International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 02 | Feb - 2025 SJIF Rating: 8.448 ISSN: 2582-3930

Twitter sentiment Analysis Using Machine Learning

[1]Ashwini Domade Department of Information and Technology from Matoshri Aasrabai Polytechnic
[2]Snehal Gadakh Department of Information and Technology from Matoshri Aasrabai Polytechnic
[3]Vaishnavi Kardile Department of Information and Technology from Matoshri Aasrabai Polytechnic
[4]Damini Pekhale Department of Information and Technology from Matoshri Aasrabai Polytechnic
[5]Ms.Vidya Kale lecturer of Information Technology from Matoshri Aasrabai Polytechnic
[6]Mr.Mahesh Bhandakkar Head of Information Technology from Matoshri Aasrabai Polytechnic

---

Abstract - abstract on page 1 With the development and expansion of web technology, a vast amount of data is generated and available to internet users, and the internet has evolved into a platform for online learning, idea exchange, and opinion sharing. Because they enable people to share and express their opinions on various topics, engage in discussions with various communities, or post messages globally, social networking sites like Facebook, Google, and Twitter are quickly becoming more and more popular. A lot of work has been done in the field of sentiment analysis of Twitter data, which is useful for analyzing the information in tweets where opinions are highly unstructured and heterogeneous and are either

This paper presents a survey and comparative analyses of current techniques for opinion mining, such as machine learning and lexicon-based approaches, along with evaluation metrics using various machine learning algorithms, such as naive bayes max entropy and support vector machine. Given the growth and advancement of web technology, there is a huge volume of data available on the internet for internet users, and a lot of data is generated too. In some cases, this data is neutral, positive, or negative. We offer data stream research on Twitter. Additionally, we talked about the basic difficulties and uses of sentiment analysis on Twitter keywords. Opinion mining machine learning using sentiment analysis on Twitter Maximum Entropy Support Vector Machine (SVM) naïve Bayes NB

***Key Words*:** Twitter, Sentiment analysis (SA), Opinion mining, Machine learning, Naive Bayes (NB), Maximum Entropy, Support Vector Machine (SVM).

1. INTRODUCTION

Social media platform such as Twitter provides a space where users share their thoughts and opinion as well as con- nect, communicate, and contribute to certain topics using short, 140 characters posts, known as *tweets*. This can be done through texts, pictures, and videos, etc., and users can interact using likes, comments, and reposts buttons. Accord- ing to Twitter (https://investor.twitterinc.com), the platform has more than 206 million daily active users in 2022, which is defined as the number of logged accounts that can be iden- tified by the platform and where ads can be shown. As more people contribute to social media, the analysis of informa- tion available online can be used to reflect on the changes in people's perceptions, behavior, and psychology (Alamoodi et al. 2021). Hence, using Twitter data for sentiment analysis has become a popular trend. The growing interest in social media analysis has brought more attention to Natural Lan- guages Processing (NLP) and Artificial Intelligence (AI) technologies related to text analysis.

Using text analysis, it is possible to determine the senti- ments and attitudes of certain target groups. Much of the available literature focuses on texts in English but there is a growing interest in multilanguage analysis (Arun and Srinagesh 2020a; Dashtipour et al. 2016; Lo et al. 2017). Text analysis can be done by extracting subjective com- ments toward a certain topic using different sentiments such as Positive, Negative, and Neutral (Arun and Srinagesh 2020b). One of the topical interests would be related to the Coronavirus (Covid-19), which is a novel disease that was first discovered in late 2019. The rapid spread of Covid-19 worldwide has affected many countries, leading to changes in people's lifestyles, such as wearing masks on public transportation and maintaining social distancing. Sentiment analysis can be implemented to social media data to explore changes in people's behavior, emotions, and opinions such as by dividing the spread trend of Covid-19 into three stages and exploring people's negative sentiments toward Covid-19 based on topic modeling and feature extraction (Boon-Itt and Skunkan 2020). Previous studies have retrieved tweets based on certain hashtags (#) used to categorize content based on certain topics such as "#stayathome" and "#socialdistanc- ing" to measure their frequency (Saleh et al. 2021). Another study has used the Word2Vec technique and machine learn- ing models, such as Naive Bayes, SVC, and Decision Tree, to explore the sentimental changes of students during the online learning process as various learning activities were moved online due to the pandemic (Mostafa 2021).

