**Social Media Sentiment Analysis for social media to differentiate the nature of the user using Machine Learning**

Archis Ramesh Khuspe

*Department of Electronics and Telecommunication (E&TC)*

*Vishwakarma Institute of Technology*

*Pune, India*

archis.khuspe20@vit.edu

***Abstract—Sentiment analysis pacts with recognising and organising views and opinions articulated in the original text. Social media generates a massive expanse of data in updates, tweets, and sentiment-rich posts. Sentiment study of this data created by the users is vital in recognising the overall judgement of the userbase, analysing conversations, and sharing views which can be implemented in determining commercial tactic, political study, and calculating community activities. Twitter sentiment analysis is tougher than overall sentiment examination due to the prevalence of misspellings, dialect words, symbols, and emoticons. This paper presents the analysis of the Twitter posts of a particular account using Python alongside Machine Learning. By carrying out sentiment study in a particular area, it is likely to classify the consequence of that area’s data in sentiment cataloguing. This paper presents a feature for organising a user’s most recent tweets and visualising them using graphs, charts, and word clouds.***

***Keywords—twitter, sentiment, dataset, accounts, algorithms, tweets***

**I. INTRODUCTION**

The method widely practised in-text quarrying is Sentiment Analysis. Utilising innovative text mining procedures, the emotion of the tweet is examined in the figure of positive, negative or neutral. R, as well as Python, are widely used in Twitter sentiment breakdown. Its functional role in the industries include:

*Business:* In marketing, various organisations use it to advance their policies, recognise customers’ opinions towards their brand and products, their response to the company’s promotions or product launches and why they do not seem to purchase some of their products.

*Public activities:* Sentiment analysis is also used to monitor and examine social sensations to detect possibly dangerous circumstances and decide the community’s overall mood.

*Politics:* Sentiment Analysis is used to understand the general public’s political views and spot the evenness and irregularity between the declarations and dealings made by the government. It can also be used to predict the results during the elections.

**II. LITERATURE REVIEW**

[1] This paper aims to implement the RNN-LSTM on the Twitter database to categorise individuals ‘views into optimistic and rejecting and compare the results of correctness with several machine learning algorithms. [2] This research focuses on evaluating the people’s sentiments and emotions during the COVID-19 pandemic. [3] DICET acquires and matches the data features and obtains essential features instead of using the one-word depiction model. [4] The study has implemented sentiment examination on Twitter data associated with the COVID-19 pandemic. [5] This research presents a method to manage aspect decrease using arithmetic study and n-grams to develop a lexicon set for Twitter outlook examination. [6] This paper introduces new clustering algorithms to analyse tweets’ sentiments based on K-means and DENCLUE. [7] In this paper, a QSR paradigm has been proposed to demonstrate the semantic data of private documents. [8] This composition studies a Twitter sentiment analysis example to initiate quick verdict production in the FTSE stock sphere with forecast abilities. [9] In this study, the VADER and the NLTK are used to perform a Twitter reaction study and classify tweets. [10] The researchers scrutinised many pre-processing methods and tested them in two datasets. [11] This work analyses sentiment dispersion on Twitter by studying emotional reversals, working on its prediction model, and designing SentiDiff, an innovative Twitter sentiment cataloguing procedure. [12]. An opinion analysis system is constructed by learning and applying machine learning processes using Maximum Entropy and Naive Bayes. [13] In this paper, a classification system has been proposed, which is presumed to advance the functioning of sentiment classification. [14] In this paper, the sentiment analysis establishes the model’s efficiency in social media utilising the standard approaches towards accuracy, recall, precision, and F1. [15] In this paper, the safety of the protected vital swap system by KLJN has been researched. [16] The paper reveals two reasons made for TSA research. The first is applying TSA to gain the information of several business and social matters and predict analytical indicators. The second is developing enhanced methods and tactics for TSA. [17] This project has applied a multistage hybrid classification scheme with a united outline to spot and categorise sentiments articulated by the operators in tweets. [18] The study has applied a framework that sees and categorises sentiments stated by Twitter consumers linking towards a commodity. [19] Using various machine learning methods, this study has explained the Twitter sentiment study of the information concerning ordinal reversion. [20] A penetration density neuronic system for Twitter outlook analysis is used in this work.

