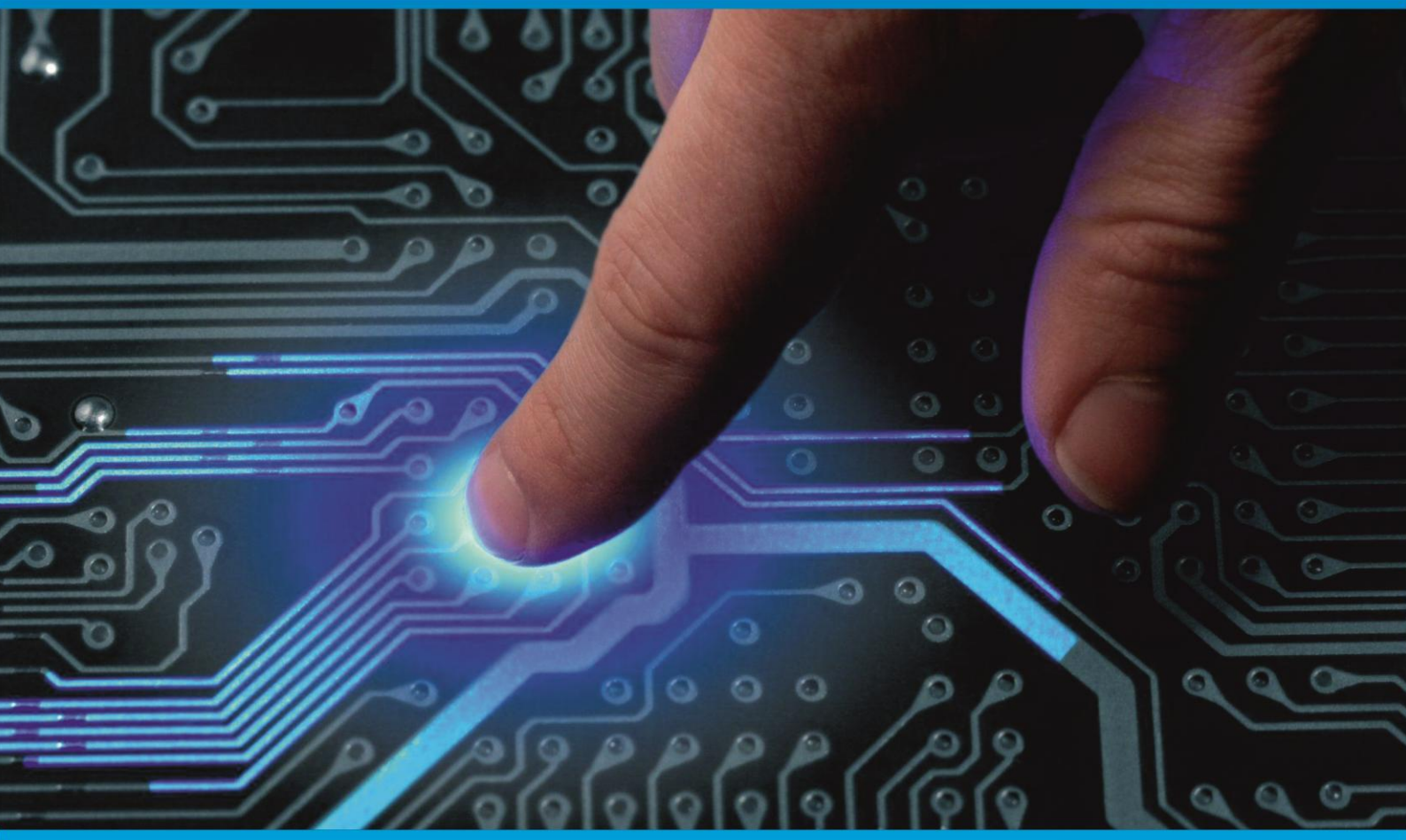
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 9, Issue 5, May 2021

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.488

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 5, May 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0905078 |

Twitter Sentiment Analysis Web App Using NLP

Atharva Kulkarni[1], Manas Mohite[2], Parin Dodhiya[3], Prof. Sukhada Aloni[4]

