

Twitter Sentiment Analysis using Machine Learning

G. Manikandan, M. Robinson Joel, S. Lidiya Jones Raj, D. Madlin Jency

Department of Information Technology, Kings Engineering College, Chennai, India

ARTICLE INFO

Article History:

Accepted: 05 April 2023

Published: 27 April 2023

Publication Issue

Volume 10, Issue 2

March-April-2023

Page Number

665-669

ABSTRACT

Social media makes it easier for people to communicate with one another online. Social media encompasses a wide range of applications and platforms, including Facebook for entertainment, Instagram for youth, Twitter for social and political, and YouTube, that let users share information, communicate online, and create communities. More than 4.7 billion individuals, or nearly 60% of the world's population, utilise social media. Twitter is a popular social media platform where users may express their feelings and opinions. In order to determine user sentiments, this Twitter sentiment analysis study uses sentiment analysis to data from tweets on the social media site. A whole new set of problems, such as the usage of slang and acronyms, are brought about by the relatively small size of the tweet format. Our objective is to carry out research on Twitter sentiment analysis while outlining the methodology, models, and generalised Python-based approach that was employed.

Keywords: Classification, Data Preprocessing, Machine Learning, Sentiment Analysis.

I. INTRODUCTION

Users of the social networking site Twitter can submit tweets, which are brief communications that can have a maximum character count of 280. Users can follow other users, and a chronological feed will display all of their tweets. Since its founding in 2006, Twitter has developed into one of the most widely used social media networks worldwide. Utilizing hashtags to categorise tweets and increase their discoverability by other users is one of Twitter's distinctive features. The pound symbol (#) is used to form hashtags, which are then followed by a word or phrase. It's simple to share and interact with material on Twitter since users may retweet, reply to, and also like other users' tweets.

Twitter has developed into a potent tool for exchanging information and facilitating communication, especially during political campaigns and breaking news events. Politicians and journalists frequently use Twitter to disseminate information and interact in real time with their audiences. Twitter has also been used to coordinate social movements and demonstrations, like Black Lives Matter and the Arab Spring. Twitter has still come under fire for its part in disseminating false information and hate speech. The platform has responded to these problems by enacting rules to identify and delete deceptive information and by banning users who breach its terms of service. Despite these initiatives, Twitter still has a difficult

time balancing the right to free expression with the need to keep its site welcoming and secure for all use. Twitter has significantly influenced contemporary society and is the origin of many popular viral phenomena and memes. It has, however, also come under fire for its part in disseminating false information and for how it handles online abuse and harassment. Despite these difficulties, Twitter continues to be a vital platform for engagement and communication among people, companies, and organisations.

II. LITERATURE REVIEW

The many machine learning methods used to analyse Twitter data, such as Naive Bayes, SVM, and the Maximum Entropy Method, are discussed. To harvest the sentiments, many parts of Twitter data are being analysed. In this study's sentiment analysis of Twitter, the term "opinion" is defined. Sentiment analysis examines opinions broken down into positive, negative, and neutral categories. The study demonstrates that the most accurate machine learning technique is Naive Bayes, which can be used as a baseline technique. Maximum Entropy techniques are also sometimes highly successful. To improve the performance measurements, further effort will be required in the future. They[1] provide an overview of a pipeline that is frequently utilised when developing social media-based apps and concentrate on talking about the various analytical approaches, including topic analysis, time series analysis, sentiment analysis, and network analysis. The effects of such applications are then presented in three distinct fields, including disaster management, healthcare, and business. They conclude by outlining current difficulties and outlining interesting future research trajectories in the areas of data privacy, 5G wireless networks, and multilingual assistance. Continuous word representations are characteristics for classifying Twitter sentiment using supervised learning. They[2] acquire the complex word representations through use

