Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 12 Issue 07, July-2023

Twitter Sentiment Analysis Using Machine Learning Algorithms

Nithya N
Computer Science and Engineering
RNS Institute of Technology
Bangalore, India

Devaraju B M
Computer Science and Engineering
RNS Institute of Technology
Bangalore, India

Dr. Girijamma H A
Computer Science and Engineering
RNS Institute of Technology
Bangalore, India

***Abstract*—Twitter has become a prominent platform for people to express their opinions, emotions, and sentiments on a wide range of topics. Extracting valuable insights from this massive pool of user-generated data is a challenging task. This project focuses on developing an effective sentiment analysis model for Twitter data, enabling opinion mining and trend analysis. The objective of this paper is to develop a robust sentiment analysis solution that can accurately classify tweets into positive, negative, or neutral sentiment categories. By analyzing the sentiment expressed in tweets, valuable insights can be gained into public opinion, customer feedback, and emerging trends. This enables organizations and individuals to make informed decisions, understand brand perception, and monitor sentiment fluctuations in real-time. The project tackles several challenges specific to Twitter sentiment analysis. These challenges include handling noise and data quality issues caused by abbreviations, misspellings, slang, and informal language commonly found in tweets. The presence of sarcasm, irony, and cultural references further complicates accurate sentiment analysis. The proposed sentiment analysis model leverages machine learning techniques, including advanced algorithms such as Naive Bayes, Support Vector Machines (SVM), Decision Tree and Logistic Regression. The model will be trained on a large dataset of labeled tweets to learn patterns and relationships between tweet features and sentiment labels. Various preprocessing techniques, feature extraction methods, and feature selection approaches will be employed to enhance the accuracy and generalization of the model. To evaluate the performance of the sentiment analysis model, standard evaluation metrics such as accuracy, precision, recall, and F1-score will be utilized. A comparison with existing approaches will be conducted to assess the model's effectiveness and competitiveness.**

Keywords— Twitter; sentiment ; VADER; NLP; polarity; subjectivity;

I. INTRODUCTION

Social media platforms like Twitter have become a significant source of information and opinions for millions of users worldwide. Twitter provides a platform for users to express their thoughts, emotions, and opinions in short text messages called tweets. Analyzing the sentiment behind these tweets can provide valuable insights into public opinion, customer feedback, and emerging trends. Sentiment analysis of Twitter data involves extracting and analyzing the sentiment expressed in tweets to understand the overall sentiment of a particular topic, brand, or event.

The paper focuses on sentiment analysis of Twitter data and does not delve into other aspects such as user demographics, geographical analysis, or network analysis. The project seeks to build a robust sentiment analysis solution that can provide valuable insights into public opinion, customer feedback, and sentiment trends on Twitter. The solution should achieve high accuracy in sentiment classification, considering the unique characteristics of Twitter data such as abbreviations, misspellings, slang, and the presence of emojis or emoticons. Additionally, the model should address challenges related to language ambiguity, sarcasm, and cultural references that can impact accurate sentiment interpretation. The goal is to develop a scalable and efficient sentiment analysis model that can handle real-time data streams and adapt to different domains or topics. The project also aims to explore techniques to address biases, ensure fairness, and protect user privacy in sentiment analysis.

The significance of Twitter sentiment analysis lies in its potential to provide actionable insights for various domains:

1. **Brands and Businesses:** Companies can analyze sentiment to gauge customer satisfaction, identify product issues, and assess the success of marketing campaigns.
2. **Politics and Elections:** Sentiment analysis on political tweets can help predict public opinion, track candidate popularity, and understand voter sentiment.
3. **Finance:** Investors and traders can leverage sentiment analysis to monitor market reactions, detect trends, and make informed decisions.
4. **Social Listening:** Brands and individuals can use sentiment analysis to monitor their reputation, respond to customer feedback, and engage with their audience.
5. **Disaster and Crisis Management:** Sentiment analysis can be applied to assess public response during emergencies, natural disasters, or public health crises.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 12 Issue 07, July-2023

II. OBJECTIVES

a) To collect a dataset of tweets related to a specific topic, brand, or event.

b) To preprocess the tweet data to remove noise, perform text normalization, and handle special characters and emoticons.

c) To apply various sentiment analysis techniques and machine learning algorithms to classify tweets into positive, negative, or neutral sentiment categories.

d) To evaluate the performance of the sentiment analysis model using appropriate evaluation metrics.

e) To analyze and interpret the results to gain insights into the sentiment trends and patterns related to the chosen topic or brand.

