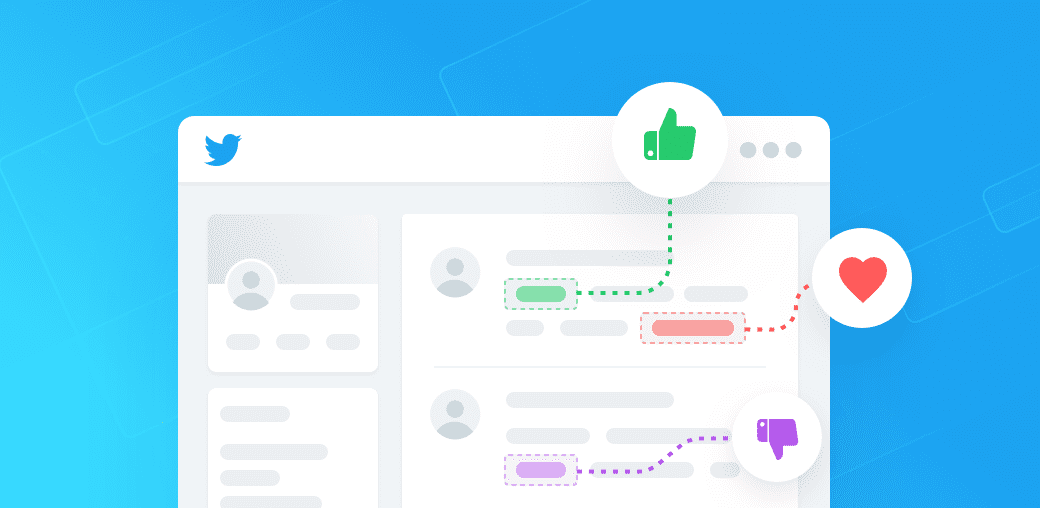
A

Project Report On

**“ LIVE TWITTER SENTIMENT ANALYSIS ”**





**Prepared by**

18DCS007

RUDRA BARAD

**Under the Supervision of**

Prof. Radhika Patel

**A Report Submitted to**

Charotar University of Science & Technology (CHARUSAT)

Bachelor of Technology (B.Tech) in

Computer Science & Engineering (CSE)

for Partial Fulfillment of the Requirements for the

4th Semester Software Group Project-II (CS255)

 **Submitted at**



**DEPSTAR - CHARUSAT**

**At: Changa, Dist: Anand – 388421**

**APRIL, 2020**



**CERTIFICATE**

This is to certify that the report entitled “ **Live Twitter Sentiment Analysis** ” is a Bonafide work carried out by **Mr. RUDRA BARAD | 18DCS007** under the guidance and supervision of **Prof. RADHIKA PATEL** for the subject **CS255** | **Software Group Project-II** (CSE) of 4th Semester of Bachelor of Technology in **DEPSTAR** at Faculty of Technology & Engineering- CHARUSAT, Gujarat.

To the best of my knowledge and belief, this work embodies the work of candidate himself, has duly been completed, and fulfills the requirement of the ordinance relating to the B.Tech. Degree of the University and is up to the standard in respect of content, presentation and language for being referred to the examiner.

|  |
| --- |
| Prof. Parth Goel  HOD  CSE  DEPSTAR, Changa, Gujarat  Prof. Radhika Patel  IT  DEPSTAR, Changa, Gujarat  Prof. Chandrasekhar Pawar  CSE  DEPSTAR, Changa, Gujarat |
| Dr. Amit Ganatra  Principal, DEPSTAR  Dean, FTE  CHARUSAT, Changa, Gujarat. | |

**ABSTRACT**

Social media is turning out to be a popular trend nowadays. People spend most of their free time surfing and scanning the social media. And due to this extra influence of social media on people in general, it has turned out to be an important factor in influencing people round the globe. Many political leaders are nowadays initiating to join social media after realizing its importance on people’s mind and their daily routine. Due to such an outrageous effect of social media on common people it is very easy to analyze how people feel or think by monitoring their social media profiles. From all the social media platforms “Twitter” differentiates itself as it plays a major role in influencing and guiding people with each other’s help. Almost all the major authorities such as presidents, prime minister, other political leaders and many business person are actively handling their twitter handles and each of their tweets may lead to major impacts on many aspects of daily routine of common people round the world. So the twitter sentiment analysis project will help us to know and analyze what are the common trends or mindset of people towards a particular person, place or topic.

