

A Project Report on
Twitter Sentiment Analysis

Submitted in partial fulfillment of the requirements

in

Computer Engineering

by

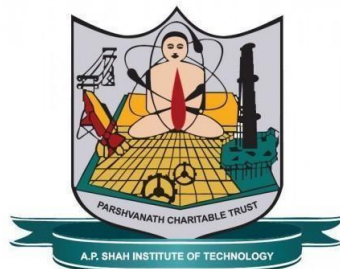
Atharva Kulkarni (17102070)

Manas Mohite (17102069)

Parin Dodhiya (17102036)

Under the Guidance of

Prof. S. S. Aloni



Department of Computer Engineering

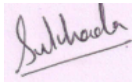
A.P. Shah Institute of Technology G.B.Road,Kasarvadavli, Thane(W), Mumbai-400615

UNIVERSITY OF MUMBAI

Academic Year 2020-2021

Approval Sheet

This Project Report entitled “**Twitter Sentiment Analysis**” Submitted by “**Atharva Kulkarni(17102070)**”, “**Manas Mohite(17102069)**”, “**Parin Dodhiya (17102036)**” is approved for the partial fulfillment of the requirement in **Computer Engineering** from **University of Mumbai** .



Prof. S.S. Aloni

Guide



Prof. S.H. Malave

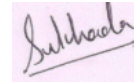
Head & Guide, Computer Engineering Department

Place: A.P. Shah Institute of Technology, Thane

Date: 29-05-2021

CERTIFICATE

This is to certify that the project entitled "**Twitter Sentiment Analysis**" submitted by "**Atharva Kulkarni(17102070)**", "**Manas Mohite(17102069)**", "**Parin Dodhiya(17102036)**" for the partial fulfillment of the requirement for award of a degree **Bachelor of Engineering in Computer Engineering**, to the University of Mumbai, is a bonafide work carried out during the academic year 2020-2021.



Prof. S. S. Aloni
Guide



Prof. S.H. Malave
Head & Guide, Computer Engineering Department

Dr. Uttam D. Kolekar
Principal

External Examiner

Place: A.P. Shah Institute of Technology, Thane
Date: 29-05-2021

Declaration

We declare that this written submission represents our ideas in our own words and where others' ideas or words have been included, We have adequately cited and referenced the original sources. We also declare that We have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in our submission. We understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

(Signature)

Atharva Kulkarni (17102070)
Manas Mohite (17102069)
Parin Dodhiya (17102036)

Date:29-05-2021

Contents

1.Project Conception and Initiation...	1
1.1 Abstract.....	1
1.2 Objectives.....	2
1.3 Literature review... ..	3
1.4 Problem Definition... ..	5
1.5 Scope... ..	5
1.6 Technology stack.....	6
1.7 Benefits for environment and society... ..	7
2.Project Design.....	8
2.1 Proposed system.....	8
2.2 Design (Flow of Modules)	9
2.3 Class Diagram... ..	10
2.4 Module-1... ..	11
Module-2... ..	11
Module-3... ..	12
Module-4... ..	12
3.Implementation	13
3.1Proposed System	13
3.1.1 Algorithms... ..	14
3.1.2 Pseudo code... ..	15
3.1.3 Platforms for execution... ..	16
4.Results... ..	17
5.1 Conclusion.....	20
5.2Future Scope... ..	20
6.References... ..	21
7.Bibliography	22

1.Project Conception and Initiation

1.1 Abstract

Twitter Sentiment Analysis Web App is a project aimed to analyze 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 analyze 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 users feedback in order to improve quality of service or product. Our project helps to know the feedback of users on Twitter.

1.2 Objectives

1. To develop a model to predict the different sentiments of twitter data.
2. To help companies analyse how a product is received by users on twitter.
3. To create a frontend GUI application for data analysis which would be easy for the user to understand .