Student, Dept. of Computer Engineering, A. P. Shah Institute of Technology, Thane, Maharashtra, India[1]

Student, Dept. of Computer Engineering, A. P. Shah Institute of Technology, Thane, Maharashtra, India[2]

Student, Dept. of Computer Engineering, A. P. Shah Institute of Technology, Thane, Maharashtra, India[3]

Professor, Dept. of Computer Engineering, A. P. Shah Institute of Technology, Thane, Maharashtra, India[4]

ABSTRACT: Twitter Sentiment Analysis Web App is a project aimed to analyse different data by twitter users. Twitter is a social media app used widely by users across the globe. This project helps to understand the sentiment of the tweets posted by users. The "Web App" uses twitter API to collect real time tweets of users. Algorithms such as text processing are used to clean the data. Natural language toolkit (NLTK) libraries are used in the project to remove unnecessary words and characters. The machine learning (ML) model is created in order to clean the data and help improve the accuracy of predictions. The cleaned data is then processed further to analyse and predict the sentiment of the Tweet. The web app uses a graphical interface for the user to enter a topic. The app then fetches tweets related to that specific topic. The tweets consist of thoughts of many different users and may have different sentiments (positive, negative or neutral). Twitter Sentiment Analysis Web App can help the user to know the overall sentiments of tweets. Multinational brands often use sentiment analysis of user's feedback in order to improve quality of service or product. Our project helps to know the feedback of users on Twitter.

KEYWORDS: Machine learning, Natural Language, Sentiment Analysis, Twitter, GUI.

I. INTRODUCTION

Social media is widely used all around the globe by millions of users. Twitter has emerged among other apps, where users put forward their thoughts and emotions openly in a textual format. In fact, many companies often listen to their customers through 'Twitter'. Twitter is an excellent platform for reviewing products and services of a company. As the users in twitter are in large numbers, the company has to listen to their consumers and improve their services. The reception of users towards a particular product or a topic may be mixed. Some may like it and some may not. Our project is aimed at overcoming this problem.

Sentiment analysis is the technique to classify textual information into emotions (positive, negative and neutral). By applying sentiment analysis on tweets by users, we can predict the emotions of the users. This will help us to understand how the topic/product is perceived by users on twitter. As many companies need twitter data from users for analysis, Twitter created an API(Application Programming Interface) which fetches real time tweets.

The tweets collected can act as a dataset for the machine learning model. Now, the tweets are in an informal language and may contain unwanted words and special characters which are not needed. To remove them Natural Language Toolkit (NLTK) is used to process the dataset. A bag of words is created after this which contains some keywords like good, like, awesome which reflects positive feedback, whereas words like bad, terrible, failure etc. reflects negative feedback. According to the keywords the polarity of the tweet can be predicted; positive tweet's polarity is assigned 1 while the negative is assigned 0. To understand the sentiments of many users, polarity is displayed graphically in the form of a pie chart. The data visualization is created using a webapp. The web app is created using HTML and CSS. With the help of the web app users can search a trending topic or a recently released product and can see sentiments of users on twitter on that particular topic.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 5, May 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0905078 |

II. PROPOSED SYSTEM

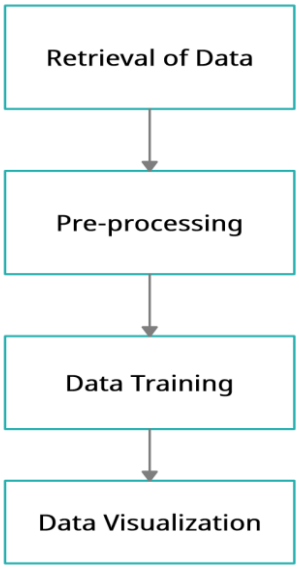


Fig.1. Workflow Diagram

- **Retrieval of Data:** Public Twitter data is mined using the existing Twitter APIs for data extraction. Tweets would be selected based on a few chosen keywords pertaining to the domain of our concern, i.e., product reviews.
- **Pre-processing:** In this stage, the data is put through a preprocessing stage in which we remove identifying information such as Twitter handles, timestamps of the message and embedded links and videos. Such information is largely irrelevant and may cause false results to be given by our system.
- **Data Training:** With a large data set available, we can train the algorithm to have a high accuracy.
- **Data Visualization:** Classify data into positive, negative and neutral tweets and visualize it in the form of Pie Charts and Graphs.

