

Sentimental Analysis in Twitter Using Python

Rukmini Ankush Dhamdhere^{1*}, Nikita Makarand Dhumal²,
Mayuri Keshav Gawali³, Ganesh Pramod Kulkarni⁴

Abstract

Social media websites are a great source of information because they have a lot of data. For instance, Twitter generates millions of packets of text. These statistics may be employed for commercial or charitable purposes. One of the hottest new buzzwords for many business strategies is the analysis of data from these social networking websites. Sentimental analysis can be used to manage election campaigns, global health problems, technical concepts, inventions, entertainment, and natural resource issues. Using Stanford NLP Libraries implemented in SaaS (cloud), which will manage all global current affairs, our proposed study assesses sentimental analysis of Twitter data. Implementing the cloud will improve speed to market, result growth, and process efficiency.

Keywords: Sentimental analysis, natural language processing, Twitter4j, NLP, JSON

INTRODUCTION

Social networks are extremely important to civilization today. Twitter and other social networking sites provide a forum for people to express their opinions on a range of subjects, current events, and global affairs. Every day, social media platforms like Twitter make millions of kilobytes of data [1].

It is essential to analyze such an overwhelming amount of data to learn exactly what the opinions of individuals are on various issues. To find out peoples' viewpoints, the emotional analysis method is employed. Business groups employ this strategy to gather feedback on their goods so they can choose their marketing tactics.

Twitter uses the # symbol, also known as a hashtag, to classify communications. Users of Twitter are permitted to post tweets that are just 140 characters long (updated to 280 characters).

The majority of the tweets consists of only words, extra punctuation, slang, and emoticons. Twitter is a key tool for expressing one's views about an incident on social media. The expression of happiness and melancholy can serve as a gauge of support for or opposition to particular laws [2].

The technique used to gauge public opinion on social networking platforms like Twitter is sentiment analysis.

*Author for Correspondence

Rukmini Ankush Dhamdhere
E-mail: rukminidhamdhere.nbnssoe.it@gmail.com

¹⁻⁴Student, Department of Information Technology, NBN
Sinhgad School of Engineering, Pune, Maharashtra, India

Received Date: June 07, 2023
Accepted Date: June 23, 2023
Published Date: July 06, 2023

Citation: Rukmini Ankush Dhamdhere, Nikita Makarand Dhumal, Mayuri Keshav Gawali, Ganesh Pramod Kulkarni. Sentimental Analysis in Twitter Using Python. International Journal of Computer Science Languages. 2023; 1(1): 28–32p.

LITERATURE REVIEW

The majority of scholars have looked into various methods and algorithms for categorizing sentiment analysis on Twitter. There are several methods for conducting opinion mining on twitter data.

Using information from Twitter about chemotherapy, Ling Zhang, Magie Hall, and Dhundy Bastola [3] have examined it. By searching the accounts using various cancer-related keywords including “cancer,” “leukemia,” “tumor,”

“lymphoma,” and “carcinoma,” they were able to extract Twitter accounts (person accounts and organization accounts) relevant to cancer and their tweets.

Support Vector Machine (SVM), Particle Swarm Optimization (PSO), and Genetic Algorithm (GA) approaches are used by Kristiyanti et al. [4] to identify the sentiments of the 2019 Indonesian presidential election. The outcome demonstrates that the SVM combination of PSO is the best for classifying data with an accuracy of 76.20% and a kappa value of 0.934.

This study shows how social media can be a useful source of healthcare data for determining how the general public discusses health-related issues. This research may also be used to help patients create individualized treatment programs. A precise approach for doing sentiment analysis on tweets concerning reviews of Bollywood and Hollywood films has been proposed by Akshay Amolik, Niketan Jivane, Mahavir Bhandari, and Dr. M. Venkatesan [2]. To correctly classify a tweet as positive, negative, or neutral, they used classifiers and feature-vectors like SVM (support vector machine) and Naive-Bayes. This study demonstrates that Naive Bayes has somewhat lower accuracy and recall than SVM, but superior precision. SVM outperforms Naive Bayes in terms of accuracy, precision, and recall [5].

METHODOLOGY

Fetching of Tweets

Unlike usual methods of fetching tweets from code written in different languages, we are making use of TAGS application [6], which is a free source application to fetch tweets from twitter using hashtags or usernames.