1.3 Literature review

1. Twitter as a Corpus for Sentiment Analysis and Opinion Mining [1]

The paper titled "Twitter as a Corpus for Sentiment Analysis and Opinion Mining" was published by Alexander Pak and P. Paroubek in 2010. Recent research based on sentiment analysis says that the analysis of opinion utilizes simultaneous learning. Pak and Paroubek utilized tweets which end with emoticons like ":)" ":-)" as positive, and ":(" ":-(" as negative. They accumulated models including Max Entropy, Support Vector Machines (SVM) and Naive Bayes and concluded that SVM performed the best amongst various others, SVM was the best performer of all the classifiers due to its increased precision.

2. Sentiment analysis of movie reviews on discussion boards using a linguistic approach[2]

The authors proposed a linguistic approach system for aspect based opinion mining, which is a clause/Sentence level sentiment analysis for opinionated texts. It creates a syntactic dependency tree for each message post sentence and divides the sentence into clauses. It then determines the contextual based sentiment score for each clause using grammar dependency of words and uses SentiWordNet which has prior sentiment scores for the words and also from domain specific lexicons.

3.Opinion Mining on Social MediaData[3]

To collect data from Twitter, they used the Twitter API. . Tweets which contain opinions were filtered out. For polarity identification, a Unigram Naive Bayes model was created. They have used the Mutual Information and Chi Square function extraction methods to remove unnecessary features. Finally, this strategy did not improve the accuracy of predicting whether tweets were positive or negative.

4.Analyzing the Political Landscape of 2012 Korean Presidential Election in Twitter [4]

Kim and Jeong initially collected the tweets. Their collection and mining techniques processed large datasets in real time. To demonstrate the usefulness of their approaches, they specifically focused on topical trend analysis and network analysis to examine presidential issues embedded in Twitter data.

1.4 Problem Definition

Twitter has a large number of users and every user has a different opinion about a particular topic. It is very difficult to analyse the sentiments of various opinionated users. Twitter only allows only a limited number of characters in one tweet (280 characters), so users try to abbreviate words and tend to skip punctuations. The proposed system will comprehend these constraints and try to predict the sentiments of tweets.

1.5 Scope

1. To help users understand the sentiment of trending topics on twitter.
2. The project will be helpful for companies to understand user sentiments on their new product or hardware released.
3. Visual representation of data for easier understanding to users.

1.6 Technology stack

1. **Twitter API:** Twitter provides an API for developers to access twitter in unique and advanced methods. It is a very reliable and easy to use API. Many advertisers around the world use it for accessing twitter data.

2. **Visual Studio Code:** Visual Studio Code is a free source-code editor made by Microsoft for Windows, Linux and macOS. Features include support for debugging, syntax highlighting, intelligent code completion, snippets, code refactoring, and embedded Git.

3. **NLTK:** Natural Language ToolKit is a group of libraries built in python language. NLTK is widely used for text processing of large datasets. Using NLTK libraries, the processing of texts is done faster and more efficiently.

1.7 Benefits for environment and society

1. Users can understand the sentiments of a particular tweet on Twitter.
2. Companies can use this platform to understand customer reviews in an easy and efficient way. Companies can hence take steps accordingly.
3. This project is not only beneficial for the companies to get a review of their product but also for the common users to search for their favourite topics and interests on twitter and know other users' opinions.

