

Twitter Sentiment Analysis Using Machine Learning

Piu Gosh¹, Ritusree Naskar², Dr. Radha Krishna Jana³

Department of Computer Science & Engineering, JIS University, Kolkata, West Bengal 700109, India^{1,2,3}

Email address: piu718627@gmail.com(Piu Gosh),ritusreenaskar04@gmail.com(Ritusree Naskar),
radhakrishnajana.jana@jisuniversity.ac.in (Radha Krishna Jana)

*Corresponding Author: radhakrishna.jana@jisuniversity.ac.in

Abstract— The people of our Society are busy in social media.The post their comments on different topics.We analysed them using different model.This worked applied machine learning model for sentiment analysis on tweets. The propped model shown good result. LSTM gives the 91% accuracy.

Keywords— Sentiment Analysis, Natural Language Processing, Machine learning Model

1. INTRODUCTION

Twitter sentiment analysis is the process of machine learning.It identify and classify opinions, emotion and attitudes (like – positive,negative,neutral) expressed in tweets. When applied to Twitter, it analyzes tweets to understand how people feel about a topic, brand, event, or public figure. We are putting different comments ,opinion on social media. We make different network.Same minded people are make a network called homophily [1].we analysed the opinion, tweets and sentiment.There are several models for analysed the sentiments [2].In this worked we applied genetic algorithms for analysis the sentiments [3].In this worked we used hybrid model for analysis the tweets related to the COVID-19 pandemic[4].The study discussed about modified Firely algorithm and application on sentiment analysis [5].This study proposed a hybrid model using XAI.This model shows the better confidence than other related model[6].

2. RELATED WORK

With the explosion of social media platforms, Twitter has become one of the most prominent sources of real-time data reflecting public opinions on various topics. The microblogging nature of Twitter, with its 280-character limit, allows users to express views on news, products, politics, entertainment, and more creating a valuable corpus for sentiment analysis.Traditional algorithms such as Naive Bayes, Support Vector Machines (SVM), and Logistic Regression were widely used initially. Later, deep learning models such as CNN, LSTM, and recently transformer-based models like BERT have been introduced for improved performance.Table 1. shown the summary of previous study.

Table 1: Summary of Previous Study

Publis hed Work	Nam e of The Mod el	Strength	Weekness	Finding
Ref. 7	LST M	Captures sequential dependencies , effective with time series data.	Prone to vanishing gradient, requires more training time.	Achieved consistent performance across datasets,robust across different domains.
Ref. 8	Naiv e Baye s	Fast, interpretable, performs well on small data.	Assumes feature independence , weak on complex language.	Performed moderately, less accurate than SVM or logistic regression.
Ref 9.	SV M	Robust to overfitting, effective in high dimensional space.	Sensitive to parameter tuning,slow with large data.	Outperformed Naive Bayes and Decision Tree in sentiment classification.
Ref. 10	Deci sion Tree	Very interpretable, handles mixed features types,useful for explainability abd rule extraction.	Prone to overfitting,un steable with small changes to data,typically lower accuracy than ensembles.	Useful for explainability and quick prototyping.

Ref. 11	Random Forest	Robust, reduces overfitting vs single trees, handles high dimensional forest.	Slower to train, less interpretable than single trees.	It often outperforms single trees on social media text.
Ref. 12	K-Nearest Neighbor	Simple, non-parametric, good when dense features represent text well.	Prediction slow on large data, sensitive to scaling, irrelevant features.	It can work well when using dense embeddings.

Table 2. showed the previous experimental worked.

Table 2. Previous Experimental Work

Published Work	Name of The Model	Dataset Details	Result
Ref. 8	Naive Bayes	Kaggle Twitter Sentiment Dataset	Fast and simple, works well on small text data. Struggles with sarcasm or negation. 80-83 % accuracy.
Ref. 9.	Support Vector Machine (SVM)	Twitter Sentiment 140	Excellent for high dimensional text, robust accuracy, slower for huge datasets. 87-89% accuracy.
Ref. 10	Decision Tree	Twitter Sentiment 140	Simple and interpretable, but tends to overfit, not ideal for large datasets. 75-79% accuracy.
Ref. 11	Random Forest	Amazon Reviews	More robust than Decision Tree handles noise well, training time is higher. 83-85% accuracy.
Ref. 12	K-Nearest Neighbors	Twitter Sentiment, Movie Reviews	Easy to implement but slow for large datasets, accuracy depends on data scaling. 78-81% accuracy.

