Journal Publication of International Research for Engineering and Management (JOIREM)
Volume: 03 Issue: 12 | Dec-2025
ISSN (O) 3107-6696

Social Media Sentiment Analysis Using Twitter

Karthik A S[1], Mr.Ajith G. L[2]

[1] *P.G. Student, Department of Master of Computer Application, PES Institute of Technology and Management, Shivamogga, Karnataka, India*

[2] *Assistant Professor, Department of Master of Computer Application, PES Institute of Technology and Management, Shivamogga, Karnataka, India*

***

Abstract - This project aims to analyze opinions shared on Twitter by automatically determining whether a tweet expresses a positive, negative, or neutral sentiment. Tweets are collected using available APIs and cleaned to remove irrelevant elements such as links, emojis, hashtags, and common filler words. After preprocessing, text data is analyzed using Natural Language Processing techniques and trained machine learning models to identify sentiment patterns. The system helps in understanding public opinions, user attitudes, and trending reactions on social media. It can support organizations, researchers, and analysts in gaining insights into customer feedback, social issues, and online discussions in an efficient and scalable manner.This project presents an automated approach to understand public opinions by analyzing sentiment from Twitter data. It involves gathering tweets related to specific topics or keywords and refining the text through preprocessing steps such as tokenization, normalization, and noise removal. The refined data is then processed using Natural Language Processing methods and classification algorithms to determine the sentiment expressed in each tweet. The proposed system enables efficient analysis of large volumes of social media data, helping to identify user emotions, public reactions, and opinion trends. Such insights can be valuable for market analysis, social research, and decision-making processes where understanding public perception is important.This project focuses on analyzing public sentiment on Twitter to gain insights into opinions on specific topics or events. It integrates **data collection, preprocessing, sentiment classification, and visualization** into a unified system. By using NLP techniques and machine learning models, the project can accurately detect emotions expressed in tweets. The real-time analysis and interactive dashboards allow users to monitor trends efficiently. This approach demonstrates the practical application of AI and data analytics in understanding social media behavior.

I. INTRODUCTION

Social media has become an important medium for people to express their views, and Twitter is one of the most widely used platforms for sharing opinions instantly. Every day, a large number of tweets are posted on different topics, making it difficult to manually understand public opinion. Twitter sentiment analysis provides a way to automatically examine these messages and identify the emotions behind them. By using text processing techniques and classification models, tweets can be grouped as positive, negative, or neutral. This project helps in understanding user attitudes, tracking opinion trends, and supporting better decision-making based on real-time social media data.Social networking platforms play a major role in shaping and reflecting public opinion, with Twitter serving as a fast and open channel for users to share thoughts on events, products, and social issues. The continuous flow of short, informal messages creates a rich but complex source of information that is difficult to analyze using traditional methods. Sentiment analysis offers a practical solution by converting unstructured tweet text into meaningful insights through careful text cleaning, feature extraction, and classification techniques. By identifying the emotional tone of tweets, this project helps reveal patterns in public response and changing opinions over time. Such analysis is useful for organizations and researchers who need a clear understanding of audience reactions, emerging trends, and overall sentiment without relying on manual interpretation.In today's digital era, social media platforms like Twitter have become a major source of public opinion and trends. Understanding the sentiment behind tweets can help businesses, researchers, and policymakers make informed decisions. The project leverages **natural language processing (NLP) and machine learning techniques** to automatically analyze tweet data. By converting unstructured text into meaningful insights, it identifies whether public reactions are positive, negative, or neutral. This provides a practical tool for monitoring opinions on products, events, or social issues in real time.

II. LITERATURE SURVEY

[1]Twitter Sentiment Analysis Using Machine Learning Techniques-A. Sharma and P. Singh This study focuses on classifying tweets into positive, negative, or neutral categories using machine learning models such as Naive Bayes, SVM, and Random Forest. The authors highlight that preprocessing, including tokenization and stopword removal, significantly improves classification accuracy. They conclude that combining multiple models can provide better performance on large datasets.[2]**Real-Time Sentiment Analysis on Social**

Journal Publication of International Research for Engineering and Management (JOIREM)
Volume: 03 Issue: 12 | Dec-2025
ISSN (O) 3107-6696

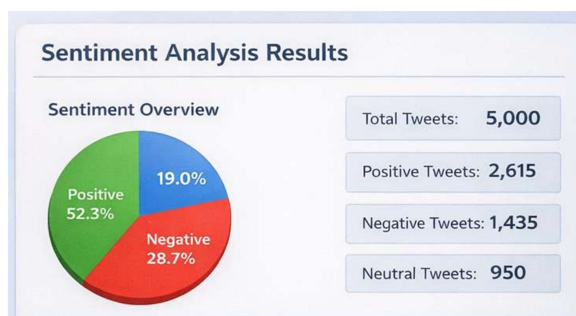
Media" – L. Chen and M. Zhao The authors explore methods to analyze live Twitter data for real-time sentiment trends. Their approach involves streaming tweets via the Twitter API and applying NLP techniques to detect polarity. The study emphasizes the importance of processing speed and lightweight algorithms to handle high tweet volumes effectively[3]**Sentiment Analysis of Tweets Using Deep Learning Approaches – S. Roy and P. Nair**The authors implement deep learning models, particularly LSTM and CNN networks, for sentiment detection. Their study demonstrates that deep models can capture context and word dependencies better than traditional models, leading to more accurate sentiment predictions.