This analysis is better with twitter because it is a serious and mature means of communication between the people of the community. This sentiment analysis helps the user to know how are people thinking about any topic be it positive, negative or neutral. It is most likely that people like to share their views on any particular topic on social media platforms like twitter, Facebook, etc. So, twitter sentiment analysis is a very efficient method to identify common trends amongst people, especially among the youths, as youths are generally more inspired by the means of social media.

**ACKNOWLEDGEMENT**

I, the developer of the project “Live twitter sentiment analysis”, with immense pleasure and commitment would like to present the project assignment. The development of this project has given me wide opportunity to think, implement and interact with various aspects of management skills as well as the new emerging technologies.

Every work that one completes successfully stands on the constant encouragement, good will and support of the people around. I hereby avail this opportunity to express my gratitude to number of people who extended their valuable time, full support and cooperation in developing the project.

I express deep sense of gratitude towards our Head of the CSE Department, Prof. Parth Goel and project guides Prof. Radhika Patel and Prof. Chandrasekhar Pawar for the support during the whole session of study and development. It is because of them, that I was prompted to do hard work, adopting new technologies.

I would also like to thank all the mentor for their guidelines throughout the development phase of the project. They encouraged me to look forward to learn and implement new and emerging technologies. They also guided me to go for some user friendly and extremely useful real life application.

They altogether provided me favorable environment, and without them it would not have been possible to achieve my goal.

**INDEX**

|  |  |  |
| --- | --- | --- |
|  | **Certificate**  **Abstract** | **2**  **3** |
|  | **Acknowledgement** | **4** |
| **1.** | PROJECT DEFINITION | **6** |
| **2.** | PROJECT DESCRIPTION | **7** |
| **3.** | SOFTWARE & HARDWARE REQUIREMENTS | **12** |
| **4.** | SYSTEM FLOW CHART | **14** |
| **5.** | IMPLEMENTATION SNAPSHOTS | **17** |
| **6.** | PROJECT OUTCOMES | **21** |
| **7.** | CURRENT SYSTEM & ENHANCEMENTS IN PROJECT | **23** |
| **8.** | LIMITATION OF PROJECT & FUTURE SCOPE | **25** |
| **9.** | REFERENCES | **27** |

**CHAPTER 1: PROJECT DEFINITION**

**Live Sentimental analysis of Twitter and classifying the tweets as Positive, Negative or Neutral.**

**CHAPTER 2: PROJECT DESCRIPTION**

**2.1 PROJECT OVERVIEW :**

It is self evident that social media is gaining a lot of popularity amongst almost all age groups and mostly among the youth generation. Twitter is also a very good social medium in which people share their motives, thoughts and feelings with others present on twitter itself. Twitter is an American microblogging and social networking service on which users post and interact with messages known as "tweets". Registered users can post, like, and retweet tweets, but unregistered users can only read them. The live twitter sentiment analysis is a project in which the sentiments or the feelings of people are recognized using the text that they post which are popularly known as tweets. Taking twitter as a platform for such analysis is necessary as twitter is one of the platforms on which any tweet from one or the other person can influence the public thinking or public reviews. So sentiment analysis of tweets can help for business and political empowerment of the society. Project basically takes two approaches to solve the sentiment analysis problem. First approach to the sentiment analysis problem is using data mining . Data mining is a much simpler and easier way to solve the problem of twitter sentiment analysis. Another approach is the data analytics approach in which we train data models based on dataset and algorithm applied to the datasets. We use tweepy module to fetch live tweets from twitter using the twitter API. Twitter provides a tool to the developers which is known as twitter API, which allows the API user to fetch and analyze. It also allows the developer to post tweets to twitter using the API. Tweepy is an interface between twitter API and python program code which enables the processing of the fetched tweets from twitter using python programs.