2. PROBLEM STATEMENT

With the explosive growth of social media, platforms like Twitter have become a vital source for capturing public sentiment on various topics ranging from products and services to political events and social issues. However, the sheer volume and unstructured nature of this data make it challenging to extract meaningful insights. Companies, governments, and organizations often struggle to manually analyze and understand public opinion in real-time, missing opportunities to respond to trends or public concerns.

This project aims to address the challenge of automatically analyzing sentiments from Twitter data. The

system will classify tweets into positive, negative, or neutral categories using machine learning techniques. The solution will help businesses, political organizations, or individuals make data-driven decisions based on public opinion by providing real-time sentiment analysis on Twitter data. Additionally, the system will visualize sentiment trends, providing a clearer understanding of the sentiment landscape around specific topics, events, or brands.

3. LITERATURE REVIEW

1] **Twitter as a Corpus for Sentiment Analysis and Opinion Mining**

Numerous sentiment analysis solutions utilize machine learning and natural language processing techniques for analyzing social media data, particularly Twitter. Studies have demonstrated that systems like sentiment classifiers or opinion mining tools are effective in classifying data into categories (positive, negative, neutral). However, many systems struggle with real-time analysis, handling of noisy data (e.g., slang, emojis), and managing large volumes of streaming data from platforms like Twitter. Furthermore, existing solutions are often limited in their ability to detect nuanced sentiments such as sarcasm and irony.

Challenges:

Research reveals several challenges in Twitter sentiment analysis systems, including:

1) **Noisy Data**: Tweets often contain informal language, special characters, and abbreviations.
2) **Short Text Length**: The 280-character limit on tweets limits context and makes classification difficult.
3) **Sarcasm & Irony**: Detecting sarcasm is difficult with basic NLP techniques.
4) **Real-Time Processing**: Managing and analyzing real-time data streams from Twitter requires robust and scalable systems.

2. Sentiment Analysis of Twitter Data

This paper develops a sentiment analysis model tailored to Twitter, focusing on how unique features like emoticons and hashtags contribute to sentiment. The study finds that accounting for these features enhances classification accuracy.

3. Convolutional Neural Networks for Sentence Classification

This research explores the use of Convolutional Neural Networks (CNNs) for sentiment classification. It shows that CNNs outperform traditional approaches in handling short texts like tweets by effectively capturing semantic nuances.

4. Sentiment Analysis and Opinion Mining

Liu provides a broad review of sentiment analysis techniques, especially their application to social media platforms. The paper outlines the challenges of working with Twitter data and offers solutions like lexicon-based approaches and advanced machine learning models.

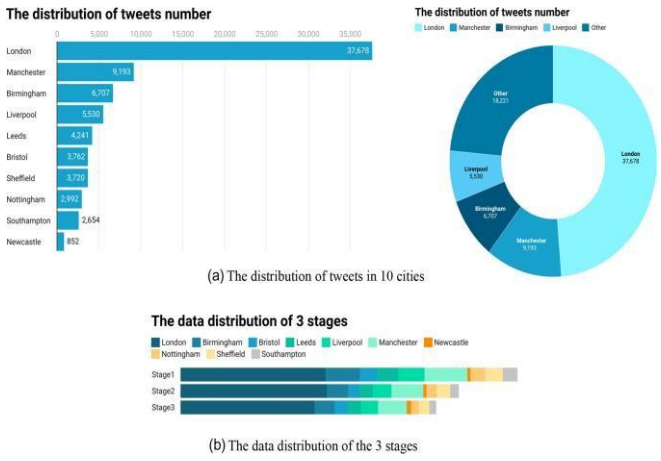
5. A System for Real-time Twitter

The authors present a real-time sentiment analysis system used to track public opinion during the 2012 U.S. Presidential Election. The study demonstrates the effectiveness of real-time sentiment tracking but also discusses the importance of data quality and volume in maintaining accuracy.