III. LITERATURE REVIEW

In [1], 7 machine learning models were implemented for identifying emotions and organizing the tweets as either happy or unhappy. A Voting Classifier is an ensemble machine learning technique that combines the predictions from multiple base classifiers to make a final prediction. It takes advantage of the diversity in predictions made by different classifiers to improve overall performance and generalization. The idea is that when multiple classifiers agree on a certain prediction, it is likely to be more accurate than the prediction of any single classifier.

[2] presents a solution to detect hate speech and offensive language in Twitter through machine learning. It uses n-gram features along with weighted TF-IDF values. Some tweets identified as offensive were misinterpreted as hate speech. 7000 tweets were extracted using the package twitteR to find out whether McDonalds had good reviews or KFC had good reviews [3]. There are various levels of sentiments such as, document level, sentence level and aspect level.

The first step in a data mining method is pre-processing. But, when it comes to text classification, the techniques such as tokenization, removal of stop words and lemmatization are taken into consideration [4]. A huge amount of tweets were extracted and the percentage of negative, null and positive hashtags were shown in a pie chart in [5]. In addition to ML Models, the different deep learning models such as LSTM, CNN and RNN have been used in [6]. The method of voting system helps in making final predictions. Sometimes, analysing the sentiments of tweets also helps in monitoring real-time events like predicting earthquakes and weather forecast.

A. SYSTEM ARCHITECTURE

System architecture refers to the structural design and organization of a complex system. It encompasses the arrangement of components, subsystems, modules, and their relationships, as well as the principles and guidelines governing their interaction and behaviour.

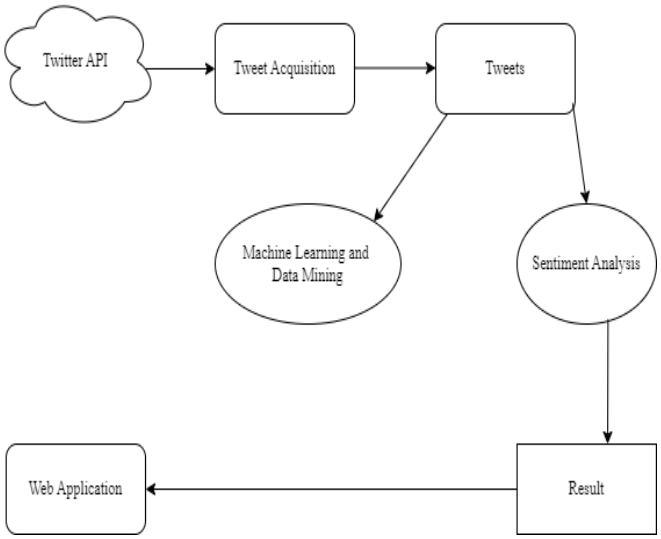

Fig 1: System Architecture

A group of tweets are collected for the purpose of sentiment analysis and the results generated after training the model is sent to web application, where the results are displayed

B. METHODOLOGY

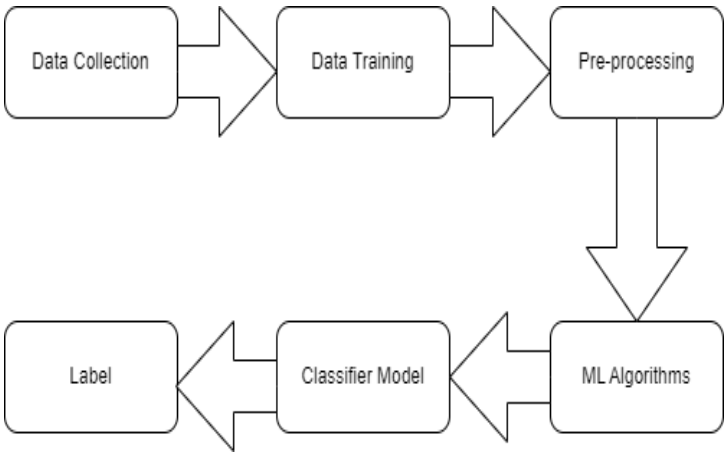

Fig 2: Methodology of System

The methodology followed by the system involves these steps: data collection, data training, data pre-processing, feature extraction, some text pre-processing techniques and sentiment analysis techniques such as rule-based methods and machine learning methods.

The twitter dataset is a labeled dataset consisting of separate train and test csv files. The train.csv dataset consists of 31,962 rows and 3 columns while the test.csv consists of 17,197 rows and 2 columns.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 12 Issue 07, July-2023

ML algorithms such as Decision Tree, Logistic Regression, Support Vector Classifier and Random Forest Classifier, were applied and the model was trained. In addition to using twitter dataset, we have considered cell phone reviews dataset analysing the sentiments of customers who use cell phones of particular brand. Among the top 10 unique brands booked by the customers, Samsung brand was frequently booked by many customers.