**2.2 INTRODUCTION TO DOMAINS :**

**WEB SCRAPPING**

It a technique used for extracting data from website or any other platform where by data is extracted and stored in local file in our computer or to a Table in Databases. Data displayed by most websites can only be viewed using a web browser. They do not offer the functionality to save a copy of this data for personal use. The only option then is to manually copy and paste the data - a very tedious job which can take many hours or sometimes days to complete. [Web Scraping](http://www.webharvy.com/) is the technique of automating this process, so that instead of manually copying the data from websites, the [Web Scraping software](http://www.webharvy.com/) will perform the same task within a fraction of the time.

**DATA MINING**

Data mining is the process of sorting through large data sets to identify patterns and establish relationships to solve problems through data analysis. Using a broad range of techniques, you can use this information to increase revenues, cut costs, improve customer relationships, reduce risks and more. Data mining tools allow enterprises to predict future trends.Data mining uses sophisticated mathematical algorithms to segment the data and evaluate the probability of future events.

**MACHINE LEARNING**

Machine learning is the subset of Data Analytics and Artificial Intelligence.Machine Learning is the field of study that gives computers the capability to learn without being explicitly programmed. ML is one of the most exciting technologies that one would have ever come across. As it is evident from the name, it gives the computer that makes it more similar to humans: The ability to learn. Machine learning is actively being used today, perhaps in many more places than one would expect.

**2.3 APPROACH TO PROBLEM :**

There are basically to approaches to the problem. The first approach to sentiment analysis is using data mining. In data mining we use simple python codes to authenticate the twitter API user , to fetch the data and for the sentiment analysis of the data. Textblob is a popular tool used in data mining projects. It is basically a predefined module which has the collection of classified positive, negative and neutral words. Textblob also provides the options for sentiment analysis from a particular point of view , for example textblob allows user to analyse any data from business or political views to make the data classification faster and efficient. No pre-processing or model training for the data mining part as the text blob module itself works as data set and the classifier model both.

Another approach to solve the live twitter sentiment analysis problem is using data analytics. Machine learning is a advanced and latest technique in data analytics to solve the classification and analysis problems. A data model needs to be trained for analysis using machine learning. The data model is trained using naïve Bayes algorithm which is much used algorithm for prediction and classification work in machine learning. The training is done using two datasets listed as positive dataset and negative data set. The Naïve bayes algorithm extracts the most informative features from the tweets and analyze it to check whether the tweet is positive or negative. The more number of features are extracted we get the better accuracy for our model.

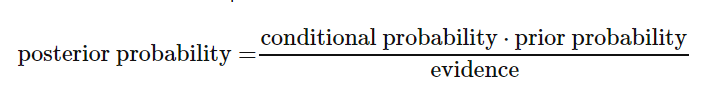
The accuracy for the output is directly proportional to the data fetched in both the approaches. If we fetch more number of tweets for processing , we get better accurate outputs. For the machine learning approach , feature extraction also plays and extremely important role in increasing accuracy of the data model. But increasing the dataset may lead to increased processing time and increase in time to train the model. Preprocessing is required only for the first time or the time when we change the dataset. Once the data model training is completed the trained classifier is stored as a pickle file on the disk.

