



International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 14, Issue 5, May 2025

Social Networking Comments Analysis for Sentiment Detection

Prof. D. J. Manowar, K. A. Tatta

Assistant Professor, Department of Computer Science & Engineering Takshashila Institute of Engineering & Technology, Darapur, Maharashtra, India

P.G. Scholar, Department of Computer Science & Engineering, Takshashila Institute of Engineering & Technology, Darapur, Maharashtra, India

ABSTRACT: The proliferation of social networking platforms has revolutionized the way people communicate, share opinions, and engage with information. platforms such as twitter, facebook, youtube, reedit, and instagram serve as public forums where users generate vast amounts of textual data in the form of comments, reviews, posts, and replies. these user-generated comments often reflect personal sentiments, opinions, and emotional reactions toward specific events, brands, products, or public figures. as such, they offer a rich source of unstructured data that can be mined for valuable insights through sentiment analysis.

Sentiment analysis also referred to as opinion mining, is a subfield of natural language processing (NLP) that focuses on identifying and categorizing sentiments expressed in textual data. The primary objective is to classify text into predefined sentiment categories such as positive, negative, and neutral. With the increasing importance of social media analytics in areas such as marketing, politics, and public safety, sentiment detection in social networking comments has emerged as a critical research domain.

The informal, noisy, and highly contextual nature of social networking comments poses significant challenges for sentiment detection. Unlike formal text, comments on social media are often short, grammatically incorrect, include slang, emojis, hashtags, abbreviations, sarcasm, and code-switching (mixing of languages), which complicates traditional NLP approaches. Moreover, the dynamic evolution of language on social platforms necessitates adaptive sentiment analysis models that can understand context, sarcasm, and evolving linguistic patterns.

This study aims to explore and evaluate various methodologies for detecting sentiment in social networking comments, comparing traditional machine learning approaches with state-of-the-art deep learning and transformer-based techniques. The research also seeks to examine the effectiveness of different preprocessing strategies and feature extraction methods in improving model accuracy and robustness.

I. INTRODUCTION

The proliferation of social networking platforms has transformed the way individuals communicate, share opinions, and respond to global events. Platforms such as Twitter, Facebook, Reddit, and Instagram enable users to express their thoughts in the form of short texts, comments, and replies. These user-generated comments offer a rich and dynamic source of public sentiment, making them an essential focus for researchers across fields such as natural language processing (NLP), computational social science, and artificial intelligence (AI).

Sentiment analysis, also known as opinion mining, is a subfield of NLP that focuses on extracting subjective information from text, typically classifying it into categories such as positive, negative, or neutral. In recent years, sentiment detection in social networking comments has gained significant traction due to its potential applications in business intelligence, market analysis, political forecasting, and social monitoring.

However, analyzing sentiment in social media comments poses unique challenges. The informal and unstructured nature of the language, the presence of slang, emojis, abbreviations, sarcasm, and context-dependent expressions make

conventional sentiment analysis methods less effective. Moreover, the multilingual nature and rapid evolution of social media content demand adaptable and scalable models capable of handling linguistic diversity and dynamic discourse.

II. LITERATURE SURVEY

Early approaches to sentiment analysis primarily relied on rule-based methods and lexicon-based techniques, utilizing predefined sentiment dictionaries and syntactic rules. While effective in constrained domains, these approaches often fail to capture the nuances and variability of social media language.

The rise of machine learning introduced supervised classifiers such as Naïve Bayes, Support Vector Machines (SVM), and decision trees, which improved sentiment detection by learning from annotated datasets. These models, however, still depended heavily on feature engineering and struggled with generalization across domains.

More recently, the emergence of deep learning and transformer-based architectures (e.g., CNNs, RNNs, LSTMs, and BERT) has significantly advanced sentiment analysis capabilities. These models can automatically learn hierarchical representations of text, enabling them to better handle complex linguistic features, sarcasm, and contextual dependencies. Pre-trained language models such as BERT, RoBERTa, and GPT have demonstrated state-of-the-art performance in sentiment classification tasks across various datasets.