2.Project Design

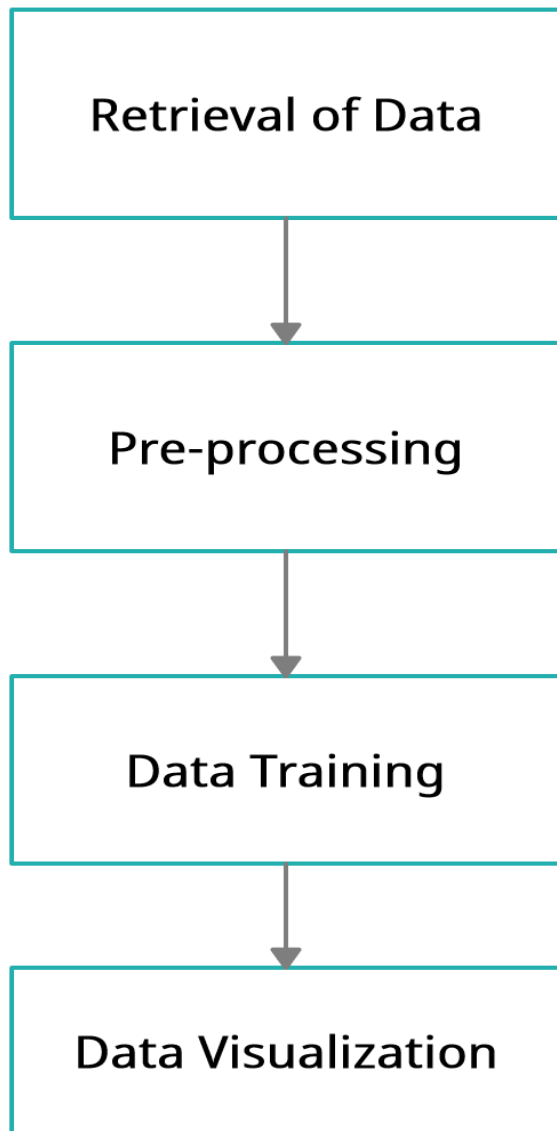
2.1 Proposed system

The proposed flow of modules can be classified into four modules:

- 1.Gathering the data
- 2.Processing the data for unwanted characters and preparing data set,
- 3.Applying machine level algorithms to training and testing data for generating outputs, and lastly
- 4.Creating a frontend for data visualization.

Gathering the tweets is a basic module for accessing the tweets using the Twitter Search API. The next step is to divide the main dataset into training and testing dataset. Then feature engineering the data to suit our project. Finally data visualization tools will be used for better understanding of sentiments of tweets.

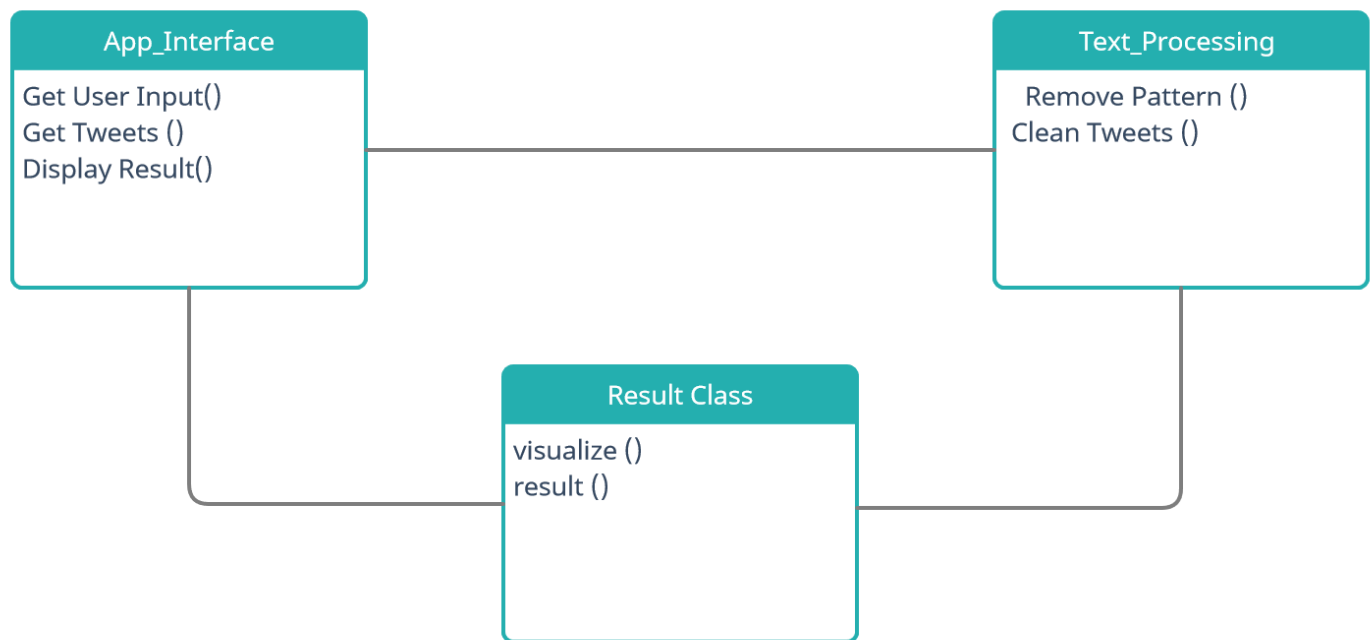
2.2 Design(Flow of Modules)