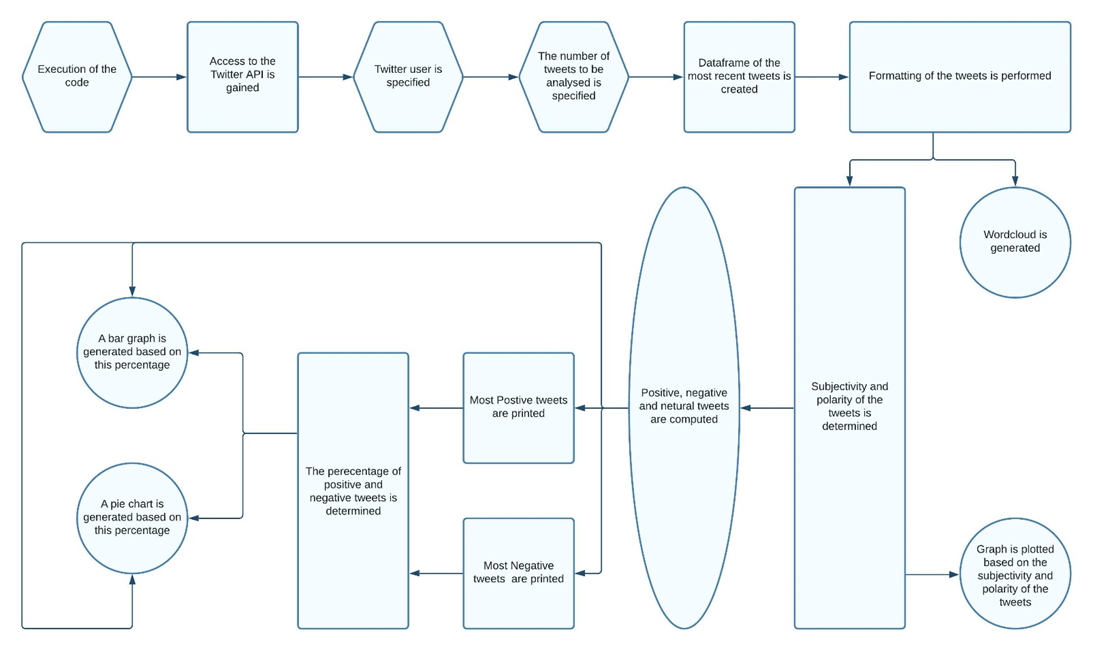
**III. PAPERS REVIEWED**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **S No.** | **Year** | **Learning techniques & Algorithms used** | **Datasets** | **Summary** |
| 1. | 2021 | Deep Neural Network  RNN-LSTM algorithm  Naive Bayes  Entropy  Support Vector Machine  WordNet  Decision Tree  K- Nearest Neighbour | IMDB, Amazon, and Airline datasets | This paper aims to implement the Twitter dataset with the RNN-LSTM to categorise the public’s views into definite and undesirable and compare the accuracy outcomes with several machine learning algorithms. |
| 2. | 2020 | Sentiment Analysis algorithms | Twitter API | This research evaluates people’s sentiments and emotions during the COVID-19 pandemic. |
| 3. | 2020 | SentiStrength, a lexicon-based algorithm  Viterbi algorithm  Metaphone algorithm | SemEval-2013 evaluation campaign.  US airlines dataset  Emirates airlines datasets  Twitter API library | DICET acquires and matches the data features and obtains essential structures as a replacement for operating the one-word depiction example. |
| 4. | 2020 | Naïve Bayes algorithm  Lexicon-based algorithms  Latent Dirichlet Allocation (LDA) | Text blob dataset  Tweepy python library  Twitter API | The study has implemented the Coronavirus pandemic emotion study on information acquired from Twitter. |
| 5. | 2020 | DAN2  Naïve Bayes  kNN  Logistic Regression  artificial neural networks (ANN)  SVM | Starbucks dataset  Verizon dataset  @GovChristie dataset  Southwest Airlines Dataset | This research presents a method to manage aspect decrease using arithmetic study and n-grams to develop a lexicon set for Twitter outlook examination. |
| 6. | 2019 | Clustering algorithm  Cuckoo search algorithm.  K-means algorithm  DENCLUE algorithm  Hill climbing algorithm  EM algorithm | Sanders’s dataset  Twitter dataset | This paper introduces new clustering algorithms to analyse tweets’ sentiments based on K-means and DENCLUE. |
| 7. | 2019 | Supervised and unsupervised learning  Novel image sentiment prediction  Globally convergent algorithm  Quantum-inspired Sentiment Representation (QSR) model | Information retrieval (IR) techniques  Obama-McCain Debate (OMD) dataset  Sentiment140 Twitter dataset | A QSR paradigm has been proposed to demonstrate the semantic data of private documents. |
| 8. | 2019 | Machine learning algorithms | Historical modelling datasets | This composition studies a Twitter sentiment analysis example to initiate quick verdict production in the FTSE stock sphere with forecast abilities. |
| 9. | 2019 | Naive Bayes  Maximum Entropy  Support Vector Machine (SVM) | Stanford University dataset | In this study, the VADER and the NLTK are used to perform a Twitter reaction study and classify tweets. |
| 10. | 2018 | Generalised Linear Models (GLM  Naïve Bayes (NB),  Support Vector Machines (SVM)  Neural Networks (NN)  Logistic Regression (LR).  Linear SVC (LSVC). | Stanford Twitter Sentiment Dataset  SemEval dataset  Sentiment Strength Twitter dataset | This study has scrutinised many pre-processing methods and tested them in two datasets. |