Data Cleaning

Here, we are utilizing the Anaconda package, which includes the Jupyter Notebook program. It is the program where we execute our Python commands. Data cleaning involves eliminating unnecessary symbols and stopping words that hinder the efficient processing of the words and make it harder to extract sentiment from tweets [7].

- *Removing RT:* Re-Tweets need to be eliminated in order to stop counting tweets with the same sentiment over and over again.
- *Removal of unwanted patterns:* Unwanted patterns include user names, URL links, HTML elements, and hashtags that don't accurately reflect the sentiment of the tweets.
- *Cleaning of punctuations:* In this step, Jupyter Notebook's data frames and regex expressions are used to remove any punctuation or symbols.
- *Cleaning of stop words:* Stop words, such as “is,” “are,” “the,” “a,” “an,” “and,” assist make sentences grammatically correct. by utilizing the NLTK (Natural Language Tool Kit) library, are eliminated from the tweets.

Tokenization and Stemming

By breaking the text up into tokens that are stored in an array, the sentence is broken up into individual words [8]. Stemming, which involves removing additional suffixes like-ing and-ed, is the process of determining the root word from a derivative word as shown in Figure 1.

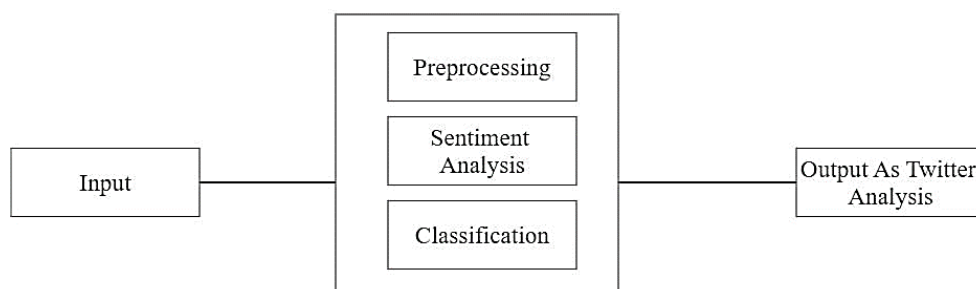


Figure 1. Data flow diagram.

Labelling and Calculation of Polarity

To classify the sentiments in the tweet, the polarity of the message must be calculated. This is accomplished by utilizing a text bob library, which accurately foretells the sentiment of each tweet [9]. The labelling is done as positive, negative, or neutral based on this polarity value, and it can be provided to the machine learning algorithms to be properly trained.

Word List Creation

By establishing an instance of the counter class, a list of the most common unique terms and their total number of occurrences in the entire data set is created. In the data sets of various Indian government programs, some of the most frequently occurring words are represented visually as shown in Figure 2.