a modified version of Delta TFIDF to weight the vectors. The experimental findings show that the suggested strategy is viable and capable of greatly enhancing sentiment analysis performance. They[3] have developed a model to sentiment analysis which allows the processing of Twitter API streaming feed in real time and to classify its polarity to provide valuable insight in industry and users. Our built classifier can be utilized as data analysis tools in NLTK. Therefore, in general we can use our proposal technique to sentiment analysis for any device, public figure or sports team that is better than any other existing model with high accuracy performance. They[4] have implemented various Machine Learning classification algorithms and NLP techniques on a large, imbalanced, multi-classed, and real-world dataset to analyse sentiment. Our best approaches provide 77% accuracy with both Support Vector Machine and Logistic Regression algorithm along with the Bag-of-Words technique. In the future, we would like to apply more advanced techniques to increase accuracy and will also try to build a generalized and robust model for similar datasets. They[5] have presented a method for an automatic collection of a corpus that can be used to train a sentiment classifier. We used Tree Tagger for POS-tagging and observed the difference in distributions among positive, negative and neutral sets. From the observation conclude that authors use syntactic structures to describe emotions or state facts. Some POS-tags may be strong indicators of emotional text. Their[6] analysis provides detailed insight into the COVID-19 datasets of both social networks, this paper can be used to comprehend people's reactions and the attention they pay to certain topics. We intend to extend our experiments to gain a deeper understanding of social network responses to the pandemic over a longer period of time. In addition, They intend to investigate the measures taken by governments and their results. They[7] performed sentiment analysis using the random forest algorithm and achieved an accuracy of around 75%. I would recommend you try and use some other machine

learning algorithms such as logistic regression, SVM, or KNN and see if you can get better results. They [8] analysed sentiments carried by Emojis have not been taken care. This can be considered as a future direction of this research by incorporating techniques which can extract sentiment/opinion information from Emojis. Emotion analysis was also considered and the anger among the people was the top emotion that dominated the tweets, followed by the fear from the first attempt to try out distance learning. To further enhance the readability of the mined reasons, They [9] select the most representative negative tweets to define the latent reasons behind the negative results of online learning. Absence of face to face communication, network system breakdown, ambiguity and games were the most significant reasons behind the negative sentiment. They [10] investigated some queried tweets of around 39 weeks of a sensational tragic issue of “Central American migrant caravans” and how the tweets played their critical roles in the overall sentiment on the twitter sphere. We used Vader Sentiment Analysis to analyze the sentiment of each tweet before estimating the sentiment on Twitter toward this news. We show that such simple estimates can be very accurate. Their study [11] intends to study and discover the public perception and sentiment towards the #LAWAN rally. The investigation also questions whether is the civil society supportive of such social movement and rally, especially when rally, protests and demonstration are traditionally and often seen as a negative and non-constructive approach in the country. Tweets with the hashtag #LAWAN over a day on and before the rally day on 31st July 2021 have been collected and analysed. Sentiment analysis helped to identify the public sentiments towards the rally, and topic modelling helped to discover common topics from the 5000+ tweets scrapped from the social media. Their [12] approach to convert unstructured WhatsApp messages to a structured form is discussed on which various data mining techniques for sentiment analysis can be performed. Our approach to deal with code-mixing, different emojis and the

emotions they depict, and finally, perform basic analysis using this algorithm is discussed. Their [13] proposed approach uses numerical attributes and textual attributes with a blend of clustering and topic modelling techniques for analyzing trends and detecting an event. The proposed model successfully detects an event up to 80% of accuracy. The comparison of a detected event and real time event statistics drawn from social media and news channels highlights the performance and accuracy of the proposed model. They [14] have taken into account various methods to perform sentiment analysis. Sentiment Analysis has been performed by using Machine Learning Classifiers. Polarity-based sentiment analysis, and Deep Learning Models are used to classify user's tweets as having ‘positive’ or ‘negative’ sentiment. The idea behind taking in various model architectures was to account for the variance in the opinions and thoughts existing on such social media platforms. These classification models can further be implemented to classify live tweets on twitter on any topic. Their [15] main contribution towards this paper is (i) to analyze and interpret public opinion and sentiments related to the post posted on the social media (ii) to calculate the effectiveness of that post corresponding to that social issue. For this they have used Senti word net to classify the words as positive or negative and used the concept of Naive Bayes to classify the feedback into positive and negative classes using probability based function and support vector machine to create decision boundary from which they can determine the effectiveness of the post.