**2.4 NAÏVE BAYES ALGORITHM :**

Naïve Bayes algorithm is a classification technique based on Bayes’ Theorem with an assumption of independence among predictors. In simple terms, a Naive Bayes classifier assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature.It is family of algorithms where all of them share a common principle. Naïve Bayes algorithm is a combination of a theorem proposed by Bayes and an assumption based on the theorem proposed by Naïve.Bayes’ Theorem finds the probability of an event occurring given the probability of another event that has already occurred. Bayes’ theorem is stated mathematically as the following equation:



We are trying to find probability of event A, given the event B is true. Event B is also termed as evidence.P(A) is the priori of A (the prior probability, i.e. Probability of event before evidence is seen). The evidence is an attribute value of an unknown instance(here, it is event B).P(A|B) is a posteriori probability of B, i.e. probability of event after evidence is seen. So, the above equation can be written in general as below.



Initially the posterior probability is very less. The posterior probability gets increased as the number of trained data elements increases.

Naïve assumption states the mutual independence of features amongst the data. The probability for feature independence is named as class probability. Based on naïve assumption the evidence part is divided into independent parts which specify the independence of features in data.

**CHAPTER 3: SYSTEM REQUIREMENTS**

**3.1 HARDWARE AND SOFTWARE THAT WERE USED IN THIS PROJECT :-**

**SOFTWARE USED:**

* Command prompt to run the code.
* Pre-installed python environment to compile and execute python code.

**SYSTEM CONFIGURATION ON WHICH THE PROJECT WAS DEVELOPED:**

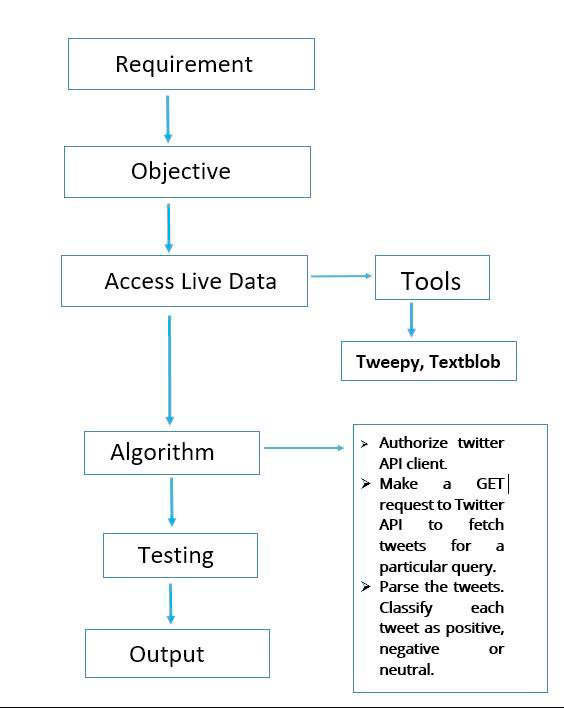
* 1TB Hard drive storage for application data storage
* 8 GB RAM,Intel i5 8th generation processor.
* Internet connectivity**.**

**OTHER REQUIREMENTS:**

* Modules such as tweepy as an interface for twitter API to fetch tweets, textblob for data mining purpose.
* Other modules such as pickle, nltk, etc for python processing , required for data model training through machine learning.

**CHAPTER 4: SYSTEM FLOWCHART**

**4.1 FLOW OF PROJECT USING DATA MINING APPROACH :**



The flow of project goes as the tweets will be fetched from twitter using the twitter API and the user will be authenticated and then the topic , user wants to search for is entered and the tweets related to entered subject are fetched for sentiment analysis. Textblob module analysis the fetched tweet data using it own classifier feature , and the output of tweet to be positive negative or neutral is returned.

**4.2FLOW OF PROJECT USING MACHINE LEARNINGAPPROACH :**

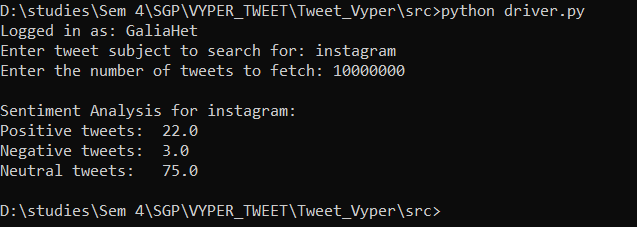


The basic flow is similar to that of data mining flow but the main difference here is the trained data model which acts as a classifier during analysis. Whenever a tweet is fetched the most informative features of the tweet are extracted by applying appropriate algorithms and based on the features any tweet is classified as a positive or a negative tweet by the classifier. The fetched tweets are stored in a text file using which the graph can also be plotted to have a graphical representation of the analysis.

