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 5, Issue 3, November 2025
Impact Factor: 7.67

Sentiment Analysis on Social Media using Machine Learning

Mr. Gavhad Abhijeet[1], Mr. Kamble Rushikesh[2], Prof. Patade J. M.[3]
Student, AI&DS Department[1-2]
Professor, AI&DS Department[3]
Jaihind College of Engineering Kuran, Pune, India

Abstract: *In today's digital world, social media has become a space where millions of people share their thoughts, emotions, and opinions every day. Understanding these sentiments can provide meaningful insights for businesses, organizations, and researchers. This project, titled "Sentiment Analysis of Social Media Using Machine Learning," aims to automatically detect and categorize the emotions expressed in user posts as positive, negative, or neutral. The system gathers data from popular social media platforms and processes it through several steps such as cleaning, tokenization, and removal of unnecessary words. Key text features are then extracted using techniques like TF-IDF and word embeddings, which help the model understand the context of each post. Machine learning algorithms like Naïve Bayes, Logistic Regression, and Support Vector Machine (SVM) are used to train and predict the sentiment of new data. The project evaluates its accuracy using performance measures such as precision, recall, and F1-score. The outcome helps in understanding public opinion trends, brand perception, and user behavior more effectively, showing how machine learning and NLP can make sense of large volumes of online data.*

Keywords: Sentiment Analysis, Machine Learning, Natural Language Processing (NLP), Social Media, Text Classification, Opinion Mining, LSTM, BERT

I. INTRODUCTION

In recent years, social media platforms such as Twitter, Facebook, Instagram, and Reddit have become powerful channels for communication, self-expression, and information exchange. Millions of users share their opinions, emotions, and experiences daily, generating vast amounts of unstructured text data. Analyzing this data can provide valuable insights into public attitudes toward events, brands, products, and social issues. However, the massive scale and informal nature of social media content—often filled with slang, abbreviations, and sarcasm—make manual analysis nearly impossible. Sentiment analysis, also known as opinion mining, is a branch of Natural Language Processing (NLP) and Machine Learning (ML) that aims to automatically identify and interpret people's emotions, opinions, and attitudes expressed in text. It categorizes text into different sentiment classes such as positive, negative, or neutral. By leveraging supervised and unsupervised machine learning algorithms, sentiment analysis enables the extraction of meaningful patterns and trends from large datasets. With the rise of artificial intelligence and data-driven decision-making, sentiment analysis has found applications in numerous fields such as marketing, politics, customer feedback analysis, brand reputation monitoring, and public opinion tracking. Organizations use these insights to understand user satisfaction, predict market trends, and make informed business decisions. This review focuses on how machine learning models and NLP techniques are applied to perform sentiment analysis on social media data. It explores existing research, compares traditional machine learning methods with deep learning approaches, and highlights the challenges and opportunities in developing accurate, real-time sentiment analysis systems. The goal is to understand how these methods contribute to effective text classification and how they can be improved for better social media insight generation.

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 5, Issue 3, November 2025
Impact Factor: 7.67

II. PROBLEMSTATEMENT

With the rapid increase in social media usage, a massive amount of user-generated text data is produced every day. However, analyzing and understanding the sentiments expressed in this unstructured data remains a major challenge. The project aims to explore how machine learning techniques can accurately identify and classify sentiments from social media content.

III. OBJECTIVES

- To study various machine learning techniques used for sentiment analysis on social media data.
- To compare the performance of traditional and deep learning models in sentiment classification.
- To understand the role of Natural Language Processing (NLP) in preprocessing and feature extraction.
- To identify the challenges involved in analyzing unstructured social media text data.
- To review existing research and tools used for sentiment detection and opinion mining.
- To evaluate different datasets commonly used for training and testing sentiment analysis models.
- To explore real-world applications of sentiment analysis in business, politics, and social media monitoring.
- To suggest improvements or future directions for building more accurate and efficient sentiment analysis systems.