III. IMPLEMENTATION PROCESS

Gathering the tweets: Twitter provides the REST search API for searching tweets from Twitter's search index. This is different than using the streaming filter API, in that the latter is real time and starts giving you results from the point of query, while the Search API will give you results from past, up to as far back as the search index goes (usually last 7 days).To start with the API Rate Limit page details the limits of various Twitter APIs, and as per the page the limit for the Search API is 180 Requests per 15 mins window for per-user authentication. The twitter Search API is limited 180 Requests/15 mins limit, and per request you can ask for maximum 100 tweets, giving you a grand total limit of 18,000 tweets/15 mins, if you download 18K tweets before 15 mins, you won't be able to get any more results until your 15minute window expires and you search again.

Pre-processing the Tweets: After gathering the tweets, we need to remove any unnecessary qualities in the data which would make the trained model a poor generaliser. Any redundant characters or words can make the training model less accurate to recognize the actual words. Text pre-processing involves many things like removing emojis, properly formatting the text to remove extra spaces or any other information in the text that we don't believe would add information to our model. We also have to make sure that the information we pass the model is in a format that

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 5, May 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0905078 |

computers can understand. After this pre-processing step, our data should be ready to use for a machine learning classification task.

Model Training and Testing: This stage is a continuation from the previous modules that manages the tweets for unwanted characters using text pre-processing for a machine learning (ML) sentiment analysis task. In this stage we'll split our data into training and test sets. After training the model, we will then use it to classify sentiment on unseen twitter data that has been pre-processed in the same manner as the training data.

Front End and Data Visualization: The objective is to make a front end for the end user to search for a tweet. The result of the front end is to visualize the data into positive, negative and neutral tweets. The end user should be analysing the tweets with the help of the graphical representation of data.

IV. RESULTS

Twitter sentiment analysis is an ML based Program which Sorts the sentiments mentioned by the people in their tweets. NLP (Natural Language Processing), ML (machine learning) & HTML/CSS were the primary technologies used for building this project. The Website will fetch data from the twitter API then analyse the data for sentiment using the machine learning model and present the processed data to the end user in the form of a pie chart sorting the tweets based on that topic as positive, negative and neutral. It is very beneficial for the companies trying to compete in the market with an ever-growing competition. It will help to obtain accurate data related to the Sentiment of the people who have used your company's product and will help companies to critically analyse their product for the next generation of their products. The Fig. 2, Fig. 3 and Fig. 4 shows the results after performing the sentiment analysis on the word "India".