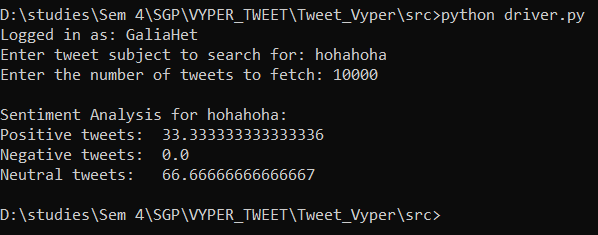
**CHAPTER 5: IMPLEMENTATION SNAPSHOTS**

**OF PROJECT**

1. Fetching and analyzing latest 10000000 tweets on Instagram using data mining at a particular instance of time.



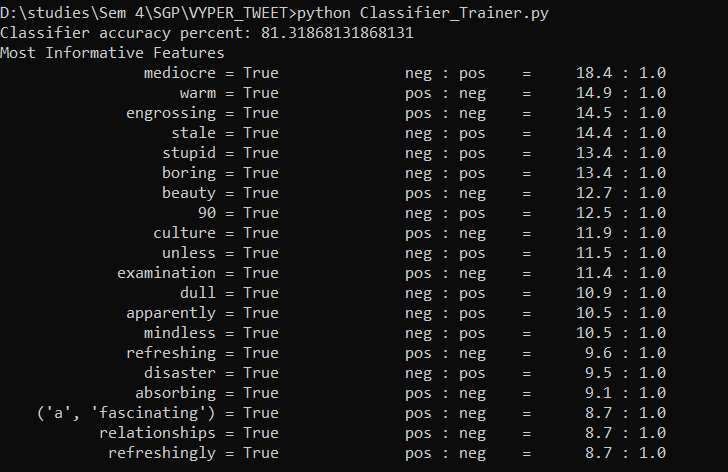
1. The program will return analysis of maximum number of tweets fetched if some anomalous data is entered.



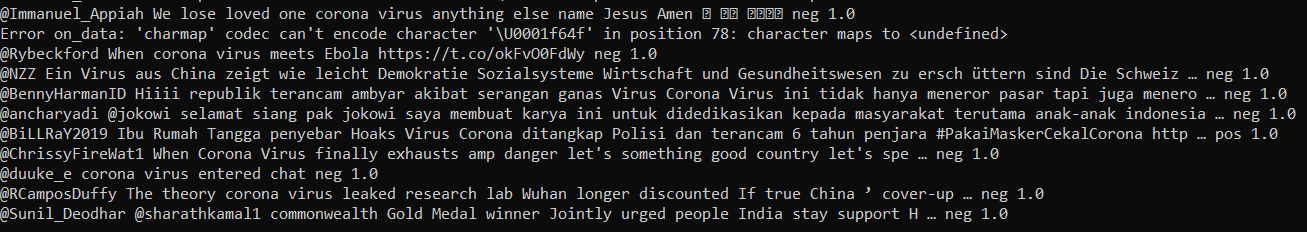
1. Fetching and analyzing latest 10000000 tweets on Instagram using data mining at some another instance of time.