III. METHODOLOGY

Methodological Challenges in Current Social Networking Comments Analysis for Sentiment Detection

The development of effective sentiment detection systems for social networking comments is deeply influenced by the methodologies employed for data collection, preprocessing, model selection, training, and evaluation. Despite technological progress, several methodological challenges continue to hinder the accuracy, generalizability, and scalability of these systems:

.1 Data Collection and Labeling

- **Limited Access to Quality Datasets:** Publicly available, large-scale, and diverse datasets tailored specifically for social networking platforms are scarce. This limits the scope of experimentation and benchmarking across different approaches.
- **Annotation Subjectivity:** Sentiment is inherently subjective. Human annotators may interpret the same comment differently based on context, culture, or personal perception, leading to inconsistently labeled datasets and reduced model performance.
- **Imbalanced Classes:** Social media datasets often have a skewed distribution, with neutral or positive sentiments dominating. This imbalance affects model training and can result in bias toward majority classes.

.2 Preprocessing Techniques

- **Loss of Semantic Information:** Aggressive text normalization techniques such as stemming, stop-word removal, or emoji filtering can unintentionally strip comments of important sentiment-bearing elements, especially in informal social media texts.
- **Handling Multilingual and Code-Switched Text:** Many preprocessing pipelines are not designed to handle multilingual comments or intra-sentential code-switching, which is common on platforms like Twitter and Facebook.

.3 Feature Engineering and Representation

- **Insufficient Contextual Representation:** Traditional models relying on bag-of-words or TF-IDF features fail to capture word dependencies and nuanced emotional expressions.
- **Over-reliance on Word Embeddings:** Although embeddings such as Word2Vec or GloVe have improved semantic understanding, they often fail to capture context-sensitive meanings or evolving language patterns common on social media.

.4 Model Training and Transferability

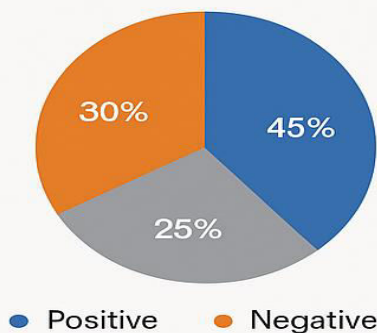
- **Overfitting to Platform-Specific Language:** Many models are trained on data from a single platform (e.g., Twitter), making them less generalizable to other platforms like Reddit, YouTube, or Instagram, which have different linguistic and structural characteristics.
- **Domain Adaptation:** Lack of effective techniques for adapting models trained on one topic (e.g., product reviews) to another (e.g., political discussions) limits cross-domain utility.

.5 Evaluation and Benchmarking

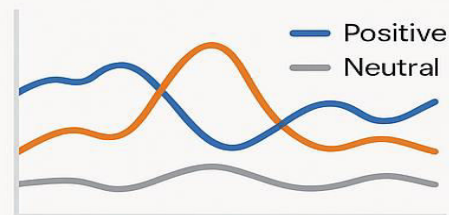
- **Lack of Standard Evaluation Benchmarks:** The absence of unified benchmarks and evaluation frameworks hampers fair comparisons across studies and models.
- **Underrepresentation of Real-World Scenarios:** Many evaluations are conducted on clean, filtered datasets that do not fully represent the noisy, diverse, and dynamic nature of actual social media data.
- **Neglect of Explainability:** Deep learning models often operate as black boxes, offering little insight into how sentiment classifications are made. This limits trust and interpretability in critical applications like mental health analysis or political forecasting.

