JRITM
JOURNAL OF RESEARCH AND INNOVATION IN TECHNOLOGY, COMMERCE AND MANAGEMENT

Journal of Research and Innovation in Technology, Commerce and Management
Vol. 3 Issue 6, June 2026, pp. 36039-36044
ISSN: 3049-3129(Online)

Sentiment Analysis of Social Media Data Using Machine Learning

Shital Lumade
ECT Dept. MGM'S College of Engineering, Nanded, Maharashtra
Email : shitallumade@gmail.com

Durga Kadam
ECT Dept. MGM'S College of Engineering, Nanded, Maharashtra
Email: kadamdurga2004@gmail.com

Dr. Mrs. K. P. Paithane
ECT Dept. MGM'S College of Engineering, Nanded, Maharashtra
Email: Jondhale_kc@mgmcen.ac.in

Ms. A. K. Salve
ECT Dept. MGM'S College of Engineering, Nanded, Maharashtra
Email: Salve_ak@mgmcen.ac.in

Cite as: Shital Lumade, Durga Kadam, Dr. Mrs. K. P. Paithane, & Ms. A. K. Salve. (2026). Sentiment Analysis of Social Media Data Using Machine Learning. Journal of Research and Innovation in Technology, Commerce and Management, Vol. 3(Issue 6), 36039–36044. https://doi.org/10.5281/zenodo.20794488

DOI : https://doi.org/10.5281/zenodo.20794488

ABSTRACT - The growth of social media platforms has created a vast amount of user-generated text data. This data provides important insights into public opinions, feelings, and attitudes. Sentiment analysis uses machine learning techniques to automatically classify and understand these sentiments. This is useful in areas like marketing, politics, public health, and crisis management. This review explores recent advances in machine learning-based sentiment analysis on social media, covering key algorithms, datasets, preprocessing methods, challenges, and future research areas. It highlights the evolving role of deep learning models and hybrid approaches to tackle the complexities of informal and noisy social media text.

Keywords: Sentiment Analysis, Social Media Analytics, Machine Learning, Opinion Mining, Natural Language Processing (NLP), Big Data, Artificial Intelligence (AI), Real-Time Sentiment Detection.

INTRODUCTION

Social media platforms like Twitter, Facebook, Instagram, and Reddit are vital spaces for people to express their opinions, feelings, and feedback on various topics. The enormous and rapidly increasing amounts of textual data generated daily on these platforms offer valuable resources for analyzing public sentiment. Sentiment analysis, or opinion mining, examines the opinions, sentiments, attitudes, and emotions conveyed in text' As big data grows and artificial intelligence advances, machine learning has become a key technology for automating sentiment analysis. Unlike traditional rule-based or lexicon-based methods, machine learning algorithms can learn intricate language patterns and subtle contextual hints from data, improving accuracy and scalability. The ability to assess social media sentiment in real time is particularly advantageous for marketing, politics, disaster response, and mental health monitoring. This paper reviews the recent progress of machine learning techniques in sentiment analysis of social media data.

LITERATURE REVIEW

Rule-based and lexicon-based approaches that relied on sentiment dictionaries were used in the early stages of sentiment analysis research; however, these were not domain- or context-adaptable (Putra & Wijaya, 2023). By identifying patterns in textual features like TF-IDF and n-grams, traditional machine learning models like Support Vector Machines (SVM), Naive Bayes (NB), and Random Forest (RF) increased classification accuracy (Rifa'i et al., 2023). More accurate sentiment classification is now possible thanks to deep learning advancements that have allowed models such as Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) to capture semantic and sequential dependencies in text. By offering contextual embeddings that perform better than previous models, transformer architectures—specifically BERT and its variations, RoBERTa and DistilBERT—revolutionized the field (Vaswani et al., 2017; Gunasekaran, 2023).

Current issues like sarcasm detection, code-mixing, and data bias continue to be important research topics, but recent studies show the increasing use of hybrid and multimodal models that integrate textual, visual, and audio data to improve sentiment detection.