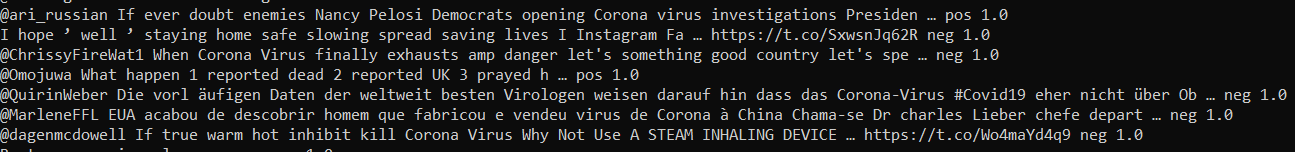


1. Accuracy of the trained model and most informative features of the data in the dataset.

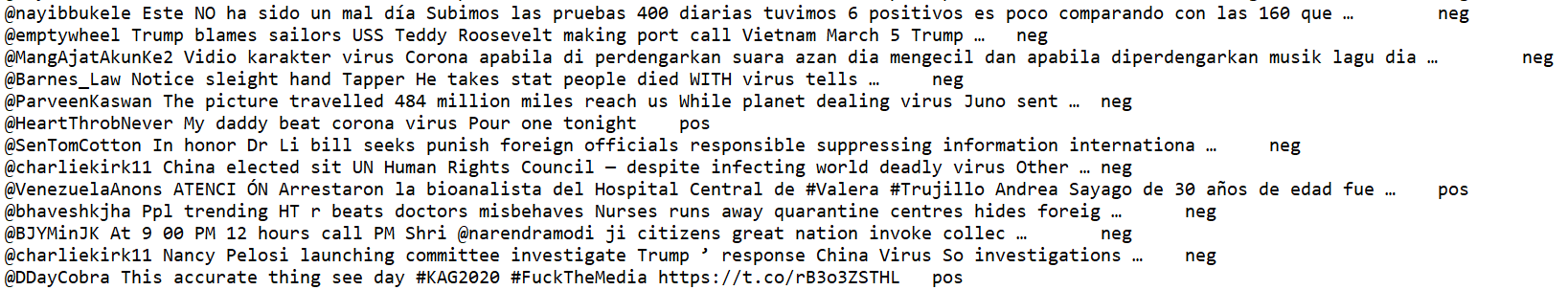


1. Live tweets fetching ongoing in command prompt.

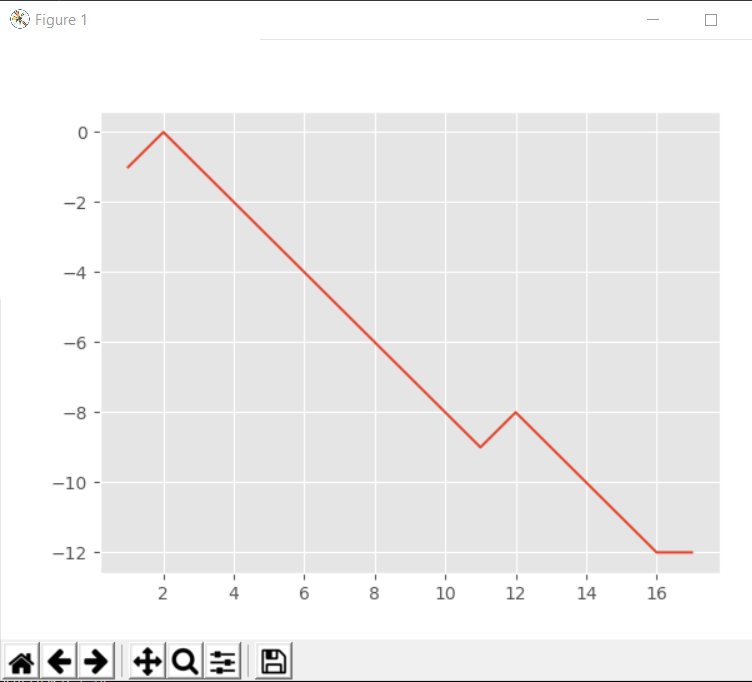


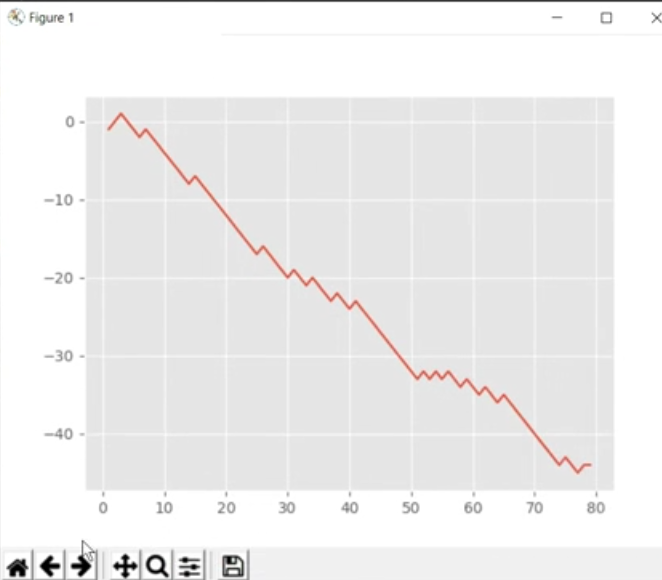


1. Fetched tweets stored in the text file .



1. Graph plots for same subject at two different time instances.





Accuracy

**81.3%**

**CHAPTER 6: PROJECT OUTCOME**

* Sentiment analysis helps businesses process huge amounts of data in an efficient and cost-effective way.
* Real-Time Analysis Sentiment analysis can identify critical issues in real-time. Sentiment analysis models can help you immediately identify different kinds of situations, so you can take action right away.
* By using a centralized sentiment analysis system, companies can apply the same criteria to all of their data, helping them improve accuracy and gain better insights.
* Monitor social media mentions of your brand and automatically categorize by urgency.
* You can develop a more insightful, data-based marketing strategy & Understand your customers.

**CHAPTER 7: CURRENT SYSTEMS & ENHANCEMENTS**

**CURRENT SYSTEMS:**

* Most of the current systems for live twitter sentiment analysis which use data mining are not able to fetch more than some particular amount of tweets.
* Some professionally developed system provides better accuracy and user interface but is extremely expensive to maintain. They also high maintenance.
* The professionally built system requires high configuration softwares and devices to maintain and process data.
* The currently existing system requires a user to have some basic understanding of the system to use it.

