IJFANS INTERNATIONAL JOURNAL OF FOOD AND NUTRITIONAL SCIENCES

ISSN PRINT 2319 1775 Online 2320 7876

 Volume 11, Sp.Iss 5 , 2022

Sentiment Analysis Using Machine Learning in Twitter: A Web-Based Approach for Real-time Emotion Detection

A Jaswanth Kiran [1*], K Nithin [1*], B Raviteja [1*], M Ajay Bhavani [1*], Dr M Kavitha [1], M Kalyani [2]

[1*] Student, [1] Associate Professor, [2] Assistant Professor

[1*, 1] Department of Computer Science & Engineering, Koneru Lakshmaiah Education Foundation (KLEF), Vaddeswaram, Green fields, Guntur, Andhra Pradesh, India -522302.
[2] Department of Mathematics, PACE Institute of Technology and Sciences, Ongole, AP, India
Email: mkavita@kluniversity.in

DOI : 10.48047/IJFANS/11/Sp.Iss5/059

Abstract. Sentiment Analysis is process or technique that determine the emotional tone of a text whether it is positive, negative, or neutral using machine learning. Nowadays, social media generates a huge amount of text data from blogs, comments, and other sources. It takes a lot of time and money for an individual to analyse the sentiment in this data, thus a classification model is utilized instead. Navie Bayes is one of the finest classification models. Navie Bayes is designed based on Bayes theorem that mostly used for text categorization, to implement this model twitter dataset from Kaggle is used. A web-based approach is implemented to check the emotion of the person based on text. The joblib library is used to save the trained model, which is then imported into the Django project where an input form is provided, and the sentiment type positive or negetive or neutral is displayed as a result. The model's performence is evaluated using accuracy metric.

Keywords: Classification, Navie Bayes, Sentiment Analysis, Social Media

1. Introduction

In our increasingly interconnected digital world, the analysis of sentiment within text data plays a pivotal role in understanding and responding to human emotions, opinions, and trends. The microblogging platform Twitter, renowned for its rapid and dynamic generation of user-generated content, represents a valuable repository of real-time textual data that encapsulates the diverse spectrum of human sentiments [1]. The area of Sentiment Analysis intends to comprehend these opinions and distribute them into the categories like positive, negative, neutral [2]. In this research endeavour, we introduce a novel approach that seamlessly amalgamates the power of machine learning with web-based interactivity to provide a user-friendly platform for instant sentiment analysis. Our primary objective is to empower individuals and organizations to effortlessly discern the emotional connotations embedded within any given sentence or text snippet, with a specific emphasis on harnessing the unique characteristics of Twitter data. While Sentiment Analysis is widely practiced getting the insight about the feelings and attitude of the people on the internet; it is done to measure the performance of various brands, products etc [3]. Our research bridges this gap by offering a comprehensive framework for the development, deployment, and utilization of an efficient sentiment analysis model, tailored to the demands of Twitter's real-time and succinct textual content. This paper delves into the intricacies of our web-based tool, elucidating the technical foundations that underlie its functionality. We provide insight into the pre-processing techniques, feature engineering methods, and machine learning algorithms meticulously crafted to enable the precise detection of sentiments, ranging from the rudimentary positive and negative polarities to the subtle nuances of human emotion. Through an intuitive user interface, our platform caters to both novices and experts, facilitating the rapid input and analysis of text,

ISSN PRINT 2319 1775 Online 2320 7876
Research paper **Volume 11, Sp.Iss 5 , 2022**

thereby democratizing real-time sentiment assessment. Furthermore, our research confronts the unique challenges presented by Twitter's idiosyncratic characteristics, including the brevity of tweets, the absence of extensive context, and the ever-evolving landscape of trending topics. Beyond individual users seeking to comprehend the emotional resonance of their tweets, our web-based platform holds the potential for a diverse array of applications, spanning brand sentiment analysis, marketing insights, and the real-time monitoring of social trends. The rationale for the choice of Twitter as a source of dataset collection lies in its ability to provide a huge number of tweets for the case study of this paper [5].