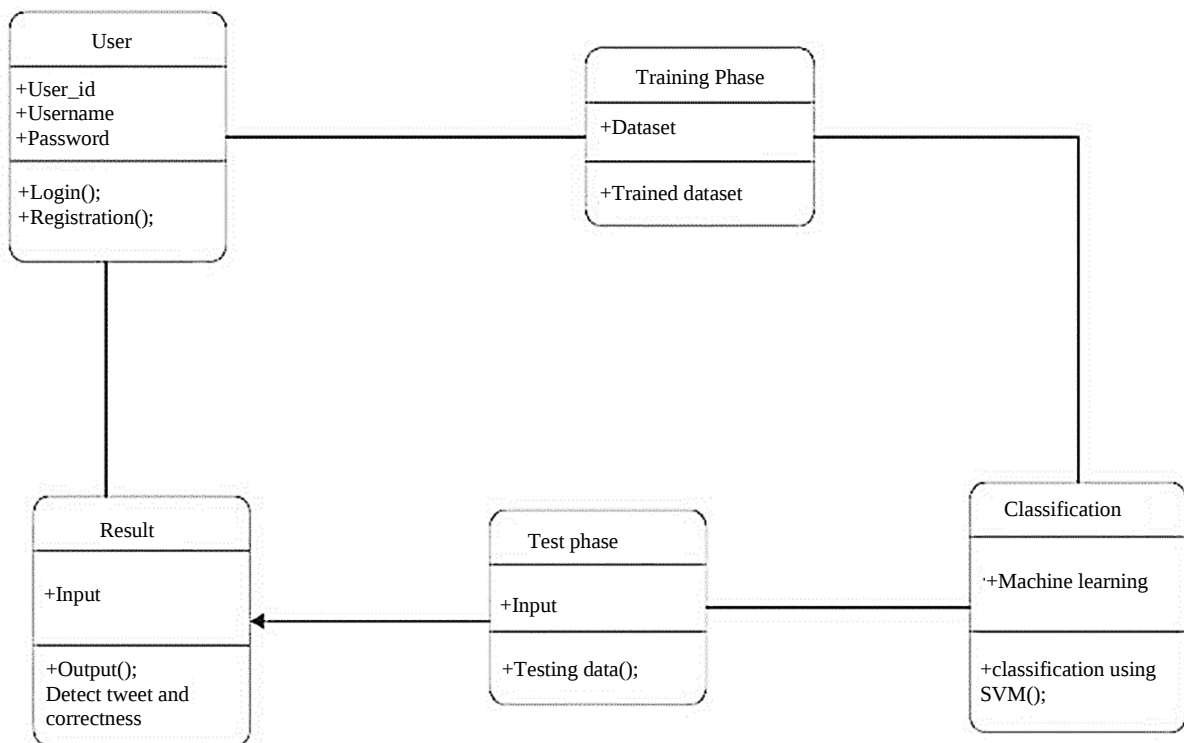


Figure 2. Class use case diagram.

Bag of Words

The word list that was previously constructed is utilized to create a bag of words model, which consists of tuples with the sentiment category and attributes with the word list's properties. In our bag-of-words model, we use the binary weight representation, which is appropriate for the machine learning techniques we have taken into consideration [10]. Bernoulli Naive Bias is one of the algorithms, and it uses binary data for training with features and labels. The other technique under consideration builds several binary decision trees with characteristics as nodes and labels as leaf nodes using this approach. a tabular excerpt from the bag of words [11].

Bernoulli naive Bayes: This classifier uses binary input, or Boolean values that define the features or attributes that are involved, to classify events that involve several variables. This model is mostly recommended for jobs involving document categorization if the attributes suggest the usage of binary terms as opposed to term frequencies.

Random forests: Decision trees serve as the foundation for random forests. This is employed in regression, classification, and other tasks as a collective learning source that aggregates decision trees built during the training phase and produces the class as an output (classification), or computes the mean

prediction (regression), for each tree. Decision forests generated randomly are a superior alternative to decision trees since they overfit to their training data [12].

Tailored random forest: is a random forest that has two additional parameters set. The number of trees in the forest, `n_estimators`, and the maximum number of features, `max_features`, should both be taken into account when determining the optimal split.

RESULT

In data sets of various programs that were prepared and analyzed through machine learning, statistical representations of positive, negative, and neutral sentiments were found as shown in Figure 3.

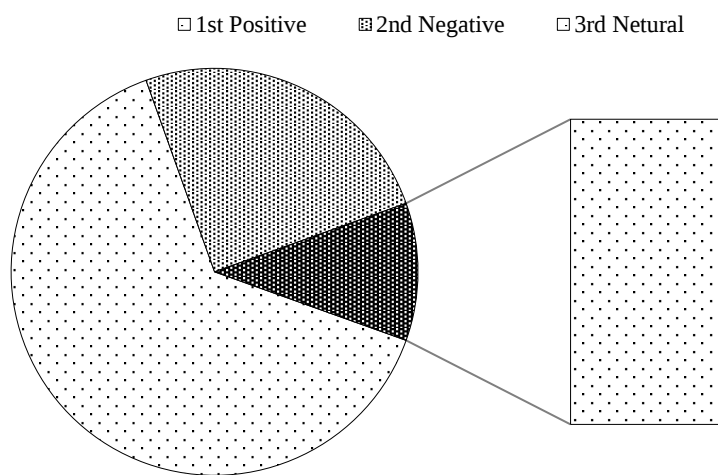


Figure 3. Pie chart of sentiments visualization.

The NLP program produces an integer value that indicates how favorable or negative the tweet was. The sentimental analyzer package is employed in this case to produce better outcomes [13]. The primary goal of the suggested effort is to improve results, which is the main aim of the proposed work.