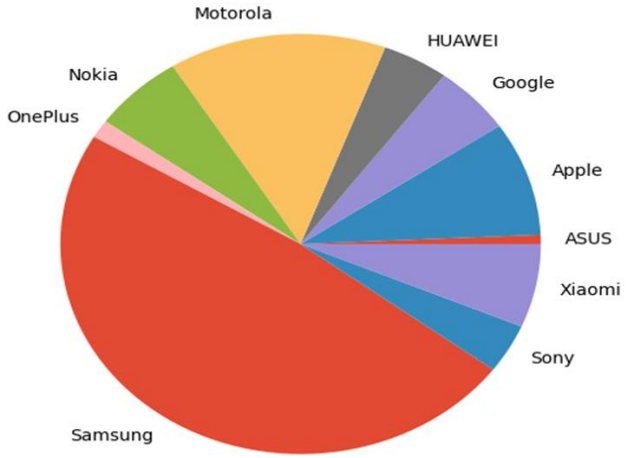

Fig 3: Pie Chart of total reviews by Brand

The pie chart shows that the total reviews given for each brand, by customers, could be either positive or negative. Samsung has more customers than any other brand.

IV. RESULTS

	Precision	Recall	F1-Score	Support
Negative	0.81	0.50	0.61	36169
Positive	0.60	0.86	0.71	31587
Accuracy			0.67	67756
Macro avg	0.70	0.68	0.66	67756
Weighted avg	0.71	0.67	0.66	67756

TABLE I.: Classification Report Table

Hence, there are overall 36,169 negative reviews and 31,587 positive reviews about different cell phone brands.

V. CONCLUSION

In this paper, we conducted sentiment analysis on Twitter data to gain insights into public opinion and sentiment trends. This involved collecting a dataset of tweets, preprocessing the data, applying various sentiment analysis techniques, and evaluating the performance of the sentiment analysis model. The findings and contributions of this project provide valuable insights and pave the way for future research and applications in the field of Twitter sentiment analysis. Through the analysis of sentiment trends, we observed that the majority of tweets in our dataset were negative, while a significant portion expressed positive or neutral sentiment. This indicates a diverse range of sentiments expressed on Twitter, highlighting the platform's significance as a medium for sharing opinions and emotions.

VI. FUTURE SCOPE

Noise can be handled more effectively and the model can be better improved by using deep learning algorithms. Ethical considerations, such as privacy concerns and responsible use of sentiment analysis, should also be prioritized in future research.

REFERENCES

[1] N. F. F. da Silva, E. R. Hruschka, and E. R. Hruschka, ‘‘Tweet senti ment analysis with classifier ensembles,’’ Decis. Support Syst., vol. 66, pp. 170–179, Oct. 2014.

[2] C. Kariya and P. Khodke, ‘‘Twitter sentiment analysis,’’ in Proc. Int. Conf. Emerg. Technol. (INCET), Jun. 2020, pp. 212–216.

[3] A. Alsaeedi and M. Zubair, ‘‘A study on sentiment analysis techniques of Twitter data,’’ Int. J. Adv. Comput. Sci. Appl., vol. 10, no. 2, pp. 361–374, 2019.

[4] A. Bandhakavi, N. Wiratunga, D. Padmanabhan, and S. Massie, ‘‘Lexicon based feature extraction for emotion text classification,’’ Pattern Recognit. Lett., vol. 93, pp. 133–142, Jul. 2017.J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68-73.

[5] J. Capdevila, J. Cerquides, J. Nin, and J. Torres, ‘‘Tweet-SCAN: An event discovery technique for geo-located tweets,’’ Pattern Recognit. Lett., vol. 93, pp. 58–68, Jul. 2017.

[6] T. Alsinet, J. Argelich, R. Béjar, C. Fernández, C. Mateu, and J. Planes, ‘‘An argumentative approach for discovering relevant opinions in Twitter with probabilistic valued relationships,’’ Pattern Recognit. Lett., vol. 105, pp. 191–199, Apr. 2018.

[7] W. Chen, Y. Zhang, C. K. Yeo, C. T. Lau, and B. S. Lee, ‘‘Unsupervised rumor detection based on users’ behaviors using neural networks,’’ Pattern Recognit. Lett., vol. 105, pp. 226–233, Apr. 2018.

[8] H. Hakh, I. Aljarah, and B. Al-Shboul, ‘‘Online social media-based sentiment analysis for us airline companies,’’ in New Trends in Information Technology. Amman, Jordan: Univ. of Jordan, Apr. 2017.