This review discussed on university students physiological parameters[13]. Financial sentiment analysis discussed here[14]. This work used BERT, NB and LR model for sentiment analysis[15]. A recommend system proposed in this work[16]. This works developed a software system for drug-related tweets analysis[17]. This study discussed on the hotel reviews [18]. In this work we analysed the health related comments. Patients can put their comments after receiving any health related service[19]. This study presents an AI-powered medication recommendation system[20].

3. METHODOLOGY

A. Working Flowchart

Figure 1. shown the working flowchart.

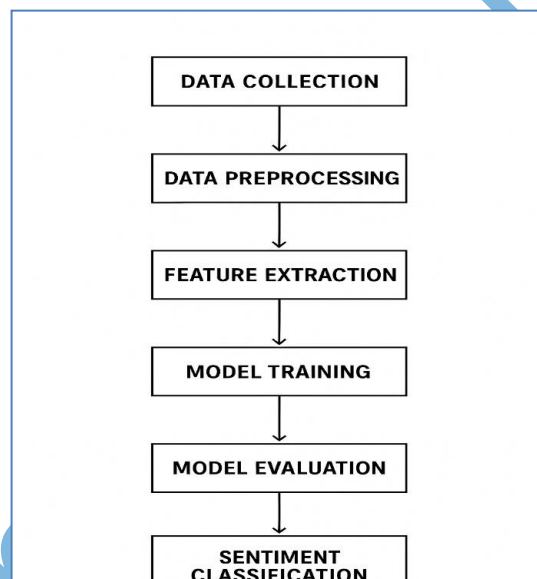


Figure 1: Working Flowchart

B. Data Collection and Preprocessing

Source: The dataset is collected from Kaggle, which provides cleaned and labeled Twitter sentiment datasets such as Sentiment140 and Twitter US Airline Sentiment.

Data Composition: Each record contains the tweet text, sentiment label, and metadata .

Size: The dataset typically consists of 1.6 million tweets (Sentiment140).

Format: CSV format with columns such as:

tweet_id , sentiment (0 = negative, 1 = neutral, 2 = positive), tweet_text.

Figure 2. showed the outflow of machine learning framework.

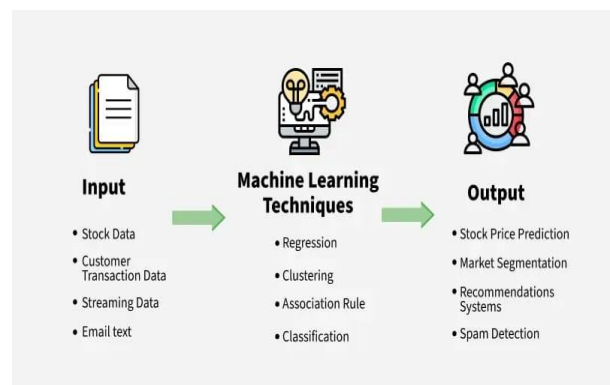


Figure 2. Machine learning workflow

4. RESULTS AND DISCUSSION

Figure 3. demonstrate the comparison with each machine learning models.

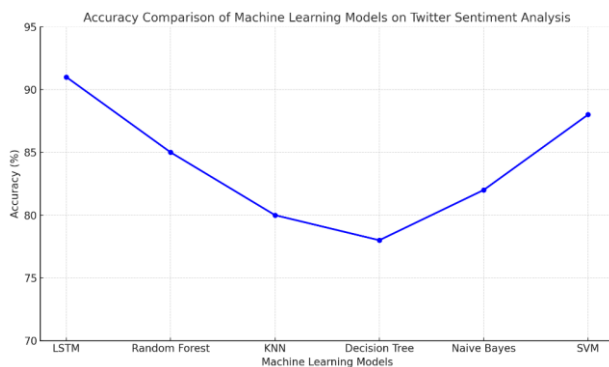


Figure 3. Accuracy Comparison

Table 3. showed the result of different models.