Sentimental analysis on Twitter data is becoming increasingly popular because to the data we obtain from movies, politics, and other sources, and many researchers use the Twitter API to gather tweets as a collection of data or to obtain datasets from platforms such as Kaggle [6].

Negative emotions are denoted by the sign ":(" and positive emotions by the symbol ":)" and ":-(". The Navi Bayes method serves as the model. The report and the text are created and shown for analysis. Authors in paper [7] presented and explained the hybrid approach using both corpus based and dictionary-based methods to determine the semantic orientation of the opinion words in tweets. Authors in paper [8] collected tweet data via the Twitter API for a while, and they used the data to perform sentiment analysis to gain the opinion polarity of the folks concerning general elections held in India. Authors in paper [9] designed a n approach to determine if the tweet is sarcastic or not, they have employed support vector machines, decision trees, Naïve Bayes, and k-nearest neighbors. And compared to other models, the SVM model has shown better outcomes with 93%. Authors in paper [10] used Hadoop Framework and MapReduce architecture for processing the dataset. Naïve Bayes is used for implementing the model. Authors in paper [11] compared various Machine Learning methods like the Naïve Bayes Classification method, Support Vector Machine Classification Method, and Maximum Entropy Classification method. The accuracy of Naïve Bayes, SVM and Maximum Entropy are 86%,74% and 83%.

2. Methodology

Figure 1 represents the workflow of the proposed work from data set collection to result analysis. Sentiment analysis model is created using Navie Bayes algorithm with a web-based approach. A framework for integrating a machine learning model with Django is shown in Figure 2. The model is trained and saved using joblib before being loaded into the Django view.py file. A html form is made to collect tweets from users and present the results based on whether the tweet is positive, negative, or neutral. The Twitter dataset was obtained from Kaggle and has 4 columns. Using feature extraction, we only consider the "selected_text" column and the "sentiment" column. The dataset contains 27.5k tweets with sentiment. The percentage of tweets comes under positive, negative and neutral is given in figure 3. The keywords found in the "text" column are included in the “selected_text" column. The number of positive, negative, and neutral forms of "selected_text" is shown in Figure 4.