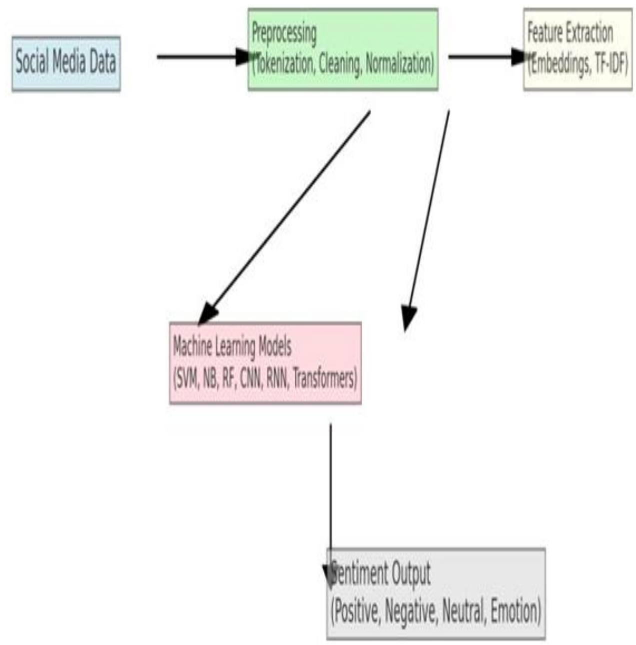

Fig.1: Sentiment Analysis Pipeline

Fig.1 Shows the pipeline demonstrates the workflow of sentiment analysis, starting from data collection to final sentiment classification.

MACHINE LEARNING TECHNIQUES FOR SENTIMENT ANALYSIS

Traditional ML models use handcrafted features, while deep learning and hybrid approaches learn patterns directly from text for more accurate sentiment analysis on social media.

a. Traditional Machine Learning Algorithms

Initially, sentiment analysis relied on classical machine learning classifiers that used manually created features. These include: Support Vector Machines (SVM): SVMs aim to find the hyperplane that best separates classes, such as positive and negative sentiment, by maximizing the distance between data points. They perform well in high-dimensional feature spaces, making them suitable for text

Journal of Research and Innovation in Technology, Commerce and Management
Vol. 3 Issue 6, June 2026, pp. 36039-36044
ISSN: 3049-3129(Online)

classification. For instance, Rifa'i et al. (2023) employed SVM to classify tweets during the 2023 Southeast Asian (SEA) Games football final, demonstrating that SVM can effectively handle large social media data streams with good accuracy. Naive Bayes (NB): This probabilistic classifier applies Bayes' theorem, assuming that features are independent. Despite its simplicity, NB performs well in many sentiment classification tasks, especially with smaller datasets. Random Forest (RF): This ensemble method uses multiple decision trees to reduce overfitting and enhance generalization. It helps identify key features that influence sentiment classification. These traditional models generally depend on textual features like bag-of-words, n-grams, Term Frequency-Inverse Document Frequency (TF-IDF), and sentiment lexicons to convert text into a numerical format suitable for learning.

b. Deep Learning Models

The rise of deep learning transformed sentiment analysis by allowing models to learn hierarchical features directly from raw text without extensive feature engineering. Convolutional Neural Networks (CNNs): Originally designed for image processing, CNNs have been adapted for text by treating word embeddings as spatial features. CNNs can effectively capture local patterns like phrases or sentiment-indicating keywords through convolution and pooling operations. Recurrent Neural Networks (RNNs): RNNs and their variants—Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU)—are meant to model sequential data. They can remember earlier words in a sentence, which helps capture context and manage long-range dependencies crucial for understanding sentiment. Transformer Models: The transformer architecture introduced by Vaswani et al. (2017) includes self-attention mechanisms, allowing models to understand the significance of different words regardless of their position in the sentence. BERT (Bidirectional Encoder Representations from Transformers) and its variants like RoBERTa and DistilBERT have set new standards in sentiment classification by offering deep contextual embeddings. Gunasekaran (2023) noted that transformers outperform earlier models, particularly in complex sentiment tasks that involve sarcasm and mixed emotions.