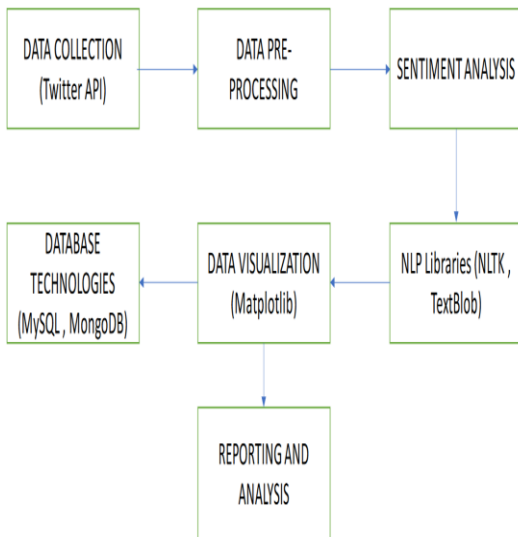
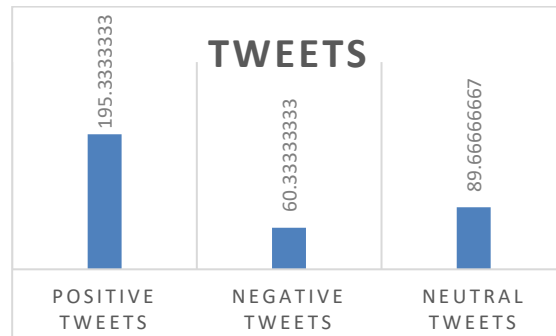
III. PROPOSED APPROACH

The proposed work focuses on a strategy to sentiment analysis on Twitter data. The data is extracted from a particular project or product, pre-processed the extracted Tweet using Natural Language Tool Kit (NLTK) and Classifier model that calculates the sentiment of each Tweet. In sentiment Analysis, the Twitter data is using data extraction , tokenization ,

stemming , lemmatization , stop word removal, named entity recognition, parts of speech tagging, create a data frame, text modelling, a classifier model and each of these processes has its own algorithm and packages to be processed. TF-IDF model has used to find out important words from tweets to predict sentiment. To uncover the sentiments, at first we have built a classifier model by using a data. In this the developed Natural Language Processing(NLP) is used based on pre-processed data framework to filter tweets and sentiment analysis . TF-IDF model helps to find out the most important word in tweets. It helps in sentiment analysis with high precision. In this analysis it evaluates whether the tweet is positive or negative or neutral.

label based on the highest score. Random forest: is a supervised algorithm for constructing multiple decision tree.

Bagging: is a classifier used to taking multiple random samples and use each sample separately to construct a prediction model.



IV. CONCLUSION AND FUTURE SCOPE

In this work, we created a model for sentiment analysis that enables real-time processing of Twitter API streaming feeds and categorises their polarity to offer insightful data about the market and users . In NLTK, our built-in classifier may be used as data analysis tools. As a result, in general, we can do sentiment analysis for any gadget, celebrity, or sports team using our proposed approach, which is superior to any other current model in terms of accuracy performance.