| 11. | 2018 | SentiDiff algorithm | Beijing Intelligent Starshine Information Technology Corporation dataset | This work analyses sentiment dispersion on Twitter by studying emotional reversals, working on its prediction model, and designing SentiDiff, an innovative Twitter sentiment cataloguing procedure. |
| 12. | 2018 | Naive Bayes  Maximum Entropy or SVM  Iterative Scaling algorithm  natural language processing (NLP)  Baseline (Evaluation Metric) | Opinion Dataset  Kaggle datasets | An opinion analysis system is constructed by learning and applying machine learning processes using Maximum Entropy and Naive Bayes. |
| 13. | 2018 | Sentiment Analysis algorithms | Stanford dataset  HCR dataset  GOP debate dataset | In this paper, a classification system has been proposed, which is presumed to advance the functioning of sentiment classification. |
| 14. | 2018 | ConSent (Context-based Sentiment Analysis)  SentiStrength  Bayes’ rule  Decision Tree  K-Nearest Neighbours | SemEval-2016 | In this paper, the sentiment analysis establishes the model’s efficiency in social media utilising the standard approaches towards accuracy, recall, precision, and F1. |
| 15. | 2018 | Machine learning algorithms | Twitter API | In this paper, the safety of the protected vital swap system by KLJN has been researched. |
| 16. | 2018 | Entropy-weighted genetic algorithm (EWGA)  SVM machine-learning algorithm  Machine-learning algorithm | Weka data mining package  TSA applications  Telco, Pharma, and Security datasets  An emoticon-labelled Twitter data set | The paper reveals two reasons made for TSA research. The first is applying TSA to gain the information of several business and social matters and predict analytical indicators. The second is developing enhanced methods and tactics for TSA. |
| 17. | 2017 | Hybrid classification scheme  Slang, emoticon classifier  Domain-specific classifier  Enhanced sentiment classification  Pre-processing | Python Tweepy  SQL Server Database  Alchemy API | This project has applied a multistage hybrid classification scheme with a united outline to spot and categorise sentiments articulated by the operators in tweets. |
| 18. | 2017 | RST-centric rule induction  LEM2 and CBR  Classification  Association rule mining (ARM)  Exhaustive, genetic, and covering algorithm | Python Tweepy  SQL Server | The study has applied a framework that sees and categorises sentiments stated by users in tweets linking to a product. |
| 19. | 2017 | Multinomial logistic regression (SoftMax)  Support Vector Regression (SVR)  Decision Trees (DTs)  Random Forest (RF)  Multinomial naive (Bayes)  Ordinal regression  Perceptron ranking (PRank)  Binary classification  Multiple output CNN learning | Natural Language Toolkit (NLTK)  Asian Face Age Dataset (AFAD)  Twitter API | Using various machine learning methods, this study has explained the Twitter sentiment study of the information concerning ordinal reversion. |
| 20. | 2017 | Convolution algorithm  Classification algorithm | Stanford Twitter Sentiment Test (STSTd)  SE2014 dataset  Stanford Twitter Sentiment Gold (STSGd)  Sentiment Evaluation Dataset (SED)  Sentiment Strength Twitter dataset (SSTd) | A penetration density neuronic system for Twitter outlook analysis is used in this work. |