ISSN PRINT 2319 1775 Online 2320 7876
Research paper Volume 11, Sp.Iss 5 , 2022

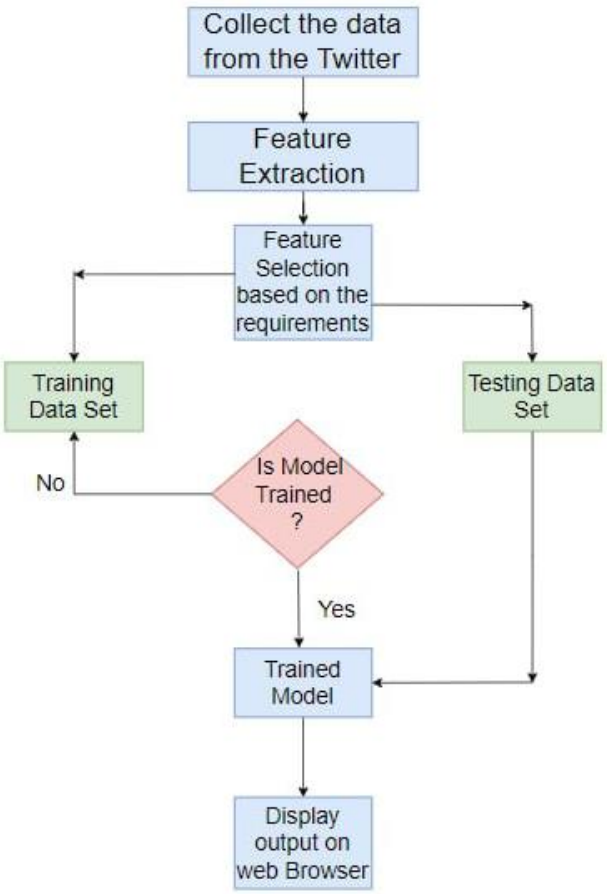

Figure 1 Workflow of the Proposed Work

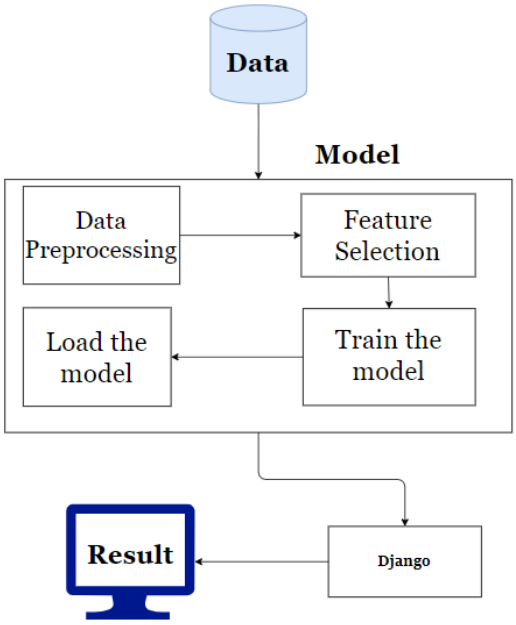

Figure 2 Framework of the Proposed Work

ISSN PRINT 2319 1775 Online 2320 7876
Research paper **Volume 11, Sp.Iss 5 , 2022**