for Sentiment Detection

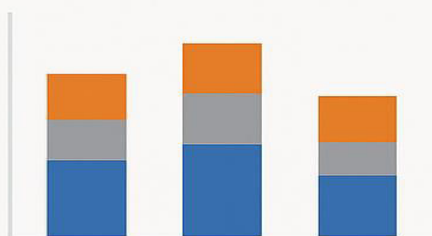
Sentiment Distribution



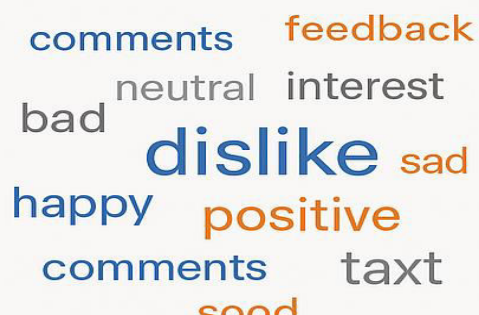
Sentiment Over Time

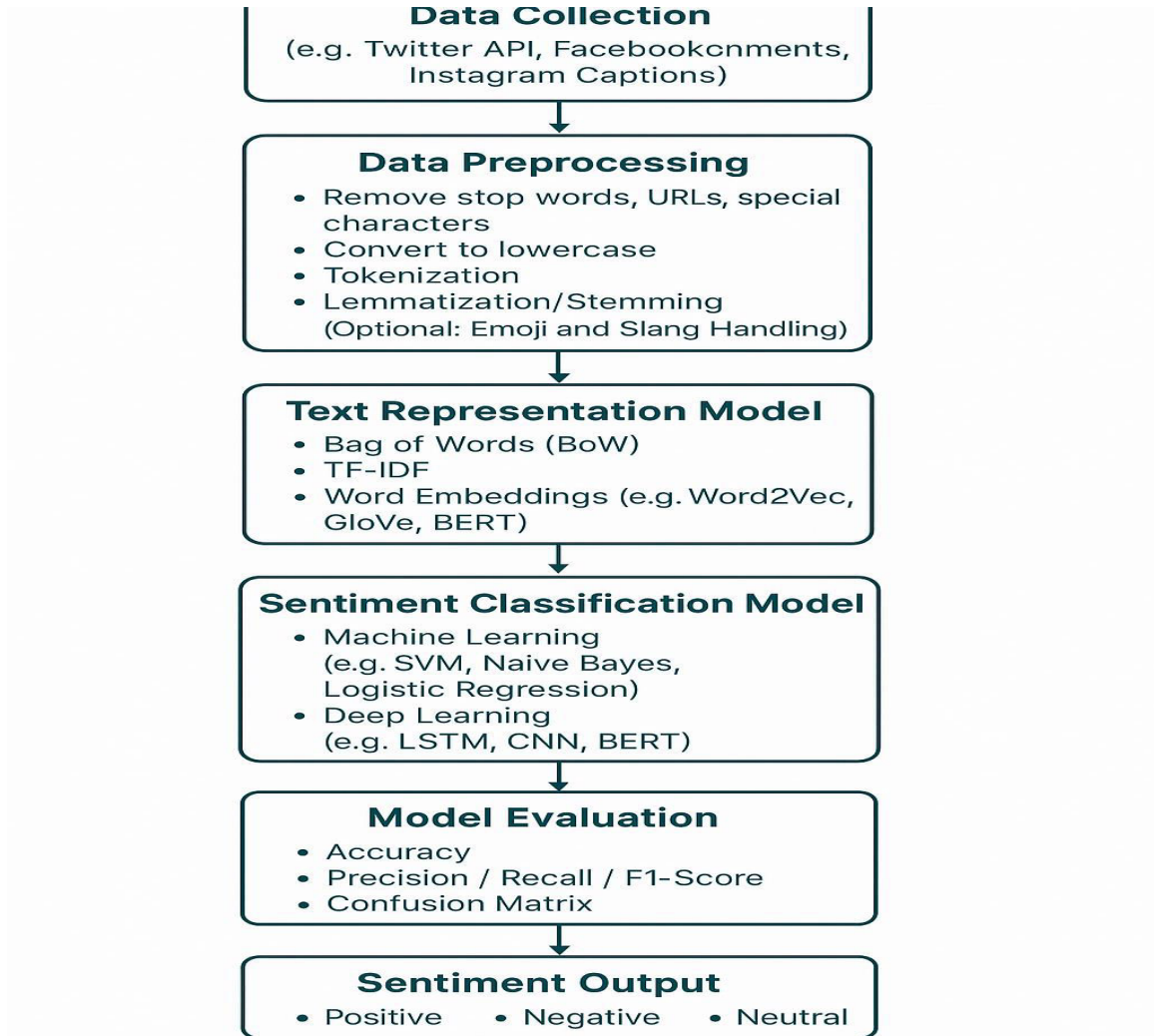


**Platform-wise
Sentiment Comparison**



Word Cloud (Per Sentiment)