Table 3. Comparison Table

Name of the Model	Accuracy
LSTM (Long Short-Term Memory)	91%
SVM (Support Vector Machine)	88%
Random Forest	85%
Naive Bayes	82%
K-Nearest Neighbors (KNN)	80%
Decision Tree	78%

Sentiment analysis on Twitter data faces many difficulties because of how people write tweets and how machine learning models work.

- Messy and informal tweets
Tweets often include emojis, hashtags, abbreviations, spelling mistakes, and slang. This makes it hard for the computer to understand the true meaning of the text.
- Short length of tweets

Tweets are very short (only up to 280 characters). Models don't get enough context or information to understand the user's feeling correctly.

- Sarcasm and irony

People often say the opposite of what they mean (for example, "Great, my phone just died!"). Models get confused and give wrong sentiment labels.

- Unbalanced data

Most datasets have more positive tweets than negative ones. Models perform well for the positive class but poorly for the negative one.

- Different topics

A model trained on tweets about movies may not work well on tweets about politics or sports. Accuracy drops when the topic changes.

5. CONCLUSION AND FUTURE SCOPE

In our work we used different machine learning model. We also used deep learning model. Our work showed that deep learning model achieved better result. There are some limitations also like noisy data, sarcasm detection, unbalanced datasets. In future we applied our model on Multimodal dataset.

Data Availability

The experimental dataset retrieve from kaggle.

Conflict of Interest

No conflict of interest

Funding Source

No funding

Authors' Contributions

All authors contributed to the conceptualization, methodology, implementation, data analysis, and preparation of the manuscript. Dr. Radha Krishna Jana supervised the research and critically reviewed the manuscript. All authors approved the final version.

ACKNOWLEDGMENT

Thanks to our Hod Sir, Dean Madam and VC Sir for motivating this research work..