IV. LITERATURE REVIEW

Sentiment analysis has become an important area of research in Natural Language Processing (NLP) and Machine Learning (ML), as it helps in understanding public opinions and emotions expressed on social media platforms. Over the past few years, several researchers have contributed to improving sentiment classification methods using advanced algorithms and deep learning models.[1]

Early studies mainly focused on traditional machine learning techniques such as Naïve Bayes, Support Vector Machine (SVM), and Logistic Regression. These models relied heavily on Bag-of-Words (BoW) and TF-IDF approaches for text feature representation. Although effective for small datasets, these models often struggled with understanding the contextual meaning of words in long or complex sentences.[2]

To overcome these limitations, recent works have shifted towards deep learning approaches such as Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM), and Convolutional Neural Networks (CNN). These models can capture sequential and contextual information, improving the accuracy of sentiment detection. For instance, researchers have shown that LSTM networks perform significantly better on datasets containing informal social media text due to their ability to handle long-term dependencies in language.[3]

Another breakthrough in sentiment analysis came with the introduction of transformer-based models, such as BERT (Bidirectional Encoder Representations from Transformers). BERT's ability to understand context in both directions of a sentence has made it one of the most powerful models for text classification tasks. Studies have reported that BERT and its variants outperform traditional ML models in classifying tweets, comments, and reviews with higher precision and recall rates.[4]

In addition to model development, researchers have explored various data preprocessing and feature extraction techniques to improve system performance. Methods such as stop-word removal, stemming, lemmatization, and the use of word embeddings (Word2Vec, GloVe) have enhanced the ability of models to interpret language patterns effectively. Some studies also focused on handling challenges like sarcasm, emojis, and multilingual text, which are common on social media.[5]

Furthermore, sentiment analysis has been applied in multiple domains — including brand monitoring, product feedback analysis, election prediction, and disaster management. This diversity highlights the growing importance of sentiment analysis as a decision-support tool for businesses and governments.[6]

Overall, the literature shows a steady evolution from simple statistical models to more advanced deep learning architectures, resulting in improved accuracy and better contextual understanding. However, challenges such as data imbalance, noise in social media text, and computational cost still remain areas of active research. Future studies are expected to focus on hybrid models, multilingual sentiment detection, and real-time analysis systems to further enhance performance and applicability.[7]

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 5, Issue 3, November 2025
Impact Factor: 7.67

V. METHODOLOGY

The methodology for this study focuses on understanding and analyzing the various techniques, models, and processes involved in performing sentiment analysis on social media data using machine learning. The review follows a systematic approach that includes data collection, preprocessing, feature extraction, model selection, and performance evaluation.

1. Data Collection :Social media data such as tweets, posts, and comments are collected from platforms like Twitter, Facebook, Instagram, and Reddit using APIs or web scraping tools. This data typically includes user-generated content containing opinions, hashtags, and emoticons. Publicly available sentiment datasets such as Twitter Sentiment140, IMDb Reviews, or Amazon Product Reviews are also used for comparative analysis.

2. Data Preprocessing : The collected data often contains noise in the form of URLs, emojis, abbreviations, and irrelevant symbols. To ensure data quality, several text preprocessing steps are performed, including tokenization, stop-word removal, stemming, lemmatization, and normalization. This process helps convert unstructured text into a clean and standardized format suitable for model training.

3. Feature Extraction : After preprocessing, meaningful textual features are extracted to represent the data numerically. Common techniques include Bag-of-Words (BoW), Term Frequency–Inverse Document Frequency (TF-IDF), and Word Embeddings such as Word2Vec, GloVe, or FastText. These methods help the model understand the semantic relationships between words and sentences.

4. Model Development and Training : The review covers both traditional machine learning models and deep learning architectures used for sentiment classification. Algorithms such as Naïve Bayes, Logistic Regression, Decision Tree, and Support Vector Machine (SVM) are compared with advanced models like Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM), and Bidirectional Encoder Representations from Transformers (BERT). Deep learning models are particularly effective in capturing contextual meaning and long-term dependencies in text.

5. Performance Evaluation : The performance of different models is assessed using evaluation metrics such as Accuracy, Precision, Recall, and F1-score. Cross-validation techniques are applied to ensure reliability, and confusion matrices are used to visualize the results. Studies also analyze the computational cost and scalability of different approaches when applied to large-scale social media datasets.