Flow of modules

- The first objective is to fetch relevant tweets from the user query.
- Applying pre-processing techniques to the data to remove the unwanted characters, words, urls etc is the second step.
- Now, the processed data must be analyzed and given polarity of sentiment accordingly.
- Finally for understanding of the data, visualization techniques are used.

2.3 Class Diagram



Class Diagram

2.4.1 Module-1

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 later 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 15 minute window expires and you search again.

2.4.2 Module-2

After gathering the tweets, we need to remove any unnecessary qualities in the data which would make the trained model a poor generalizer. Any redundant characters or words can make the training model less accurate to recognize the actual words. Text preprocessing 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 computers can understand. After this pre-processing step, our data should be ready to use for a machine learning classification task.

2.4.3 Module-3

This stage is a continuation from the previous modules that manages the tweets for unwanted characters using text preprocessing 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 preprocessed in the same manner as the training data.

2.4.4 Module-4

The objective is to make a front end for the end user to search for a tweet. The end result of the front end is to visualize the data into positive, negative and neutral tweets. The end user should be able to analyze the tweets with the help of the graphical representation of data.

3.Implementation

3.1 Proposed System

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.

Preprocessing: 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.

3.1.1 Algorithm

Natural language processing (NLP) is a branch of artificial intelligence that helps computers understand, interpret and manipulate human language. NLP draws from many disciplines, including computer science and computational linguistics, in its pursuit to fill the gap between human communication and computer understanding. Natural language processing includes many different techniques for interpreting human language, ranging from statistical and machine learning methods to rules-based and algorithmic approaches. This project uses raw twitter data which is informal language and contains many unnecessary words which are not useful for the model. In general terms, NLP tasks break down language into shorter, elemental pieces, try to understand relationships between the pieces and explore how the pieces work together to create meaning.

Original Tweet	RT @YourKayBol: 283 samples from the 301 that were pending have all tested negative
Removal of RT and username	283 samples from the 301 that were pending have all tested negative
Removal of figures	samples from the that were pending have all tested negative
Removal of stop words	sample pending tested negative

3.1.2 Pseudo code

```
program start
take input text from user
connect API and download tweets
send downloaded tweets to processing model
function(text-processing-operations)
    1.removal of special characters
    2.removal of urls and links
    3.removal of stopwords
assigning polarity to tweets
analysis of polarity
data visualization of analysis
```