6. Deep Learning in Sentiment Analysis:

Recent advancements in deep learning have further improved sentiment classification performance. Studies like Socher et al. (2013) used Recursive Neural Networks (RNN) and Convolutional Neural Networks (CNN) to achieve state-of-the-art results on sentiment analysis tasks. These models can capture contextual and semantic information from text, making them ideal for handling complex sentiment patterns in tweets

4. PROPOSED SYSTEM



(a) The distribution of tweets in 10 cities

(b) The data distribution of the 3 stages

Fig: Distribution of collected tweets based on the selected cities and different stages

the data were pre-processed before sentiment analysis (Naseem et al. 2021). We implemented a basic data-cleaning process as follows:

(1) Replacing upper-case letters to avoid recognizing the same word as different words because of capitalization.
(2) Removing hashtags (#topic), mentioned usernames (@username), and all the links that start with "www," "http," and "https." Removing stop words and short words (less than two characters). The stop words are mostly very common in the text but hardly contain any sentiment polarity. However, in sentiment analy- sis, "not" and "no" should not be listed as stop words,

because removing these negations would change the real meaning of entire sentences.

(3) Reducing repeated characters from some words. Some users will type repeated characters to express their strong emotions, so these words that are not in the lexicons should be converted into their corresponding correct words. For example: “sooooo goooood” becomes “so good.”

(4) Expanding contractions in tweets such as “isn't” or “don't” as these will become meaningless letters or words after punctuations have been removed. Therefore, all contractions in the tweets are expanded into their formal forms, such as “isn’t” become “is not.”

(5) Clearing all non-alphabetical characters or symbols including punctuation, numbers, and other special symbols that may affect the feature extraction of the text.

(6) Removing duplicated or empty tweets and creating a clean dataset.

(7) Converting emojis to their real meaning as many Twitter users use emojis in their tweets to express their sentiments and emotions. Hence, using the demojize() function in the emoji module of Python and transforming emojis into their true meaning may improve the accuracy of the sentiment analysis (Tao and Fang 2020).

5. Application and Future Scope of the Project

1. Applications that use Reviewsfrom Websites:

Today Internet has a large collection of reviews and feedbacks on almost everything. This includes product reviews, feedbacks on political issues, comments about services, etc.

2. Applications as a Sub-component Technology:-

A sentiment predictor system can be helpful in recommender systems as well. The recommender system will not recommend items that receive a lot of negative feedback or fewer ratings.

3. Applications in Business Intelligence:-

It has been observed that people nowadays tend to look upon reviews of products which are available online before they buy them. And for many businesses, the online opinion decides the success or failure of their product.

4. Applications across Domains:

Recentresearches in sociology and other fields like medical, sports have also been benefitted by Sentiment Analysis that show trends in human emotions especially on social media.

5. Applications In Smart Homes Smart homes

are supposed to be the technology of the future. In future entire homes would be networked and people would be able to control any part of the home using a tablet device.

Scopes :-

1]This project will employ supervised learning methods like logistic regression support vector machine (SVM) or naive bayes for classification.

2]Its primary aim is sentiment analysis for English-language tweets.

3] The system can be expanded to monitor shifts in public opinion and analyze sentiment patterns over time.

4] Businesses, political organizations, and individuals may find it helpful to know how the general public feels about a product, event, or policy.