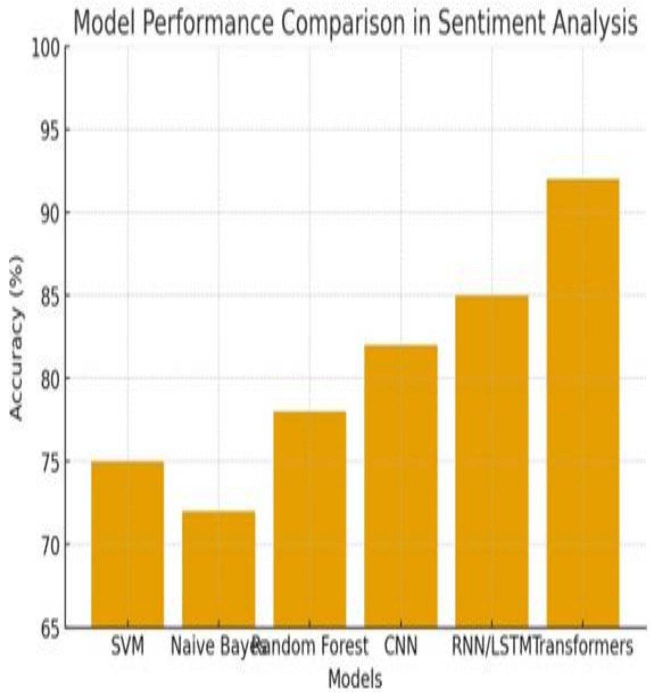

Fig .2 graph shows how well different models perform in understanding sentiment (like positive or negative feelings).

c. Hybrid and Ensemble Models

To leverage the strengths of different approaches, hybrid models that merge lexicon-based features with machine learning classifiers or combine multiple ML models are becoming more common. For instance, using ensembles of SVM and deep learning models can improve resilience by tackling individual weaknesses. Furthermore, multimodal sentiment analysis, which merges text with images or audio, is emerging as a significant area for deeper

Journal of Research and Innovation in Technology, Commerce and Management
Vol. 3 Issue 6, June 2026, pp. 36039-36044
ISSN: 3049-3129(Online)

sentiment detection.

DATASETS AND PREPROCESSING TECHNIQUES

a. Commonly Used Datasets

SemEval datasets: These popular benchmark datasets consist of annotated tweets for sentiment classification, emotion detection, and irony.

Sentiment140: This dataset contains 1.6 million tweets labeled as positive or negative, using emoticons for distant supervision.
Twitter US Airline Sentiment dataset: This focuses on customer feedback about airlines, with positive, negative, and neutral labels. Domain-specific datasets: Product reviews from Amazon, political tweets, and disaster-related posts provide various contexts for training models.

Table 1: Commonly Used Datasets for Sentiment Analysis

b.

Dataset	Domain	Size	Labels
SemEval	Twitter (general)	Variable (thousands)	Positive, Negative, Neutral, Irony, Emotion
Sentiment140	Twitter	1.6M tweets	Positive, Negative
Twitter US Airline	Airline reviews	14K tweets	Positive, Neutral, Negative
Amazon Reviews	Product reviews	Millions	1-5 star ratings
Political/Disaster Datasets	Events, crisis tweets	Variable	Positive, Negative, Neutral

Preprocessing

Social media text is often noisy and informal, requiring careful preprocessing steps like:
Tokenization: Breaking the text into tokens or words.
Stop-word removal: Eliminating common words that add little sentiment (e.g., "the", "is").
Normalization: Converting text to lowercase, clarifying slang, abbreviations, and elongated words (e.g., “soooo”).
Emoji and Hashtag Handling: Emojis have strong emotional weight and can be mapped to sentiment scores. Hashtags can provide context.
Noise Removal: Removing URLs, mentions (@user), and special characters that do not contribute to sentiment.
Effective preprocessing helps models learn valuable features from social media text.