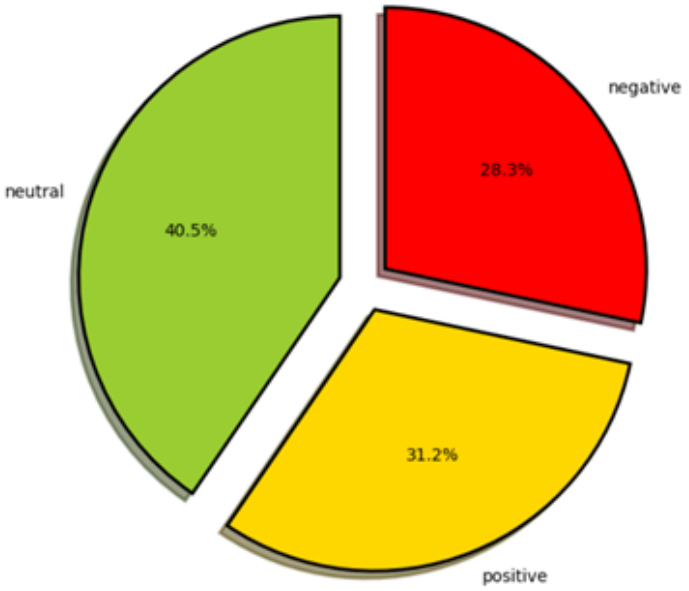

Figure 3 Distribution of Sentiments in dataset

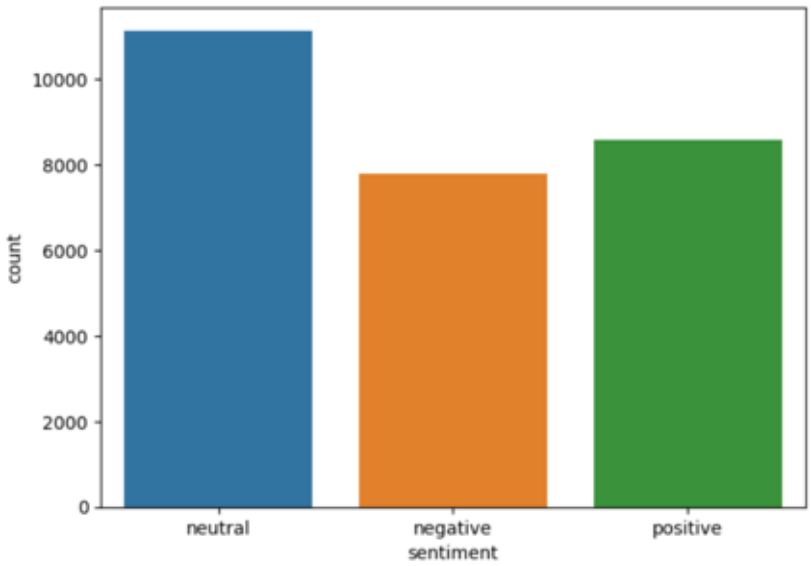

Figure 4 Category wise number of tweets in dataset

Naïve Bayes (NB) is quick and performs well even with small datasets. It is a probabilistic classifier and applies Bayes statement with strong independent hypotheses [12].

Classification approach:

Input: Dataset

Output: Predicted result
BEGIN
Divide dataset into two parts as training, testing sets and four classes with sets of features.
Mean and standard deviation calculation on each class features
Calculate summary for each feature of a class
Calculate the probabilities of features using normal distribution
Calculate probability of each class by multiplying probabilities of its features
In test dataset, the instance class is predicted by calculating probability.
END

3. Results and Discussion

We utilized a Twitter dataset for this project because of its unique qualities, such as brevity and the use of hashtags, mentions, and slang. To clean and prepare the data for analysis,

ISSN PRINT 2319 1775 Online 2320 7876
Research paper **Volume 11, Sp.Iss 5 , 2022**

preprocessing procedures were used. To analyse the performance of the Nave Bayes algorithm, the dataset was divided into two parts for training and testing. We made predictions on the test set after training the model and evaluated its performance. The model's accuracy was determined, providing a measure of how effectively it could classify tweets as positive, negative, or neutral. Word cloud for positive tweets is shown in figure 5 and word cloud for negative tweets is given in figure 6.

Figure 5 Word cloud for positive tweets

Figure 6 Word cloud for negative tweets

After building a model that categorize the sentiment of the given text/tweet as either positive, negative, or neutral. We integrated the model with the web application we created which is shown in figure 7. The model is trained well using the training part of the tweets. For testing the performance of the implemented model, a sample tweet is feed to web application and the model is run on the tweet. It classifies the sentiment as either positive, negative or

ISSN PRINT 2319 1775 Online 2320 7876
Research paper Volume 11, Sp.Iss 5 , 2022

neutral and returns the result to user. The accuracy of the model is shown in figure 7. Table 1 represents tweets and its classification report.