CONCLUSION AND FUTURE WORK

It can be deduced from the proposed study that greater attempts were made to conduct the inquiry on a sizable volume of data and to analyze tweets in order to forecast the cluster's quality, whether it be favorable, negative, or neutral. For analysis, numerous algorithms are employed. To ascertain the person's behavior on Twitter, real-time data were gathered from the service. In the future, a website may be created where users could log in and enter the Twitter search term. Create a connection to Twitter later to gather real-time data and automate data science processes, like sentiment analysis visualization.

Acknowledgments

It is indeed a great pleasure and moment of immense satisfaction for we to present project report on "SENTIMENTAL ANALYSIS IN TWITTER USING PYTHON" among a wide panorama that provided us inspiring guidance and encouragement, we take the opportunity to thanks to thanks those who gave us their indebted assistance. We wish to extend our cordial gratitude with profound thanks to our internal guide Prof. K.M. Karthickbabu for her everlasting guidance. It was his inspiration and encouragement which helped us in completing our project. Our sincere thanks and deep gratitude to Head of Department, Prof. Dr. R.M. Samant and other faculty members and also to all those individuals involved both directly and indirectly for their help in all aspect of the project. Last but not least, we express our sincere thanks to our Institute's Principal Dr. Shivprasad P. Patil, for providing us infrastructure and technical.

REFERENCES

1. Hu Y, John A, Wang F, Kambhampati S. Et-lda: joint topic modeling for aligning events and their twitter feedback. AAAI. Proceedings of the AAAI conference on artificial intelligence 2012; 26 (1): 59–65. doi: 10.1609/aaai.v26i1.8106.
2. Amolik A, Jivane N, Bhandari M. Dr MV. Twitter sentiment analysis of movie. Int J Eng Technol. 2016; 7 (6): 2038–2044.
3. Zhang L, Hall M, Bastola D. Utilizing Twitter data for analysis of chemotherapy. Int J Med Inform. 2018; 120: 92–100. doi: 10.1016/j.ijmedinf.2018.10.002, PMID 30409350.
4. Kristiyanti DA, Wahyudi M. Feature selection based on Genetic algorithm, particle swarm optimization and principal component analysis for opinion mining cosmetic product review. Proceedings of the 5th International Conference on Cyber and IT Service Management (CITSM); 2017 Aug 8; Denpasar, Indonesia: IEEE Publications; 2017. 1–6 pp. doi: 10.1109/CITSM.2017.8089278.
5. Zvarevashe K, Olugbara OO. A framework for sentiment analysis with opinion mining of hotel reviews. Proceedings of the 2018 Conference on information communications technology and society (ICTAS); 2018 Mar 8; Durban, South Africa: IEEE; 2018. 1–4 pp.
6. Ahuja S, Dubey G. Clustering and sentiment analysis on Twitter data. Proceedings of the 2017 2nd International Conference on Telecommunication and Networks (TEL-NET); 2017 Aug 10; Noida, India: IEEE; 2018. 1–5 pp.
7. Gupta B, Negi M, Vishwakarma K, Rawat G, Badhani P, Tech B. Study of Twitter sentiment analysis using machine learning algorithms on Python. Intl J Comp Appl. 2017; 165 (9): 29–34.
8. Farghaly A, Shaalan K. Arabic natural language processing: challenges and solutions. ACM Transactions on Asian Language Information Processing (TALIP). 2009; 8 (4): 1–22.
9. Do TN, Poulet F. Parallel learning of local SVM algorithms for classifying large datasets. In: Transactions on Large-Scale Data-and Knowledge-Centered Systems XXXI: Special Issue on Data and Security Engineering 2017 Berlin, Germany: Springer; 2017. pp. 67–93.
10. Cao N, Shi C, Lin S, Lu J, Lin YR, Lin CY. Targetvue: visual analysis of anomalous user behaviors in online communication systems. IEEE Trans Vis Comput Graph. 2015; 22 (1): 280–289.
11. Antinasari P, Perdana RS, Fauzi MA. Analisis sentimen tentang opini film pada dokumen twitter berbahasa indonesia menggunakan naive bayes dengan perbaikan kata tidak baku. Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer. 2017; 1 (12): 1733–1741.
12. Huq MR, Ahmad A, Rahman A. Sentiment analysis on Twitter data using KNN and SVM. Int J Adv Comput Sci Appl. 2017; 8 (6): 19–25.
13. Kharde V, Sonawane P. (2016). Sentiment analysis of twitter data: a survey of techniques. [Online] arXiv preprint. Available at <https://arxiv.org/abs/1601.06971> [Accessed on July 2023]