3.1.3 Platforms for execution

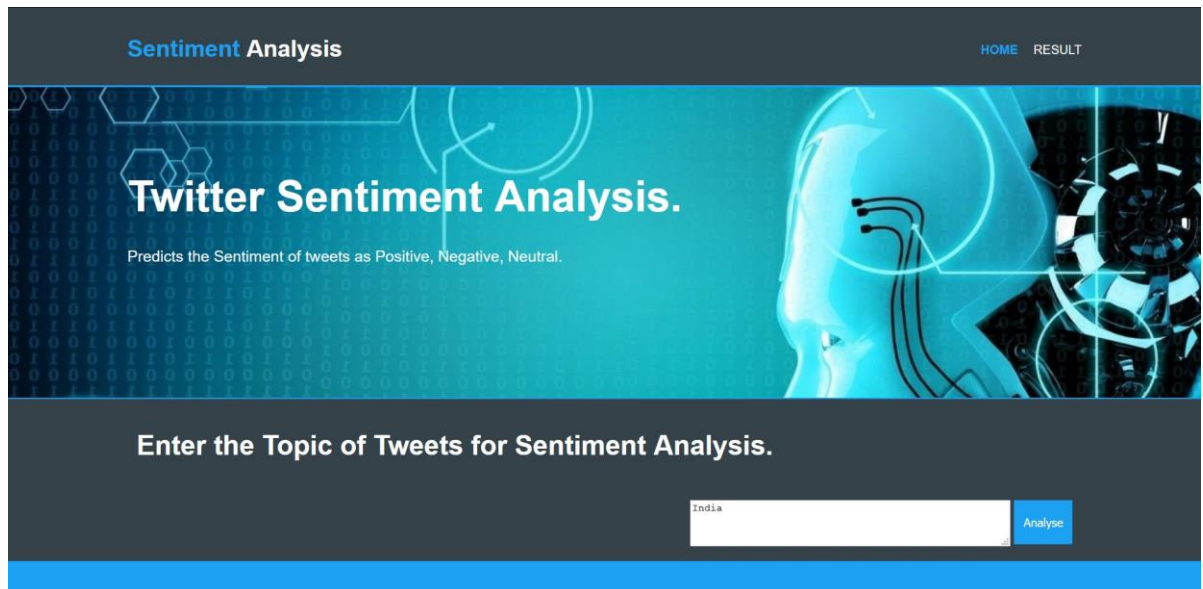
1. **Google Collab:** Colab allows anybody to write and execute arbitrary python code through the browser, and is especially well suited to machine learning and data analysis. Collab is well suited for this project for model training and testing purposes.

2. **Visual Studio Code:** Visual Studio Code is a free source-code editor made by Microsoft for Windows, Linux and macOS. Features include support for debugging, syntax highlighting, intelligent code completion, snippets, code refactoring, and embedded Git. VS Code is used for coding all the front end pages and the main app file.

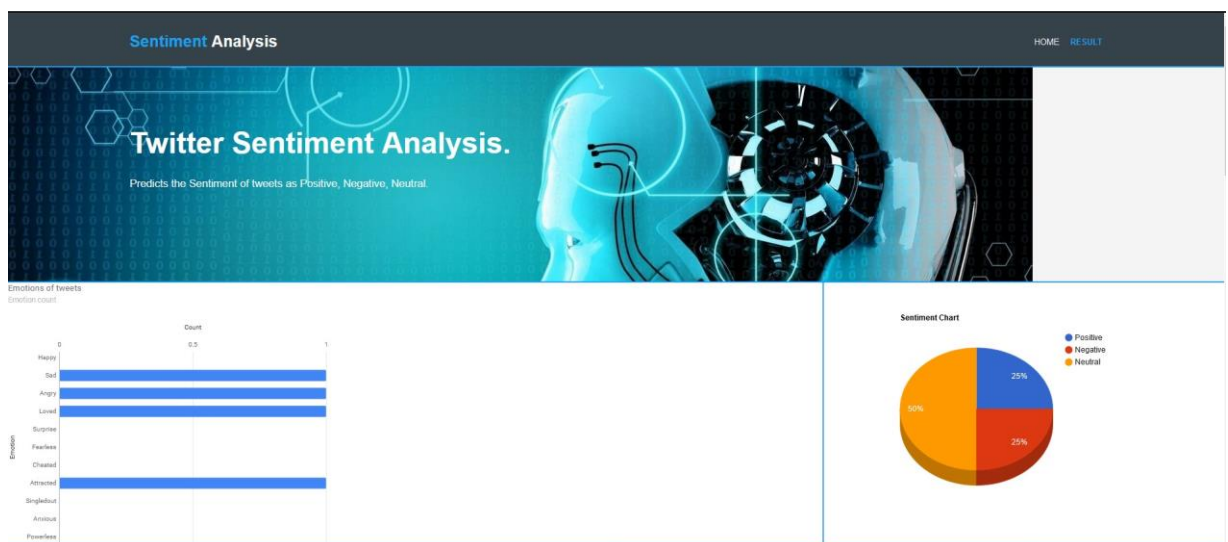
4.Results

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

Input Tweet	Sentiment Analysis
Loved the New Apple phone absolutely worth the money	Positive
Spent a Day with my friends at a theme park	Neutral
Not at all satisfied with the customer service of the brand	Negative



Web App Home Page



Result Page and Data Visualization

Seniment of Tweets of the given topic.