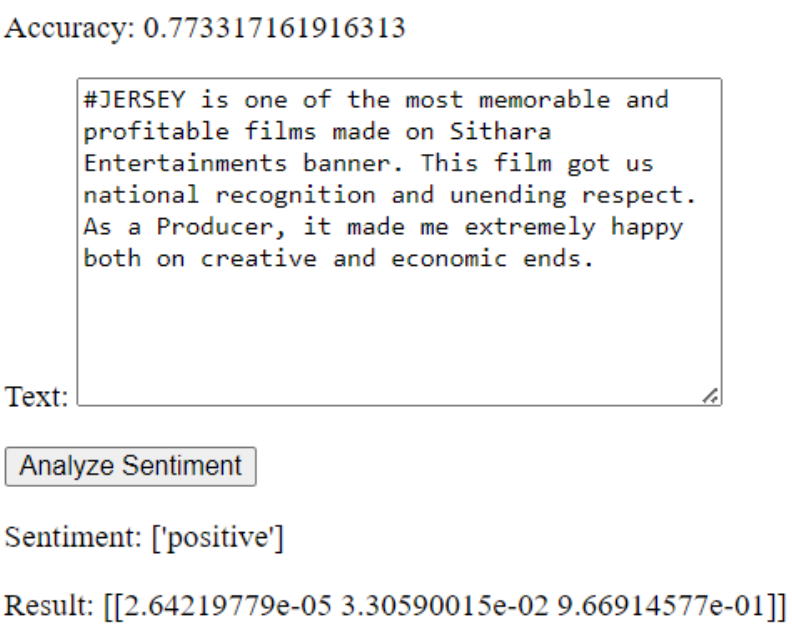

Figure 6 Tweet classification at web

Table 1 Tweets and its classification report

Sno	Feeded Tweet	Result
1	#JERSEY is one of the most memorable and profitable films made on Sithara Entertainments banner. This film got us national recognition and unending respect. As a producer, it made me extremely happy both on creative and economic ends.	positive
2	I am sorry I did a mistake, I won't repeat it	negative
3	Megha is a poor girl. She is not able to join at good college. That's why she missed job opportunity at MNC company	negative
4	My mom prepared my favorite food today. I have to go home early.	positive
5	APRET notification is released in the month of October, 2023.	neutral

4. Conclusions

Nave Bayes approach is used in this work and done sentiment analysis using Twitter dataset. The research enabled us to categorize tweets as positive, negative, or neutral, allowing us to gain a deeper grasp of the emotional content on this social media site. The results revealed that, despite its simplicity, the Nave Bayes algorithm can give reliable sentiment classification for tweets. The accuracy score of the project and the metrics in the classification report proved the model's effectiveness in classifying attitudes inside tweets. The project's findings have a variety of practical ramifications, particularly in the field of social media analytics. Businesses and organizations can utilize sentiment analysis on Twitter to track brand perception and customer feedback.

In future work, improvements can be made by exploring more advanced machine learning models, considering context and sarcasm in tweets, and adapting the approach to handle real-time sentiment analysis. Nonetheless, this project represents a significant step towards

ISSN PRINT 2319 1775 Online 2320 7876
Research paper **Volume 11, Sp.Iss 5 , 2022**

harnessing the power of sentiment analysis in the dynamic landscape of social media, and it offers valuable insights into the emotions expressed by Twitter users which can be known through a better web-based technology.

Acknowledgements

Sincere thanks to the management of KL University for providing good infrastructure to do extensive research.

References

1. Pak, A., & Paroubek, P. (2010, May). Twitter as a corpus for sentiment analysis and opinion mining. In LREc (Vol. 10, No. 2010, pp. 1320-1326).
2. SEBASTIAN, T. (2012). Sentiment analysis for Twitter (Doctoral dissertation).
3. Sharma, A., & Ghose, U. (2020). Sentimental analysis of twitter data with respect to general elections in India. Procedia Computer Science, 173, 325-334.
4. Dehkharghani, R., Mercan, H., Javeed, A., & Saygin, Y. (2014). Sentimental causal rule discovery from Twitter. Expert Systems with Applications, 41(10), 4950-4958.
5. Parveen, H., & Pandey, S. (2016, July). Sentiment analysis on Twitter Data-set using Naive Bayes algorithm. In 2016 2nd international conference on applied and theoretical computing and communication technology (iCATccT) (pp. 416-419). IEEE.
6. Mandloi, L., & Patel, R. (2020, June). Twitter sentiments analysis using machine learninig methods. In 2020 international conference for emerging technology (INCET) (pp. 1-5). IEEE.
7. Goel, A., Gautam, J., & Kumar, S. (2016, October). Real time sentiment analysis of tweets using Naive Bayes. In 2016 2nd International Conference on Next Generation Computing Technologies (NGCT) (pp. 257-261). IEEE.
8. Vadivukarassi, M., Puviarasan, N., & Aruna, P. (2017). Sentimental analysis of tweets using Naive Bayes algorithm. World Applied Sciences Journal, 35(1), 54-59.
9. Sahayak, V., Shete, V., & Pathan, A. (2015). Sentiment analysis on twitter data. International Journal of Innovative Research in Advanced Engineering (IJIRAE), 2(1), 178-183.
10. Hazra, S. Mandal, and A. Gupta" Study and Analysis of Breast Cancer Cell Detection using Naïve Bayes, SVM and Ensemble Algorithms," International Journal of Computer Applications. 2016, vol. 145, no.2, pp. 0975 – 8887.