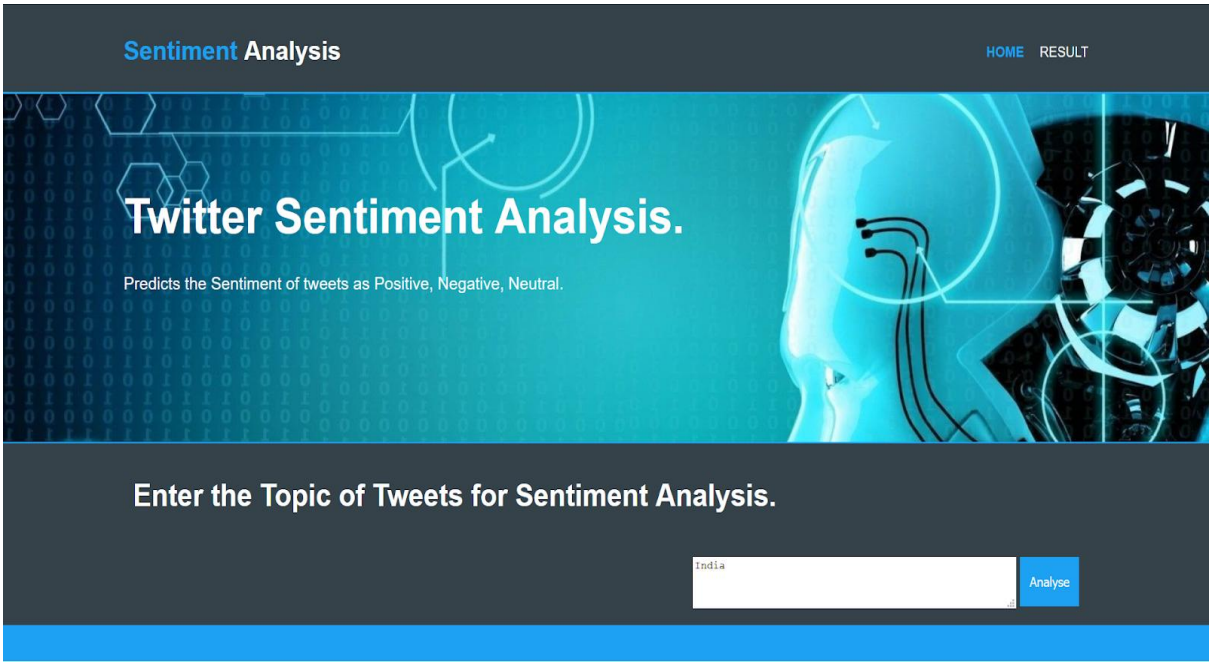


Fig.2. Home Page

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 5, May 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0905078 |