RT @TheAnuragT: Prime Minister Narendra Modi is a broker who is selling India bit by bit. Wake Up Folks. #PMSales - [Neutral](#)

@dharshvardhan @yogishramdev @BJP4India @BJP4Dehi @PMOIndia @MoF_W_INDIA @WHOSEARO @AyushmanNHA @PB_India @PBHindi @PypAyurved Where was your common sense when you promoted Coronil and asked people to eat dark chocolate. Is that not an insult to patients, doctor and other health workers working day and night. - [Negative](#)

If you inoculate co vaccine of India Biotech, you can not fly for America- Europe, because this #vaccine is not in index of #WHO. Biotech Vaccine is anticipated to be recognised by WHO in June 2021. - [Neutral](#)

RT @worldstarjsoo: JISOO THE FIRST EVER KPOP IDOL TO BE A COVER OF ELLE INDIA - [Negative](#)

@suntouchflower INDIA. Hby? - [Neutral](#)

RT @GODTIERJISOO: is jisoo the first kpop idol to be on the cover of elle india? - [Negative](#)

Covid-19 cases fall, but Karnataka reports record deaths, 11 districts report deaths in double digit... (<https://t.co/A1RmFWHzGj>) - [Positive](#)

@chetan_bhagat @chetan_bhagat You can go out of India if you have problems with Indian constitutional provisions. And if this is not possible you should open your mouth for #EWS Quota Reservation too? Oops! You can't speak about that cause it won't hamper merit according to you! @SMatreeSangh - [Negative](#)

RT @BJP_Gayathri_R: Sonia to Rahul beta "break the India" only agenda. #CongressToolKitExposed <https://t.co/vBxDEEqg9f> - [Neutral](#)

Umm, interesting <https://t.co/j5A0ZPMWb> - [Positive](#)

RT @iamvivekanand7: Chief Economic Advisor of India ❌ #PMSales - Chief Sales Manager ✔️ #PMSales <https://t.co/yHjHRWDRH1> - [Negative](#)

RT @p_sahibsingh: Guess who was the prime minister during 2004-2014 ruling india ? <https://t.co/U2gFL0e01> - [Neutral](#)

Live Tweets with Sentiment Polarity

5.1 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 are represented in a pie chart and html page.

5.2 Future Scope

1. We plan to develop a Mobile application which would make the project more accessible to the general public as most of them are more active on mobile platforms such as Android/iOS
2. We also plan to implement more data visualization techniques for in depth analysis of sentiments.

6. References

1. 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.
2. T. T. Thet, I.-C. Na, C. S. Khoo, and S. Sakthikumar, "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.
3. Po-Wei Liang, Bi-Ru Dai, "Opinion Mining on Social Media Data", IEEE 14th International Conference on Mobile Data Management, Milan, Italy, June 3- 6, 2013.
4. Min Song, Meen Chul Kim, Yoo Kyung Jeong "Analyzing the Political Landscape of 2012 Korean Presidential Election in Twitter" Published in: IEEE Intelligent Systems (Volume: 29, Issue: 2, Mar.-Apr. 2014)
5. 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.
6. G. Vinodini and RM. Chandrashekar, "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.
7. 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.
8. Apoorv Agarwal Boyi Xie Ilia Vovsha Owen Rambow Rebecca Passonneau, "Sentiment Analysis of Twitter Data", Columbia University, New York.
9. P. D. Turney, "Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews," , Association for Computational Linguistics, 2002.

7. Bibliography

[1] M. Trupthi, Suresh Pabboju, G. Narasimha, "Sentiment Analysis on Twitter Using Streaming API" 2017 IEEE 7th International Advance Computing Conference (IACC).

[2] Reference Dataset - <https://www.kaggle.com/kazanova/sentiment140>

[3] Step by Step: Twitter Sentiment Analysis in Python | by Yalin Yener

[4] <https://medium.com/analytics-vidhya/twitter-sentiment-analysis-b9a12dbb2043>