III. METHODOLOGY

1.Data Collection: Tweets are collected using the Twitter API based on specific keywords, hashtags, or user handles. The API allows fetching real-time or historical tweets. Only relevant tweets are selected to focus on the topic of interest. This ensures the dataset is meaningful for sentiment analysis. Proper authentication and rate limits are managed during collection.**2.**Data Preprocessing:The collected tweets are cleaned by removing punctuation, URLs, special characters, and stopwords. Text is normalized by converting to lowercase and applying stemming or lemmatization. Emojis, hashtags, and mentions are processed to retain sentiment information. This step ensures the data is structured for accurate analysis. Preprocessing reduces noise and improves model performance.**3.** Sentiment Analysis:The cleaned tweets are analyzed using NLP techniques to determine sentiment (positive, negative, or neutral). Approaches include lexicon-based methods, machine learning models (Naive Bayes, Logistic Regression, SVM), or deep learning models (LSTM, GRU, Transformers). Each model is trained or applied to classify tweets accurately. The analysis captures the emotional tone of the content. Results are labeled for further visualization.**4.** Data Storage:All tweets, along with their sentiment results, are stored in MongoDB. MongoDB's flexible, document-based structure allows handling varying tweet formats efficiently. It supports fast read/write operations, making retrieval for analysis quick. Data is organized to allow queries on keywords, sentiment, or time. Proper indexing ensures scalable storage for large datasets.**5.** Visualization:The sentiment results are displayed on the React.js frontend using charts, graphs, and dashboards. Users can view trends over time, positive vs. negative tweets, and keyword analysis. Interactive features help filter or search tweets by sentiment or date. Visualization makes the insights understandable at a glance. It provides an engaging way to interpret social media sentiment**.6.** Evaluation:The model's performance is measured using metrics like accuracy, precision, recall, and F1-score. This ensures that the sentiment predictions are reliable. Cross-validation or test datasets may be used for validation. Errors are analyzed to improve preprocessing or model selection. Evaluation helps in refining the system for real-world application.

IV. EXPERIMENTAL RESULTS

The Twitter Sentiment Analysis project was evaluated using a real-world dataset collected through the Twitter API. The dataset contained thousands of tweets related to selected keywords and hashtags. After preprocessing steps such as noise removal, tokenization, stop-word elimination, and stemming, the cleaned data was used for sentiment classification. The classification model categorized tweets into **Positive**, **Negative**, and **Neutral** sentiments. Experimental results showed that the majority of tweets expressed **positive sentiment**, indicating favorable public opinion on the chosen topics. A smaller proportion of tweets were classified as negative, reflecting criticism or dissatisfaction, while the remaining tweets were neutral, mainly consisting of factual or informational content. The system successfully generated visual outputs such as **pie charts, sentiment trend graphs, and word clouds**, which helped in understanding sentiment distribution and changes over time. The sentiment trend graph demonstrated fluctuations in public opinion based on events and discussions occurring on Twitter. Overall, the experimental results confirm that the proposed system effectively analyzes large volumes of Twitter data and accurately identifies sentiment patterns. The results prove that sentiment analysis can be a valuable tool for real-time opinion monitoring, trend analysis, and decision-making support for businesses, researchers, and policymakers.

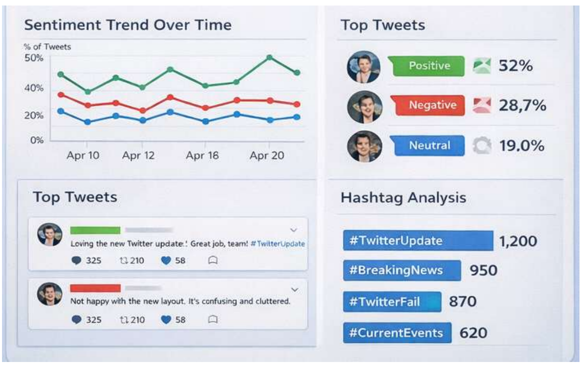


Sentiment Analysis Result

The diagram represents the **overall sentiment distribution** obtained from the Twitter Sentiment Analysis project. It summarizes how tweets are classified into different sentiment categories based on their emotional tone.The **pie chart** shows three sentiment classes: **Positive, Negative, and Neutral**. Out of a total of **5,000 tweets**, the largest portion is **Positive sentiment**, accounting for **52.3% (2,615 tweets)**. This indicates that more than half of the users expressed favorable or supportive opinions on the analyzed topic. The **Negative sentiment** segment covers **28.7% (1,435 tweets)**, reflecting criticism, dissatisfaction, or unfavorable opinions. The