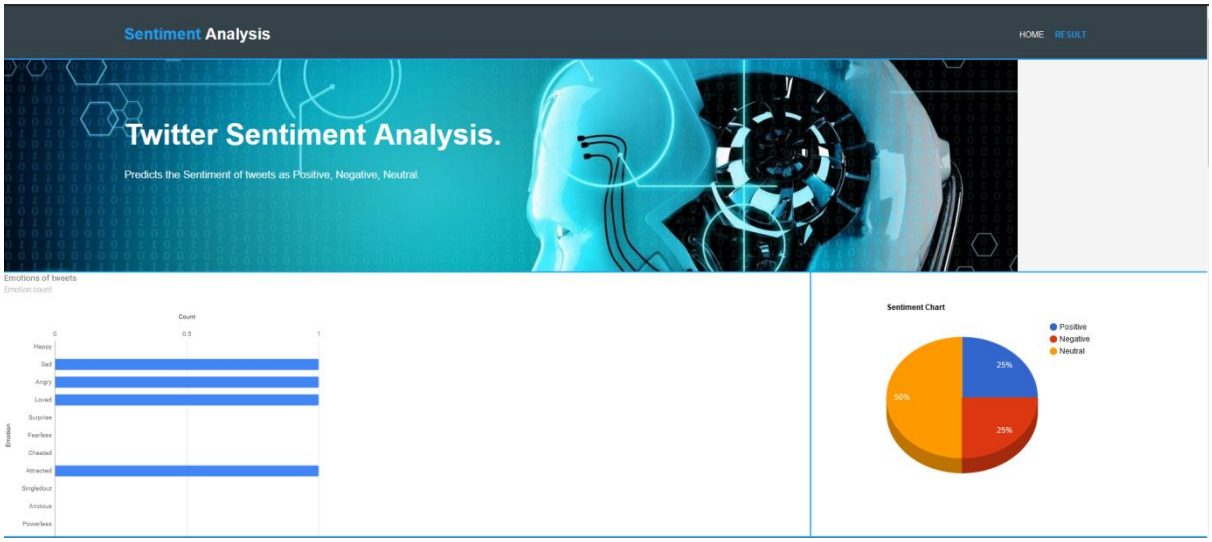


Fig.3. Result of Sentiment Analysis of tweet India

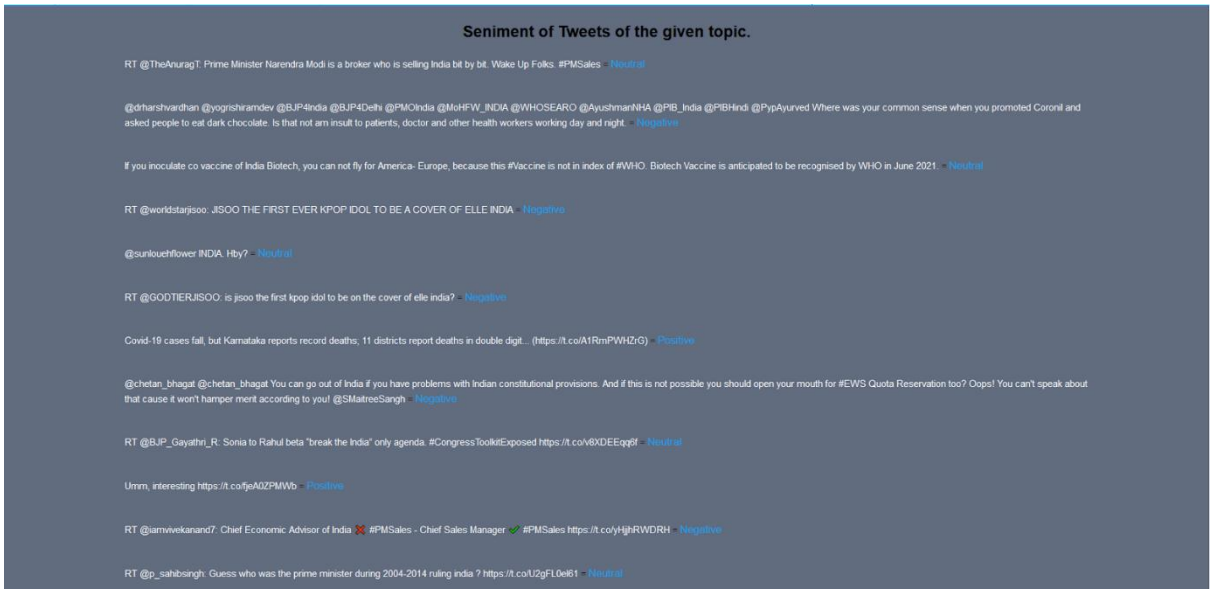


Fig.4. Live tweets results

V. CONCLUSION

Twitter sentiment analysis is developed to analyse customers' views toward the crucial to success within the marketplace. The program is employing a machine-based learning approach which is more accurate for analysing a sentiment mentioned in a specific sentence; along with the Machine Learning Model language process techniques are used. As a result, the program will categorize sentiment into positive, negative and neutral tweets which is represented in a pie chart and html page.

REFERENCES

1. P. D. Turney, “Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews,” in Proceedings of the 40th annual meeting on association for computational linguistics, pp. 417–424, Association for Computational Linguistics, 2002.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 5, May 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0905078 |

2. 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, pp.1320-1326.
3. Po-Wei Liang, Bi-Ru Dai, "Opinion Mining on Social MediaData", IEEE 14th International Conference on Mobile Data Management,Milan, Italy, June 3- 6, 2013, pp 91-96, ISBN: 978-1- 494673-6068-5.
4. T. T. Thet, l.-C. Na, C. S. Khoo, and S. Shakthikumar, "Sentiment analysis of movie reviews on discussion boards using a linguistic approach," in Proceedings of the 1st international CIKM workshop on Topic-sentiment analysis for mass opinion. ACM, 2009, pp. 81-84.
5. Hussein, D.-M.E.D.M. A survey on sentiment analysis challenges. Journal of King Saud University – Engineering Sciences (2016).
6. A Kowcika and Aditi Guptha "Sentiment Analysis for Social Media", International Journal of Advanced Research in Computer Science and Software Engineering, 216-221,Volume 3,Issue 7, July 2013.
7. G.Vinodini and RM.Chandrashekaran, "Sentiment Analysis and Opinion Mining: A Survey", International Journal of Advanced Research in Computer Science and Software Engineering, 283-294, Volume 2, Issue 6, June 2012.
8. Apoorv Agarwal Boyi Xie Ilia Vovsha Owen Rambow Rebecca Passonneau, "Sentiment Analysis of Twitter Data",Columbia University,New York.
9. Pablo Gamallo and Marcos Garcia "A Naive-Bayes Strategy for Sentiment Analysis on English Tweets" Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014), pages 171–175, Dublin, Ireland, Aug 23-24 2014.
10. Harry Zhang "The Optimality of Naive Bayes", FLAIRS2004 conference.

INNO SPACE

SJIF Scientific Journal Impact Factor

Impact Factor:
7.488

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details