**IV. METHODOLOGY**

*a. Implementation*

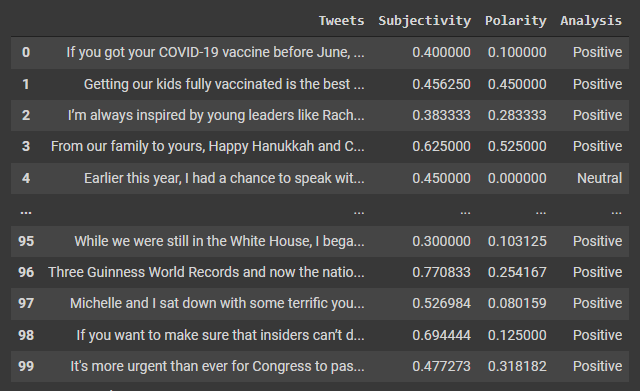
All the essential Python libraries required for the program are applied. An API object is created to add and authenticate Twitter API credentials provided by the Twitter developers forum. The user then enters the Twitter handle of the account required and the number of recent tweets to extract from the dataset. A data frame is generated to show 5 of the most recent tweets to verify that the correct account is analysed. The text is then formatted to remove any @mentions, symbols, and hyperlinks that may be present in those tweets.



*Fig. 1: Flowchart of the project*

*b. Classification and Analysis*

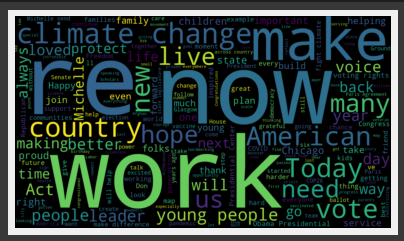
Using the text blob library, the function to obtain the subjectivity and polarity of the tweet is created. Subjectivity and polarity are defined with the help of numerical digits to comprehend their magnitude. A function is created to compute the positive, negative and neutral analysis from the subjectivity and polarity scores. Scores equal to zero are classified as neutral, above zero as positive, and below zero as unfavourable. The tweets that most inclines towards positive and negative sides are further segregated and printed. This segregation is done based on their scores obtained.



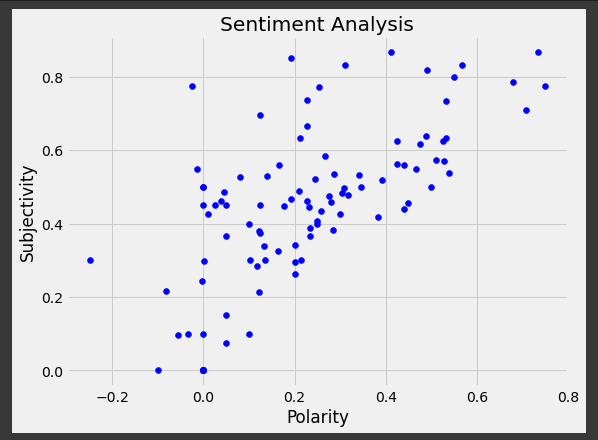
*Fig. 2: Table showing the subjectivity and polarity of the tweets*

*c. Visualisation*

A word cloud is generated, a data visualisation technique used to indicate text data in which the size of each word denotes its frequency or significance. The dimensions of the word cloud are specified using the word cloud library.

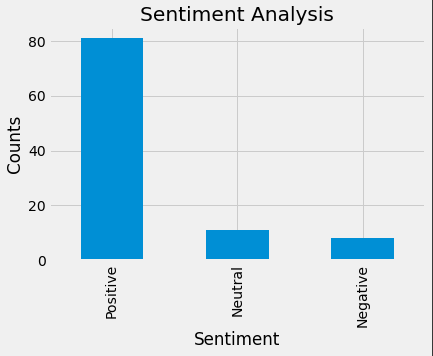


*Fig. 3: Word cloud showing the significant words used by the user*

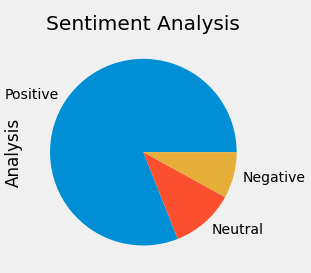


*Fig. 4: Graph showing the relative subjectivity and polarity*

Using the Matplotlib library, the tweets are plotted onto a graph according to their subjectivity and polarity. To further simplify this visualisation, the percentage of positive and negative tweets is obtained. This data is colour-coded and again plotted using the Matplotlib library in a bar graph and a pie chart displaying the amount of neutral, positive and negative tweets.



*Fig. 5: Bar Graph visualisation*



*Fig. 6: Pie Chart visualisation*

**V. RESULTS**

In this project, a sentiment analysis program has been developed using Python. The program would fetch tweets from the Twitter API and give out the following data.

1. Recent relevant tweets made by that user displayed after formatting
2. Subjectivity and polarity of their tweets and their visualisation
3. Most prominent words used by the user in the form of a word cloud
4. Top positive and negative tweets
5. Fraction of positive and negative tweets
6. Visualising the positive, negative and neutral tweets with bar graph & pie chart.

**VI. CONCLUSION**

Using advanced text mining procedures and the Python coding language, the emotion of the tweet is examined in the form of neutral, positive or negative.