CHALLENGES IN MACHINE LEARNING-BASED SENTIMENT ANALYSIS

Despite advancements, several ongoing challenges hinder the effectiveness of ML models in social media sentiment analysis:
Sarcasm and Irony Detection: Sarcasm can reverse the literal meaning of words, leading to confusion for models that focus on superficial features. Recognizing sarcasm requires understanding context and tone, along with some world knowledge.

Multilingual and Code-Mixed Text: Social media users often mix languages in one post (code-switching), which complicates matters for models trained on single-language datasets.Class Imbalance: Sentiment datasets generally favor certain classes (e.g., more positive than negative), leading to biased models.Short and Informal Text: Tweets and posts are often brief and contain slang, misspellings, and informal grammar, making language modeling difficult. Data Privacy and Ethics: Collecting and analyzing user-generated content raises privacy concerns. Moreover, bias in training data can cause unfair or discriminatory outcomes.

APPLICATIONS

Machine learning-based sentiment analysis has revolutionized various fields by offering valuable insights from social media:

Marketing and Brand Monitoring: Companies use sentiment analysis to track customer opinions, manage brand reputation, and adjust advertising strategies.

Political Campaigns and Public Opinion: Governments and political groups analyze voter sentiment to gauge support, identify issues, and plan communication strategies.

Public Health Surveillance: Mental health professionals monitor social media for signs of depression, anxiety, or suicidal thoughts to provide timely support.

CrisisManagement and Disaster Response: Real-time sentiment monitoring during emergencies helps authorities understand public mood and needs.

FUTURE DIRECTIONS

a.Multimodal Sentiment Analysis

Future systems will combine text, images, video, and audio data from social media posts for richer and more accurate sentiment insights. Using multiple data types can clarify sentiment when text alone is inadequate.

b .Explainable AI (XAI)

Creating interpretable machine learning models is crucial for building trust and accountability. Techniques like attention visualization, feature importance, and rule extraction can help explain model decisions to users and stakeholders.

c. Real-Time Sentiment Analytics

Developing scalable architectures that process social media streams in real-time will allow for immediate insights for crisis response and marketing.

d. Bias and Fairness

Research focusing on detecting and minimizing biases in training data and model predictions will help ensure fair and ethical sentiment analysis systems.

e. Enhanced Sarcasm and Irony Detection

Incorporating world knowledge, pragmatics, and user behavior analysis can improve the identification of sarcastic and ironic statements.

CONCLUSION

Machine learning has significantly improved sentiment analysis on social media by enabling automated, scalable, and increasingly accurate interpretations of user emotions. From traditional classifiers to deep learning and transformer models, ML techniques continue evolving to address the complexities of informal, noisy, and multilingual social media text. Despite ongoing challenges like sarcasm detection and ethical issues, future developments promise more robust, multimodal, and interpretable sentiment analysis systems, unlocking new opportunities across various industries and social sectors.

REFERENCES

1. Rifa'i, M., et al. (2023). Sentiment classification of social media posts during the SEA Games football final using SVM. Journal of Artificial Intelligence and Engineering Applications, 9(1), 23-30.

2, Gunasekaran, S. (2023). Deep learning for sentiment analysis: A comprehensive survey. arXiv preprint arXiv:2305.14842.

3. Putra, I. M., & Wijaya, A. (2023). Lexicon-based sentiment analysis of public opinions on AI in Twitter data. Journal of Urban and Technological Studies, 14(2), 1-15.

4. Vaswani, A., et al. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30.

5. 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.

6. Sun, C., Huang, L., & Qiu, X. (2019*).* Utilizing BERT for aspect-based sentiment analysis via constructing auxiliary sentence. NAACL-HLT 2019, 380–385.

7. .Luo, Z., et al. (2021). Multimodal sentiment analysis using cross-modal attention and fusion networks. Information Fusion, 68, 1–10.

8. Zhang, L., Wang, S., & Liu, B. (2018). Deep Learning for Sentiment Analysis: A Survey