6. Comparative Analysis and Discussion : The results of various research papers are compared to identify the most effective models and preprocessing techniques. The review discusses how hybrid and ensemble models combining multiple classifiers can improve performance. Additionally, it examines challenges such as sarcasm detection, multilingual data handling, and data imbalance, which continue to affect sentiment analysis accuracy.

VI. CONCLUSION

Sentiment analysis using machine learning provides an effective way to understand public opinions and emotions shared on social media. The review shows how techniques have evolved from traditional models like Naïve Bayes and SVM to advanced deep learning methods such as LSTM and BERT, improving accuracy and context understanding. Although progress has been made, challenges like sarcasm, noisy text, and multilingual data still affect performance. Future research should focus on hybrid and real-time models that can handle large-scale data efficiently, making sentiment analysis more reliable for real-world applications.

REFERENCES

[1]. Hamid, A. I., & Abdulazeez, A. M. (2024). Sentiment Analysis Based on Machine Learning Techniques: A Comprehensive Review. Indonesian Journal of Computer Science, 13(3), 4080-4092.

[2]. Tan, K. L., Lee, C. P., & Lim, K. M. (2023). A Survey of Sentiment Analysis: Approaches, Datasets, and Future Research. Applied Sciences, 13(7), 4550.

[3]. Singh, N., & Jaiswal, U. C. (2023). A Detailed Sentiment Analysis Survey Based on Machine Learning Techniques. ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal, 12(1), e29105.

[4]. Gunasekaran, K. P. (2023). Exploring Sentiment Analysis Techniques in Natural Language Processing: A Comprehensive Review. arXiv.

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 5, Issue 3, November 2025 Impact Factor: 7.67

[5] .Kumar, K. P., Baswaraj, D., Parvathi Devi, G., Thanugundala, S., Changala, R., & G. V. R. (2024). Sentiment Analysis in Social Media Using Deep Learning Techniques. International Journal of Intelligent Systems and Applications in Engineering, 12(3), 1588-1597.
[6] ."Challenges and future in deep learning for sentiment analysis: a comprehensive review and a proposed novel hybrid approach." (2024). Artificial Intelligence Review, 57, 62.
[7] .Ranjan, M., Tiwari, S., Sattar, A. M., &Tatkar, N. S. (2023). A New Approach for Carrying Out Sentiment Analysis of Social Media Comments Using Natural Language Processing. Engineering Proceedings, 59(1), 181.
[8] .Radha, G., & Chandrashekhar, K. (2025). Sentiment Analysis on Social Media Opinions: A Survey of Machine Learning and Lexicon-Based Approaches. Journal of Neonatal Surgery, 14(6S), 24-29.
[9] .Dev, C., & Ganguly, A. (2023). Sentiment Analysis of Review Data: A Deep Learning Approach Using User-Generated Content. Asian Journal of Electrical Sciences, 12(2), 28-36.
[10] ."A systematic review of social network sentiment analysis with comparative study of ensemble-based techniques." (2023). Artificial Intelligence Review, 56, 13407-13461.
[11] .Suryawanshi, N. S. (2024). Sentiment analysis with machine learning and deep learning: A survey of techniques and applications. International Journal of Science and Research Archive, 12(02), 005-015.
[12] .Dhole, A. M., & Sahu, N. (2023). Sentiment Analysis Of Social-Media Using Machine Learning Approach. International Journal of Creative Research Thoughts, 11(4), –.
[13] .Mrs. Rashmi T., Prof. Veena M. N. (2023). Sentiment Analysis of Social Media Data: Perspectives and Avenues. JETIR, 10(4).
[14] .Narayan, R., & Samanta, P. (2024). A Machine Learning Approach for Sentiment Analysis Using Social Media Posts. International Journal of Information Technology and Computer Science, 16(5), 23-35.
[15] .Hussain, H. H., Bangare, S., Nirmal, A., &Bangal, S. (2023). Sentiment Analysis: A Comprehensive Review of Techniques and Applications. JETIR, 10(8).