Also known as opinion quarrying, the technique is widely used to analyse the discussions, views put forward in business planning, public activities, and politics.

Thus, the data generated by this sentiment analysis will be crucial in understanding the perspective and mindset of the masses, which has massive potential for the betterment of our society.

**VII. FUTURE SCOPE**

Implementing a presentable Graphical User Interface or GUI has been proposed to simplify the end-user experience. Being more user friendly, one would only be required to enter the Twitter user handle and the number of recent tweets to be analysed.

As mentioned before, the output would be presented using word cloud, data frames, and graphical visualisations. The addition of more types of visualisations can also be considered.

Changes to the algorithm are also under consideration; the motive will be to increase the accuracy further and provide the user with even more optimised results.

**VIII. ACKNOWLEDGEMENT**

I would like to express gratitude and sincerest regards to my guide, Prof. Rahul Diwate, who made this work possible. His direction and guidance helped me through all the stages of drafting this paper.

I would also like to acknowledge my family and friends for their constant support and understanding throughout my research.

**IX. REFERENCES**

[1] Twitter Sentiment Analysis with a Deep Neural Network by Yaser Maher Wazery, Hager Saleh Mohammed and Essam Halim Houssein, Year – 2021

[2] Twitter Sentiment Analysis during COVID19 Outbreak by Akash D. Dubey, Year – 2020

[3] Transformer based Deep Intelligent Contextual Embedding for Twitter Sentiment Analysis by Usman Naseem, Imran Razzak, Katarzyna Musial and Muhammad Imran, Year - 2020

[4] Twitter Sentiment Analysis on Worldwide COVID-19 Outbreaks by Kamaran H. Manguri, Rebaz N. Ramadhan and Pshko R. Mohammed Amin, Year – 2020

[5] A Domain Transferable Lexicon Set for Twitter Sentiment Analysis Using a Supervised Machine Learning Approach by M. Ghiassi and S. Lee, Year – 2020

[6] New Clustering Algorithms for Twitter Sentiment Analysis by Hajar Rehioui and Abdellah Idrissi, Year – 2019

[7] A quantum-inspired sentiment representation model for Twitter sentiment analysis by Yazhou Zhang, Dawei Song, Peng Zhang, Xiang Li and Panpan Wang, Year – 2019

[8] A Novel Twitter Sentiment Analysis Model with Baseline Correlation for Financial Market Prediction with Improved Efficiency by Xinyi Guo and Jinfeng Li, Year – 2019

[9] Twitter Sentiment Analysis Using Natural Language Toolkit and VADER Sentiment by Shihab Elbagir and Avid Jing Yang, Year – 2019

[10] A comparative evaluation of pre-processing techniques and their interactions for Twitter sentiment analysis by Symeon Symeonidis, Dimitrios Effrosynidi and Avi Arampatzis, Year – 2018

[11] SentiDiff Combining Textual Information and Sentiment Diffusion Patterns for Twitter Sentiment Analysis by Lei Wang, Jianwei Niu and Shui Yu, Year – 2018

[12] Twitter Sentiment Analysis System by Shaunak Joshi and Deepali Deshpande, Year – 2018

[13] An Ensemble Classification System for Twitter Sentiment Analysis by Ankit and Nabizath Saleena, Year – 2018

[14] Twitter Sentiment Analysis using Deep Neural Network by Ahmed Sulaiman M. Alharbi and Elise Doncker de, Year – 2018

[15] Generalised Attack Protection in the Kirchhoff-Law-Johnson-Noise Secure Key Exchanger by Gergely Vadai, Zoltan Gingl and Robert Mingesz, Year – 2018

[16] The State-of-the-Art in Twitter Sentiment Analysis A Review and Benchmark Evaluation by David Zimbra, Ahmed Abbasi, Daniel Zeng and Hsinchun Chen, Year – 2018

[17] Twitter sentiment analysis framework using a hybrid classification scheme by Muhammad Zubair Asghar, Fazal Masud Kundi, Shakeel Ahmad, Aurangzeb Khan and Furqan Khan, Year – 2017

[18] RIFT A Rule Induction Framework for Twitter Sentiment Analysis by Muhammad Zubair Asghar, Aurangzeb Khan, Furqan Khan and Fazal Masud Kundi, Year – 2017

[19] Twitter Sentiment Analysis based on ordinal Regression by Shihab Elbagir and Jing Yang, Year – 2017

[20] Deep Convolution Neural Networks for Twitter Sentiment Analysis by Zhao Jianqiang and Gui Xiaolin, Year – 2017