**ENHANCEMENTS IN OUR PROJECT:**

* Our system does not require any high configuration devices or softwares to execute or for processing.
* In our system we have tried to reduce the data model training time to get instant and efficient results for the machine learning part.
* In this system the live tweet feed is fetched directly from twitter using twitter API which overcomes the limited tweets fetching problem.
* The project can be easily implemented by even the users with no knowledge regarding the project ,so it is extremely user friendly

**CHAPTER 8: LIMITATIONS AND FITURE SCOPE**

* In this project we have used cluster of words in the File to train the model so as to predict whether the tweet is positive or negative. In order to make our model more efficient and better we can use single words instead of cluster of words.
* If a tweet contains 2 parts one positive and the other half negative than our model will consider as negative tweet rather than neutral.
* Accuracy is directly proportional to available data set. There is always a scope to increase the Accuracy.
* We can add a feature that identifies whether a correct hash tag is used with the tweet. Now-a-days people uses trending hash tags, which is not at all related with their tweet just to promote their tweet.

**CHAPTER 9: REFERENCES**

* <https://towardsdatascience.com/live-sentiment-analysis-on-twitter-data-using-tweepy-keras-and-django-99c344e90488>

* <https://pythonprogramming.net/graph-live-twitter-sentiment-nltk-tutorial/>
* <https://monkeylearn.com/blog/sentiment-analysis-of-twitter/>

* <https://www.csc2.ncsu.edu/faculty/healey/tweet_viz/tweet_app/>
* <https://towardsdatascience.com/creating-the-twitter-sentiment-analysis-program-in-python-with-naive-bayes-classification-672e5589a7ed>