Social Media Sentiment Analysis Using Machine Learning: Frameworks, Methods, and Challenges

Sanu Kumar
B.Tech Student, Global Institute of Technology, Jaipur, Rajasthan, India
23egjcs803@gitjaipur.com

Sara Kanwar
B.Tech Student, Global Institute of Technology, Jaipur, Rajasthan, India
23egjcs804@gitjaipur.com

Sara Kumawat
B.Tech Student, Global Institute of Technology, Jaipur, Rajasthan, India
23egjcs805@gitjaipur.com

Laxmikant Vashishtha
Assistant Professor, Global Institute of Technology, Jaipur, Rajasthan, India
laxmikant.vashishtha@gitjaipur.com

Yoganand Sharma
Assistant Professor, Global Institute of Technology, Jaipur, Rajasthan, India
yoganand.sharma@gitjaipur.com

ABSTRACT: The exponential growth of social media platforms has led to the generation of enormous volumes of user-generated textual data, reflecting public opinions, emotions, and behavioral trends across diverse domains. Analyzing this data effectively is crucial for applications such as brand monitoring, political analysis, public health surveillance, and financial forecasting. Sentiment analysis, also referred to as opinion mining, focuses on automatically identifying and classifying sentiments expressed in text as positive, negative, or neutral. In recent years, Machine Learning (ML) techniques have emerged as a dominant approach for social media sentiment analysis due to their capability to learn complex patterns from large-scale and unstructured datasets. This paper presents a concise review of machine learning–based sentiment analysis frameworks for social media data. It outlines the key components of the framework, discusses commonly used machine learning algorithms and evaluation metrics, highlights major challenges, and identifies potential future research directions in this evolving field.

KEYWORDS: Sentiment Analysis, Social Media, Machine Learning, Natural Language Processing, Opinion Mining.

1. INTRODUCTION

The widespread adoption of social media platforms such as Twitter (X), Facebook, Instagram, and Reddit has transformed the way people communicate, share information, and express opinions. These platforms generate vast amounts of real-time textual data, offering valuable insights into public sentiment on topics ranging from consumer products and services to politics, healthcare, and global events. However, the sheer volume, velocity, and diversity of social media data make manual analysis impractical and inefficient.

Sentiment analysis has emerged as an essential technique for extracting subjective information from textual data by identifying users' emotions, attitudes, and opinions. Also

known as opinion mining, sentiment analysis enables automated classification of text into sentiment categories such as positive, negative, or neutral. This capability plays a critical role in decision-making processes for businesses, governments, and researchers seeking to understand public perception and behavioral trends.

Early sentiment analysis approaches primarily relied on rule-based and lexicon-based methods, which use predefined sentiment dictionaries and linguistic rules. Although these methods are simple and interpretable, they suffer from several limitations when applied to social media data. Informal language, spelling variations, slang, abbreviations, emojis, sarcasm, and context-dependent meanings significantly reduce their effectiveness and accuracy.

Machine learning–based approaches address these challenges by learning patterns directly from labeled datasets rather than relying on manually crafted rules. By leveraging algorithms such as Naïve Bayes, Support Vector Machines, Logistic Regression, and ensemble methods, machine learning models can effectively capture complex linguistic structures and contextual information. These approaches have demonstrated superior performance and scalability, making them the preferred choice for social media sentiment analysis. Consequently, machine learning has become a foundational component in the development of robust and efficient sentiment analysis frameworks.

2. MACHINE LEARNING BASED SENTIMENT ANALYSIS FRAMEWORK

Data science integrates data collection, preprocessing, statistical modeling, and machine learning to analyze complex healthcare datasets. Techniques such as regression analysis, classification, clustering, and deep learning are used to identify disease patterns, predict health risks, and optimize resource allocation. Advanced analytics help clinicians make evidence-based decisions while reducing diagnostic errA typical machine learning–based social media sentiment analysis framework consists of the following stages:

A. Data Collection

Social media data is collected using platform APIs or web scraping techniques. The data may include posts, tweets, comments, hashtags, emojis, and metadata such as timestamps and user information.

B. Data Preprocessing

Preprocessing is a crucial step due to noisy and unstructured text. Common techniques include:

- Tokenization
- Removal of stop words, URLs, and special characters
- Stemming or lemmatization
- Handling emojis, slang, and abbreviations

C. Feature Extraction

Textual data is transformed into numerical representations using methods such as:

- Bag of Words (BoW)
- Term Frequency–Inverse Document Frequency (TF-IDF)
- Word embeddings (Word2Vec, GloVe)

D. Machine Learning Classification

Various supervised ML algorithms are used to classify sentiments, including:

- Naïve Bayes (NB)
- Support Vector Machines (SVM)
- Logistic Regression (LR)
- Decision Trees and Random Forests

These models are trained on labeled datasets to predict sentiment polarity.

E. Model Evaluation

Performance is evaluated using metrics such as accuracy, precision, recall, F1-score, and confusion matrix.

3. APPLICATIONS OF SOCIAL MEDIA SENTIMENT ANALYSIS

Machine learning based sentiment analysis is widely used in:

- Brand and product reputation monitoring
- Political opinion analysis and election forecasting
- Customer feedback analysis
- Stock market and financial sentiment prediction
- Public health and crisis monitoring.

