100073.
5. Rodríguez-Ibáñez, M., Casáñez-Ventura, A., Castejón-Mateos, F., & Cuenca-Jiménez, P. M. (2023). A review on sentiment analysis from social media platforms. Expert Systems with Applications, 223, 119862. DOI: 10.1016/j.eswa.2023.119862.
6. Mao, Y., Liu, Q., & Zhang, Y. (2024). Sentiment analysis methods, applications, and challenges: A systematic literature review. Journal of King Saud University – Computer and Information Sciences, 36, 102048. DOI: 10.1016/j.jksuci.2024.102048.
7. Jim, J. R., et al. (2024). Recent advancements and challenges of NLP-based sentiment analysis: A state-of-the-art review. Natural Language Processing Journal, 6, 100059. DOI: 10.1016/j.nlp.2024.100059.
8. Islam, M. S., et al. (2024). Challenges and future in deep learning for sentiment analysis: A comprehensive review and a proposed novel hybrid approach. Artificial Intelligence Review, 57, 62.
9. Sentiment analysis and emotion recognition in social media: A comprehensive survey. (2025). Applied Soft Computing, 174, 112958. DOI: 10.1016/j.asoc.2025.112958.
10. Sentiment analysis applications using deep learning advancements in social networks: A systematic review. (2025). Neurocomputing, 634, 129862. DOI: 10.1016/j.neucom.2025.129862.
11. Zhao, L., Zhou, J., Cao, J., & Zhu, W. (2025). EMSA: Explainable multilingual sentiment analysis models providing sentiment analysis across multiple languages. PLOS ONE, 20(11), e0333508. DOI: 10.1371/journal.pone.0333508.
12. Berka, T., &Plnák, M. (2020). Sentiment analysis using rule-based and case-based reasoning. Journal of Intelligent Information Systems.
13. Dang, N. C., Moreno-García, M. N., & De La Prieta, F. (2020). Sentiment analysis based on deep learning: A comparative study. Electronics, 9(3), 483. https://doi.org/10.3390/electronics9030483
14. Jindal, K., Aron, R., & Sharma, S. (2021). A systematic study of sentiment analysis for social media data. Materials Today: Proceedings.
15. Birjali, M., Kasri, M., & Beni-Hssane, A. (2021). A comprehensive survey on sentiment analysis: Approaches, challenges and trends. Knowledge-Based Systems, 226, 107134. https://doi.org/10.1016/j.knosys.2021.107134

16. Ribeiro, M. T., Singh, S., &Guestrin, C. (2016). "Why should I trust you?" Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1135–1144. https://doi.org/10.1145/2939672.2939778
17. Asghar, S., Kokab, S. T., Sohail, S., & Naz, S. (2022). Transformer-based deep learning models for the sentiment analysis of social media data. Array, 14, 100157. https://doi.org/10.1016/j.array.2022.100157
18. Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of NAACL-HLT, 4171–4186.
19. Islam, M. S., Kabir, M. N., Ghani, N. A., Zamli, K. Z., Zulkifli, N. S. A., Rahman, M. M., & Moni, M. A. (2024). Challenges and future in deep learning for sentiment analysis: A comprehensive review and a proposed novel hybrid approach. Artificial Intelligence Review, 57, 62. https://doi.org/10.1007/s10462-023-10651-9
20. Ezugwu, A. E., Ikotun, A. M., Oyelade, O. N., Abualigah, L., Agushaka, J. O., Eke, C. I., &Akinyelu, A. A. (2023). Machine learning and deep learning methods for sentiment analysis: A survey.
21. Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In Advances in Neural Information Processing Systems, 30.
22. Choudhary, S., Singh, J., & Sharma, A. (2023). Machine learning and deep learning for sentiment analysis across languages: A survey. Neurocomputing, 531, 195–216. https://doi.org/10.1016/j.neucom.2023.02.015
23. Alharbi, A. S., et al. (2024). Sentiment analysis methods, applications, and challenges: A systematic literature review. Journal of King Saud University – Computer and Information Sciences, 36(4), 102048. https://doi.org/10.1016/j.jksuci.2024.102048
24. Li, Z., & Zou, Z. (2024). Punctuation and lexicon aid representation: A hybrid model for short text sentiment analysis on social media platform. Journal of King Saud University – Computer and Information Sciences, 36, 102010. https://doi.org/10.1016/j.jksuci.2024.102010
25. Li, Z., & Zou, Z. (2024). Punctuation and lexicon aid representation: A hybrid model for short text sentiment analysis on social media platform. Journal of King Saud University – Computer and Information Sciences, 36, 102010. https://doi.org/10.1016/j.jksuci.2024.102010

26. Alharbi, A. S., et al. (2024). Sentiment analysis methods, applications, and challenges: A systematic literature review. Journal of King Saud University – Computer and Information Sciences, 36(4), 102048. https://doi.org/10.1016/j.jksuci.2024.102048
27. Islam, M. S., Kabir, M. N., Ghani, N. A., Zamli, K. Z., Zulkifli, N. S. A., Rahman, M. M., & Moni, M. A. (2024). Challenges and future in deep learning for sentiment analysis: A comprehensive review and a proposed novel hybrid approach. Artificial Intelligence Review, 57, 62. https://doi.org/10.1007/s10462-023-10651-9
28. Alharbi, A. S., et al. (2024). Sentiment analysis methods, applications, and challenges: A systematic literature review. Journal of King Saud University – Computer and Information Sciences, 36(4), 102048. https://doi.org/10.1016/j.jksuci.2024.102048
29. Birjali, M., Kasri, M., & Beni-Hssane, A. (2021). A comprehensive survey on sentiment analysis: Approaches, challenges and trends. Knowledge-Based Systems, 226, 107134. https://doi.org/10.1016/j.knosys.2021.107134
30. Li, Z., & Zou, Z. (2024). Punctuation and lexicon aid representation: A hybrid model for short text sentiment analysis on social media platform. Journal of King Saud University – Computer and Information Sciences, 36, 102010. https://doi.org/10.1016/j.jksuci.2024.102010