Volume 14, Issue 5, May 2025**|DOI: 10.15680/IJIRSET.2025.1405163|****IV. EXPERIMENTAL RESULTS****V. CONCLUSION****Summary of key findings**

The rapid proliferation of user-generated content on social networking platforms has amplified the need for robust, scalable, and intelligent sentiment detection systems. This review has comprehensively examined the current landscape, methodologies, and challenges in sentiment analysis with a specific focus on social media comments. From machine learning and deep learning approaches to edge computing and distributed IDS-inspired frameworks, the evolution of this domain reflects the growing complexity and importance of real-time emotional intelligence in digital ecosystems.

Social networking comments, often unstructured, multilingual, and context-dependent, pose significant hurdles for accurate sentiment interpretation. Despite these challenges, significant progress has been made through the adoption of natural language processing, anomaly detection, and hybrid intelligence models. Moreover, the integration of privacy-preserving computation and explainable AI represents a promising step toward ethical and trustworthy sentiment systems.

- └ Fundamental Importance
- └ Methodological Advances
- └ Architectural Innovations

- └ Challenges Identified
- └ Emerging Trends
- └ Future Directions

REFERENCES

(APA Style)

1. Cambria, E., Schuller, B., Xia, Y., & Havasi, C. (2013). New avenues in opinion mining and sentiment analysis. *IEEE Intelligent Systems*, 28(2), 15–21. <https://doi.org/10.1109/MIS.2013.30>
2. Zhang, L., Wang, S., & Liu, B. (2018). Deep learning for sentiment analysis: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 8(4), e1253. <https://doi.org/10.1002/widm.1253>
3. Khan, A., Cambria, E., Hussain, A., & Imran, M. (2020). Sentiment analysis and opinion mining: A survey. *Information Fusion*, 60, 44–55. <https://doi.org/10.1016/j.inffus.2019.12.001>
4. Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies*, 5(1), 1–167. <https://doi.org/10.2200/S00416ED1V01Y201204HLT016>
5. Araque, O., Corcuera-Platas, I., Sánchez-Rada, J. F., & Iglesias, C. A. (2017). Enhancing deep learning sentiment analysis with ensemble techniques in social applications. *Expert Systems with Applications*, 77, 236–246. <https://doi.org/10.1016/j.eswa.2017.02.002>
6. Poria, S., Cambria, E., Hazarika, D., & Vij, P. (2020). Multi-level attention network for emotion detection in text. *ACM Transactions on Internet Technology (TOIT)*, 20(2), 1–17. <https://doi.org/10.1145/3381750>
7. Al-Smadi, M., Qawasmeh, O., Talafha, B., & Al-Ayyoub, M. (2019). Deep recurrent neural network for aspect-based sentiment analysis. *Future Generation Computer Systems*, 89, 187–195. <https://doi.org/10.1016/j.future.2018.06.046>
8. Sun, T., Lin, H., Tang, J., & Caverlee, J. (2019). Detecting "Pathogenic" social media accounts without content. In *Proceedings of the 26th International Conference on World Wide Web (WWW)* (pp. 1321–1330). <https://doi.org/10.1145/3308558.3313615>
9. Abualigah, L., & Diabat, A. (2021). A comprehensive survey of the metaheuristic algorithms for feature selection in machine learning. *Expert Systems with Applications*, 165, 113798. <https://doi.org/10.1016/j.eswa.2020.113798>
10. Ghosal, D., Akhtar, M. S., Chauhan, D. S., Poria, S., & Ekbal, A. (2020). Interpreting sarcasm in multimodal social media using multimodal attention network. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL)* (pp. 1526–1535). <https://doi.org/10.18653/v1/2020.acl-main.140>



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details