4. CHALLENGES

Despite its effectiveness, several challenges remain:

- Handling sarcasm and irony
- Multilingual and code-mixed text
- Imbalanced datasets
- Rapidly evolving social media language
- Privacy and ethical concerns

5. CONCLUSION

This paper reviewed machine learning–based frameworks for social media sentiment analysis, highlighting their structure, commonly used algorithms, and applications. Machine

learning techniques have significantly improved sentiment classification accuracy compared to traditional methods. However, addressing linguistic complexity and ethical challenges remains essential for building robust and reliable sentiment analysis systems.

REFERENCES

[1] P. Upadhyay, K. K. Sharma, R. Dwivedi and P. Jha, "A Statistical Machine Learning Approach to Optimize Workload in Cloud Data Centre," 2023 7th International Conference on Computing Methodologies and Communication (ICCMC), pp. 276-280, 2023.

[2] P. Jha, K. K. Sharma, B. Jain, V. Sharma, " Encryption Using AES Algorithm", EIJO Journal of Engineering, Technology And Innovative Research (EIJO–JETIR), Vol. 4, Issue. 2, 2019.

[3] M. K. Sain and N. Sharma, "A study of research issues and challenges of big data analytics," Journal of Advances and Scholarly Researches in Allied Education, vol. 16, no. 5, pp. 1699–1707, 2019.

[4] N. Sharma, "An analytical study of distributed data store using big data analysis technique," Research Methods, Imparc, 2019.

[5] K. Kanhaiya, A. K. Sharma, K. Gautam, P. S. Rathore, "AI Enabled-Information Retrival Engine (AI-IRE) in Legal Services: An Expert-Annotated NLP for Legal Judgements", 2023 Second International Conference on Augmented Intelligence and Sustainable Systems (ICAISS), 2023.

[6] M. K. Jha, "Recent Trends and Emerging Applications of the Internet of Things: Transforming the Way We Live and Work", International Journal of Engineering Trends and Applications (IJETA), Vol. 12, Issue. 4, pp. 239-244, 2025.

[7] A. Sharma and K. Gautam, "Flood prediction using machine learning technique," 2nd International Conference on Pervasive Computing Advances and Applications (PerCAA 2024), pp. 319-327, 2024.

[8] R. Joshi, M. Farhan, U. Sharma, S. Bhatt, "Unlocking Human Communication: A Journey through Natural Language Processing", International Journal of Engineering Trends and Applications (IJETA), Vol. 11, Issue. 3, pp. 245-250, 2024.

[9] R. Joshi, A. Maritammanavar, "Deep Learning Architectures and Applications: A Comprehensive Survey", International Conference on Recent Trends in Engineering & Technology (ICRTET 2023), pp. 1-5, 2023.

[10] S. A. Saiyed, N. Sharma, H. Kaushik, P. Jain, G. K. Soni and R. Joshi, "Transforming portfolio management with AI and ML: shaping investor perceptions and the future of the Indian investment sector," Parul University International Conference on Engineering and Technology 2025 (PiCET 2025), pp. 1108-1114, 2025.

[11] P. Jha, T. Biswas, U. Sagar and K. Ahuja, "Prediction with ML paradigm in Healthcare System," 2021 Second International Conference on Electronics and Sustainable Communication Systems (ICESC), pp. 1334-1342, 2021.

[12] H. Arora, M. Kumar, T. Rasool and P. Panchal, "Facial and Emotional Identification using Artificial Intelligence," 2022 6th International Conference on Trends in Electronics and Informatics (ICOEI), pp. 1025-1030, 2022.

[13] S. Thapar, G. K. Soni, H. Kaushik, R. Singh, S. Bisht and S. K. Bansal, "A Comparative Machine Learning Framework for Detecting Fake Accounts on Facebook," 2025 4th International Conference on Innovative Mechanisms for Industry Applications (ICIMIA), pp. 1567-1571, 2025.

[14] A. Maheshwari, R. Ajmera and D. K. Dharamdasani, "Unmasking Embedded Text: A Deep Dive into Scene Image Analysis," 2023 International Conference on Advances in Computation, Communication and Information Technology (ICAICCIT), pp. 1403-1408, 2023.

[15] H. Sharma and R. Ajmera, "Comprehensive review and analysis on machine learning based Twitter opinion mining framework," Tuijin Jishu/Journal of Propulsion Technology, vol. 44, no. 5, 2023.

[16] I. Yadav, V. Shekhawat, K. Gautam, G. Kumar Soni and R. Yadav, "Artificial Intelligence for Cybersecurity: Emerging Techniques, Challenges, and Future Trends," 2025 3rd International Conference on Sustainable Computing and Data Communication Systems (ICSCDS), pp. 1176-1180, 2025.

[17] M. Kumar, R. Ajmera, and D. Kumar, "Statistical analysis and accuracy assessment of improved machine learning based opinion mining framework," Advances in Nonlinear Variational Inequalities, vol. 27, no. 1, 2024.

[18] V. Joshi, S. Patel, R. Agarwal and H. Arora, "Sentiments Analysis using Machine Learning Algorithms," IEEE 2023 Second International Conference on Electronics and Renewable Systems (ICEARS), pp. 1425-1429, 2023.