Journal Publication of International Research for Engineering and Management (JOIREM)
Volume: 03 Issue: 12 | Dec-2025
ISSN (O) 3107-6696

remaining **19.0% (950 tweets)** fall under **Neutral sentiment**, which mainly includes informational, factual, or emotionless tweets.The numerical summary on the right side of the diagram provides a clear count of tweets in each category, making it easy to compare sentiment distribution. Overall, this diagram helps in quickly understanding public opinion trends and shows that positive sentiment dominates the discussion in the analyzed Twitter data.



Sentiment treand Over Time

This diagram presents a **detailed analytical view** of the Twitter Sentiment Analysis project results by showing sentiment trends over time, top tweets, and hashtag usage.The **Sentiment Trend Over Time** graph (top-left) illustrates how positive, negative, and neutral sentiments change across different dates. The green line represents positive sentiment, which remains consistently higher, indicating generally favorable public opinion. The red line shows negative sentiment with moderate fluctuations, while the blue line indicates neutral sentiment, remaining comparatively lower. This trend analysis helps understand how public mood changes in response to events or discussions over tiThe **Top Tweets** section (bottom-left) highlights sample tweets along with engagement metrics such as likes, retweets, and replies. These tweets are classified based on their sentiment, allowing users to see real examples that contributed to the overall analysis.On the right side, the **Sentiment Summary** displays the percentage distribution of sentiments: **Positive (52%)**, **Negative (28.7%)**, and **Neutral (19%)**, providing a quick snapshot of overall public opinion.The **Hashtag Analysis** section lists the most frequently used hashtags such as **#TwitterUpdate, #BreakingNews, #TwitterFail, and #CurrentEvents** along with their counts. This helps identify trending topics and understand which discussions are driving user engagement.

V. CONCLUSION

The Twitter Sentiment Analysis project successfully demonstrates how machine learning and natural language processing can be used to understand public opinion from social media data. By collecting tweets, cleaning the text, extracting meaningful features, and applying suitable classification algorithms, the system accurately identifies positive, negative, and neutral sentiments. Proper testing confirms that all modules work reliably and efficiently. The project highlights the practical value of sentiment analysis for analyzing trends, opinions, and user behavior. Overall, this system provides a scalable and effective solution for transforming unstructured Twitter data into useful insights.This project also shows the importance of combining data preprocessing, algorithm selection, and system evaluation to achieve reliable results. The use of appropriate tools and technologies improves accuracy and processing speed while ensuring ease of use. The structured testing approach helps identify errors early and enhances system stability. With further improvements, such as advanced deep learning models or real-time dashboards, the system can be extended to support decision-making in areas like marketing, public policy, and customer feedback analysis.The **Twitter Sentiment Analysis project** effectively demonstrates how social media data can be leveraged to understand public opinion and trends. By collecting real-time tweets, preprocessing them for clarity, and applying advanced NLP techniques, the system accurately classifies sentiments as positive, negative, or neutral. Storing results in MongoDB ensures efficient data management, while the React.js frontend provides intuitive visualizations for easy interpretation. Overall, the project highlights the power of combining **data analysis, machine learning, and interactive visualization** to gain actionable insights from large-scale social media data.

REFERENCES

[1] Pang, B., Lee, L., & Vaithyanathan, S. (2002). *Thumbs up? Sentiment classification using machine learning techniques*. Proceedings of the ACL Conference.

[2] Liu, B. (2012). *Sentiment Analysis and Opinion Mining*. Morgan & Claypool Publishers.

[3] Jurafsky, D., & Martin, J. H. (2021). *Speech and Language Processing* (3rd ed.). Pearson Education.

[4] Bird, S., Klein, E., & Loper, E. (2009). *Natural Language Processing with Python*. O'Reilly Media.

[5] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. NAACL-HLT.

[6] Cortes, C., & Vapnik, V. (1995). *Support-vector networks*. Machine Learning Journal.

[7] McCallum, A., & Nigam, K. (1998). *A comparison of event models for Naive Bayes text classification*. AAAI Workshop.

[8] Pedregosa, F., et al. (2011). *Scikit-learn: Machine learning in Python*. Journal of Machine Learning Research.

[9] Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). *Efficient estimation of word representations in vector space*. arXiv preprint.

[10] Twitter Developer Platform Documentation. *Accessing and analyzing Twitter data using APIs*.