REFERENCES

- [1] R. K. Jana, S. Maity, and S. Maiti, "An empirical study of sentiment and behavioural analysis using homophily effect in social network," *2022 6th International Conference on Intelligent Computing and Control Systems (ICICCS)*, pp. 1508–1515, IEEE, 2022.
- [2] R. K. Jana and S. Maity, "An accuracy-based comparative study on different techniques and challenges for sentiment analysis," *Pervasive Computing and Social Networking: Proceedings of ICPCSN 2022*, pp. 601–619, 2022.
- [3] K. Radha, S. Dharmapal, M. Saikat, and P. Hrithik, "Perception analysis of social network tweets using a genetic algorithm-based approach," *Journal of Software Engineering Tools & Technology Trends*, vol. 10, no. 2, pp. 27–38, 2023.
- [4] R. K. Jana, D. Singh, S. Maity, and H. Paul, "A hybrid approach to analyse the public sentiment on COVID-19 tweets," *Indian Journal of Science and Technology*, vol. 17, no. 7, pp. 610–616, 2024.
- [5] R. K. Jana, D. Singh, and S. Maity, "Modified firefly algorithm and different approaches for sentiment analysis," *The Scientific Temper*, vol. 15, no. 1, pp. 1745–1754, 2024.
- [6] R. K. Jana, D. Singh, S. Maity, H. Paul, A. Mallick, S. Ghatak, S. Mallik, and M. Wang, "Public perception analysis on COVID-19 tweets using hybridised method," *International Journal of Computational Biology and Drug Design*, vol. 16, no. 1, pp. 19–41, 2024.
- [7] A. Mollah and A. A. Munshi, "Sentiment Classification Using LSTM With Combined Datasets," arXiv preprint arXiv:2212.01791, 2022.
- [8] R. Nithya et al., "Twitter Sentiment Analysis using Machine Learning Algorithms," *International Journal of Engineering Research & Technology (IJERT)*, vol. 12, no. 03, pp. 1-5, 2023.
- [9] P. Bhushan and A. Mahato, "Twitter Sentiment Analysis Using Machine Learning and Deep Learning Approaches," *International Journal of Modern Education and Computer Science (IJMECS)*, vol. 16, no. 6, pp. 11-21, 2024.
- [10] Comparative Study of Classification Algorithms for Sentiment Analysis," *IJERT*, 2024 (Decision Tree comparison).
- [11] I. Setiawan et al., "Utilizing Random Forest Algorithm for Sentiment Prediction," Atlantis Press, 2022.
- [12] F. Suandi et al., "Enhancing Sentiment Analysis Performance Using SMOTE," Atlantis Press, 2024 (KNN/RF/SVM experiments).
- [13] Das, S., Nag, S. and Jana, R., 2025. A Multimodal Systematic Review of Stress Detection in University Students Based on Machine Learning and Physiological Measures.
- [14] Samarпита Roy, Dr. Radha Krishna Jana, Dr. Bidisha Bhabani, Dr. Tanaya Das, "From Tweets to Trades: Applications of Financial Sentiment Analysis in Market Forecasting", *Journal of Computer Science and Engineering in Innovations and Research* (ISSN No: 3049-1762 online), June 15, 2025, Vol2(1), pp 1-6.
- [15] Shabbir Ahamad Molla, Sarmistha Sarkar, Dr. Radha Krishna Jana, Dr. Bidisha Bhabani, Dr. Tanaya Das "The Sentiment Behind the Tweet", *Journal of Computer Science and Engineering in Innovations and Research* (ISSN No: 3049-1762 online), June 15, 2025, Vol2(1), pp 7-12.
- [16] Dr. Radha Krishna Jana, Dr. Atanu Kotal, Asit Kumar Nayek "Healthcare Made Smarter: AI-Driven Drug Recommendation Systems", *Journal of Computer Science and Engineering in Innovations and Research* (ISSN No: 3049-1762 online), June 15, 2025, Vol2(1), pp 25-30.
- [17] Debanjan Pan, Dr. Radha Krishna Jana, Tirthajyoti Nag, Achyut Mitra "Analyzing Sentiments to Detect Drug-Related Tweets: A Software EngSN No: 3049-1762 online), June in engineering Framework", *Journal of Computer Science and Engineering in Innovations and Research* (IS15, 2025, Vol2(1), pp 31-36.
- [18] Sandip Mandal, Shemanti Pal, Dr. Radha Krishna Jana, Asit Kumar Nayek "Uncovering Guest Sentiment: Analyzing Hotel Reviews", *Journal of Computer Science and Engineering in Innovations and Research* (ISSN No: 3049-1762 online), June 15, 2025, Vol2(1), pp 37-42.
- [19] Neelanjan Ghosh, Radha Krishna Jana, "Tweeting Health: Multimodal Sentiment Analysis for Healthcare Insights", *4th International Conference on Innovations in Data Analytics (ICIDA 2025)*, Eminent College of Management and Technology (ECMT), West Bengal, India, 5th - 6th December 2025.
- [20] Neelanjan Ghosh, Radha Krishna Jana, "Emotion-Driven Medication Recommendations for Patients", *4th International Conference on Innovations in Data Analytics (ICIDA 2025)*, Eminent College of Management and Technology (ECMT), West Bengal, India, 5th - 6th December 2025.

AUTHORS PROFILE (All authro profiles are mandateory)

Author-1 Piu Ghosh is a final-year B.Tech student specializing in Computer Science and Engineering with a focus on Artificial Intelligence and Machine Learning. She is developing her technical knowledge and skills in AI-driven technologies and emerging innovations in the field of computing.



Author-2 Ritusree Naskar is a B.Tech student specializing in Computer Science and Engineering with a focus on Artificial Intelligence and Machine Learning. She is passionate about exploring emerging technologies, developing technical skills, and gaining knowledge in AI-driven innovations and intelligent computing systems.



Author-3 Dr. Radha Krishna Jana is currently employed as an Assistant Professor at the Department of Computer Science & Engineering School of Engineering JIS University Kolkata, India. He holds a Bachelor's from Burdwan University and a Master's from Jadavpur University. He obtained his Ph.D. in Computer Science & Engineering from JIS University, India. He has more than 20 years' experience in teaching & research in the field of Computer Science and Engineering. The research interest of the authors is in the area of AI in Medicine and Healthcare, AI & ML in Disease Prediction, Big Data Analytics in Healthcare. He has published more than 50 research papers and 3 Books. He has served as reviewer in different international conferences.