V. REFERENCES

SENTIMENT ANALYSIS

After that, multiple supervised learning algorithms applied for the purpose of training: Naive Bayes, support vector machine (SVM), maximum entropy, decision tree, random forest and bagging. Naïve Bayes: is defined as classifier used to determine the most probable class label for each object. Support vector machine: is defined as supervised model, used for classification, regression analysis. Maximum entropy: is a classifier used for large variety of text classification. Decision tree: are flexible algorithms used to assign

[1]. Q. Hou, M. Han and Z. Cai, "Survey on data analysis in social media: A practical application aspect," in Big Data Mining and Analytics, vol. 3, no. 4, Dec. 2020,

[2]. R. Othman, Y. Abdelsadek, K. Chelghoum, I. Kacem and R. Faiz, "Improving Sentiment Analysis in Twitter Using Sentiment Specific Word Embeddings," 2019 10th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS), Metz, France, 2019,

- [3]. M. R. Hasan, M. Maliha and M. Arifuzzaman, "Sentiment Analysis with NLP on Twitter Data," 2019 International Conference on Computer, Communication, Chemical, Materials and Electronic Engineering (IC4ME2), Rajshahi, Bangladesh, 2019,
- [4]. L. Mandloi and R. Patel, "Twitter Sentiments Analysis Using Machine Learning Methods," 2020 International Conference for Emerging Technology (INCET), Belgaum, India, 2020, pp.
- [5]. A. Pak and P. Paroubek. „Twitter as a Corpus for Sentiment Analysis and Opinion Mining". In Proceedings of the Seventh Conference on International Language Resources and Evaluation, 2010,
- [6]. R. Singh and P. Sharma, "An Overview of Social Media and Sentiment Analysis," 2021 5th International Conference on Information Systems and Computer Networks (ISCON), Mathura, India, 2021
- [7]. Wen, Shuo et al. "Sentiment Analysis of Social Media Comments based on Random Forest and Support Vector Machine." 2022 IEEE International Conference on Advances in Electrical Engineering and Computer Applications (AEECA) (2022)
- [8]. Afroz, Nida et al. "Sentiment Analysis of COVID-19 Nationwide Lockdown effect in India." 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS) (2021).
- [9]. S. Daulatkar and A. Deore, "Post Covid-19 Sentiment Analysis of Success of Online Learning: A Case Study of India," 2022 9th International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, India, 2022,
- [10]. S. Muttineni and J. Deng, "Estimating Sentiment in Social Media - a Case Study of the Migrant Caravans News on Twitter," 2021 IEEE 15th International Conference on Semantic Computing (ICSC), Laguna Hills, CA, USA, 2021,
- [11]. S. L. Lau, M. Y. H. Lee and M. X. Teoh, "Understanding Public Sentiment Towards a Public Rally Using Text and Social Media Analytic," 2022 IEEE International Conference on Computing (ICOCO), Kota Kinabalu, Malaysia, 2022
- [12]. A. Mohta, A. Jain, A. Saluja and S. Dahiya, "Pre-Processing and Emoji Classification of WhatsApp Chats for Sentiment Analysis," 2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), Palladam, India, 2020
- [13]. P. Tijare and P. Jhansi Rani, "A Sentiment Driven Approach to Detect an Offline Event on Social Media Platform," 2020 2nd PhD Colloquium on Ethically Driven Innovation and Technology for Society (PhD EDITS), Bangalore, India, 2020
- [14]. Chandra and A. Jana, "Sentiment Analysis using Machine Learning and Deep Learning," 2020 7th International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, India, 2020,
- [15]. K. Saini, R. K. Mishra, K. Bhati and S. Kawoosa, "Social media based linking approach for finding the effectiveness of the social issues," 2016 International Conference on Computing, Communication and Automation (ICCCA), Greater Noida, India, 2016

Cite this article as :

G. Manikandan, M. Robinson Joel, S. Lidiya Jones Raj, D. Madlin Jency, "Twitter Sentiment Analysis using Machine Learning ", International Journal of Scientific Research in Science, Engineering and Technology (IJSRSET), Online ISSN : 2394-4099, Print ISSN : 2395-1990, Volume 10 Issue 2, pp. 665-669, March-April 2023. Available at doi : <https://doi.org/10.32628/IJSRSET2310281>
Journal URL : <https://ijsrset.com/IJSRSET2310281>