6. CONCLUSIONS

In conclusion, Six inferences can be drawn from this. Three phases make up this project, which collects data on COVID-19 from Twitter users in key English cities. By first cleaning the data and then using unsupervised lexicon-based methods to categorize the sentiment orientations of the tweets at each stage, we can identify the three stages of public sentiment changes regarding the COVID-19 pandemic for the majority of cities. Next, we use supervised machine learning methods with a sample of annotated data to train the random forest classifier multinomial nave bayes classifier and svc, respectively, from lexicon-based approaches, the percentage of favorable sentiments rises initially before falling as the percentage of negative opinions changes in a different way. However, lexicon-based approaches adapt the current lexicon to better suit the linguistic patterns of contemporary social media. enhancing this method's accuracy Furthermore, it is possible to draw the conclusion that the rise in confirmed cases and the fall in vaccination volume could be the cause of the rise in unfavorable attitudes by examining the number of deaths, confirmed cases, and vaccination conditions. Moreover, an annotated dataset can be created for comparison. The discrepancies between anticipated and actual results Future research can combine the sentiment classification results with other variables, such as mortality and vaccination rates, and develop a regression model to investigate which factors influence sentiment shifts. Time series can also be used as the foundation for COVID-19 studies to evaluate how people's attitudes and perceptions have changed over time. Overall, the study showed that several sentiment analysis techniques, such as svc employing bow or tfidf, outperformed the model accuracy overall.

REFERENCES

1. Adwan Oy Al-Tawil, Huneiti, Shahin Ra Abu Zayed, and Al-Dibsi rh 2020 Twitter sentiment analysis methods for a survey in Emerging Technologies
2. . https://doi.org/10.3991/ijet. v15i15.14467
3. al Amrani Y, Lazaar M, el Kadirp KE (2018) Random forest and support vector machine based hybrid approach to sentiment analysis. Proc Comput Sci. https://doi.org/10.1016/j.procs. 2018.01.150
4. Alamoodi AH, Zaidan BB, Zaidan AA, Albahri OS, Mohammed KI, Malik RQ, Almahdi EM, Chyad MA, Tareq Z, Albahri AS, Hameed H, Alaa M (2021) Sentiment analysis and its applica- tions in fighting COVID-19 and infectious diseases: a system- atic review. Expert Syst Appl. https://doi.org/10.1016/j.eswa. 2020.114155
5. Aqlan AAQ, Manjula B, Lakshman Naik R (2019) A study of senti- ment analysis: Concepts, techniques, and

challenges. In Lecture notes on data engineering and communications technologies, vol 28. https://doi.org/10.1007/978-981-13-6459-4_16

6. Arun K, Srinagesh A (2020a) Multi-lingual Twitter sentiment analy- sis using machine learning. Int J Electr Comput Eng. https://doi. org/10.11591/ijece.v10i6.pp5992-6000
7. Arun K, Srinagesh A (2020b) Multi-lingual Twitter sentiment analy- sis using machine learning. Int J Electr Comput Eng. https://doi. org/10.11591/ijece.v10i6.pp5992-6000
8. Baccianella S, Esuli A, Sebastiani F (2010) SENTIWORDNET 3.0: an enhanced lexical resource for sentiment analysis and opinion mining. In: Proceedings of the 7th international conference on language resources and evaluation, LREC 2010
9. Barbosa L, Feng J (2010) Robust sentiment detection on twitter from biased and noisy data. In: Coling 2010—23rd interna- tional conference on computational linguistics, proceedings of the conference, 2
10. Bonta V, Kumaresh N, Janardhan N (2019) A comprehensive study on Lexicon based approaches for sentiment analysis. Asian J Comput Sci Technol 8(S2):1–6. https://doi.org/10.51983/ajcst- 2019 8.s2.2037
11. Boon-Itt S, Skunkan Y (2020) Public perception of the COVID-19 pandemic on Twitter: sentiment analysis and topic modeling study. JMIR Public Health and Surveillance, 6(4), e21978. https://doi.org/10.2196/21978
12. Breiman L (2001) Random forests. Mach Learn. https://doi.org/10.
13. 1023/A:1010933404324