

# A UDP Project Report On

## Sentiment Analysis

**Prepared by**

Tirth Patel      (**Enrollment No: 170020107041**)

Ravi Sahani    (**Enrollment No: 170020107049**)

*Guided By:*

Prof. Bansari Thakkar  
(Assistant Professor, CE Department)



**DEPARTMENT OF COMPUTER ENGINEERING**

**AHMEDABAD INSTITUTE OF TECHNOLOGY**

GUJARAT TECHNOLOGICAL UNIVERSITY

Summer-2021

**AHMEDABAD INSTITUTE OF TECHNOLOGY**  
**COMPUTER ENGINEERING**  
**Summer-2021**



**CERTIFICATE**

This is to certify that the project entitled “[Sentiment Analysis](#)” has been carried out by [Tirth Patel \(170020107041\)](#), [Ravi Sahani \(170020107049\)](#) under my guidance in fulfillment of the degree of Bachelor of Engineering in COMPUTER ENGINEERING (8th Semester) of Gujarat Technological University, Ahmedabad during the academic year 2020-21.

Date of Submission: 30/04/2021

**Guided By:**

**Prof. Bansari Thakkar**  
Asst. Prof. CE Department  
AIT

**Dr. Ajay.N.Upadhyaya**  
Head of CE-IT  
AIT



## GUJARAT TECHNOLOGICAL UNIVERSITY

CERTIFICATE FOR COMPLETION OF ALL ACTIVITIES AT ONLINE PROJECT PORTAL

B.E. SEMESTER VIII, ACADEMIC YEAR 2020-2021

Date of certificate generation : 04 May 2021 (18:29:40)

This is to certify that, *Patel Tirth Pravinbhai* ( Enrolment Number - 170020107041 ) working on project entitled with *Sentiment Analysis* from *Computer Engineering* department of *AHMEDABAD INSTITUTE OF TECHNOLOGY, GOTA, AHMEDABAD* had submitted following details at online project portal.

|                                 |                  |
|---------------------------------|------------------|
| Periodic Progress Reports (PPR) | <b>Completed</b> |
| Business Model Canvas (Image)   | <b>Completed</b> |
| Business Model Canvas (Report)  | <b>Completed</b> |
| Patent Drafting Exercise (PDE)  | <b>Completed</b> |
| Final Plagiarism Report         | <b>Completed</b> |
| Final Project Report            | <b>Completed</b> |

Name of Student : Patel Tirth Pravinbhai

Name of Guide : Miss. Bansari rameshbhai thakkar

Signature of Student : \_\_\_\_\_

\*Signature of Guide : \_\_\_\_\_

**Disclaimer :**

This is a computer generated copy and does not indicate that your data has been evaluated. This is the receipt that GTU has received a copy of the data that you have uploaded and submitted as your project work.

\*Guide has to sign the certificate, Only if all above activities has been Completed.



## GUJARAT TECHNOLOGICAL UNIVERSITY

### CERTIFICATE FOR COMPLETION OF ALL ACTIVITIES AT ONLINE PROJECT PORTAL

B.E. SEMESTER VIII, ACADEMIC YEAR 2020-2021

Date of certificate generation : 05 May 2021 (15:08:17)

This is to certify that, *Sahani Ravikumar Ramsajivan* (Enrolment Number - 170020107049) working on project entitled with *Sentiment Analysis* from *Computer Engineering* department of *AHMEDABAD INSTITUTE OF TECHNOLOGY, GOTA, AHMEDABAD* had submitted following details at online project portal.

|                                 |                  |
|---------------------------------|------------------|
| Periodic Progress Reports (PPR) | <b>Completed</b> |
| Business Model Canvas (Image)   | <b>Completed</b> |
| Business Model Canvas (Report)  | <b>Completed</b> |
| Patent Drafting Exercise (PDE)  | <b>Completed</b> |
| Final Plagiarism Report         | <b>Completed</b> |
| Final Project Report            | <b>Completed</b> |

Name of Student : Sahani Ravikumar  
Ramsajivan

Name of Guide : Miss. Bansari rameshbhai  
thakkar

Signature of Student : \_\_\_\_\_

\*Signature of Guide : \_\_\_\_\_

#### Disclaimer :

This is a computer generated copy and does not indicate that your data has been evaluated. This is the receipt that GTU has received a copy of the data that you have uploaded and submitted as your project work.

\*Guide has to sign the certificate, Only if all above activities has been Completed.

## PLAGIARISM SCAN REPORT

Words 994 Date April 30,2021

Characters 6680 Excluded URL

**9%**  
Plagiarism

**91%**  
Unique

**5**  
Plagiarized  
Sentences

**51**  
Unique Sentences

## Content Checked For Plagiarism

Sentiment Analysis is deep mining of text which extracts subjective information in source material and it also helps business to build product or service while monitoring online conversations.

In the age of artificial intelligence and machine learning, competition is between best and best. So, inorder to gain control over market, it is essential to understand market condition especially during covid-19 situation. For that sentiment of market is very important and sentiment of market is what consumer think of certain product.

Apply the following Machine Learning algorithms to arrive at the best result:

- K-Means Clustering
- Support Vector Machine
- Logistic Regression
- K Nearest Neighbours

There is a need of proper and formal comparisons between these results arrived through features and classification in order to select the best features and most efficient classification techniques for particular applications.

The scope of the project is to provide a simple web app which helps to extracts people's sentiment feelings toward certain services, products, organizations, political or non-political topics and any influential people on social media.

The project aims to:

Provide an accurate sentiment analysis results. fast and easy to use web-based tool.

Providing a good entertainment visualization capabilities. Having options in term of filtering and viewing information according to user's needs. Technology Review

Frontend Technology: HTML, CSS, Javascript

Backend Technology : Python, Flask, PHP

Libraries: Numpy, Pandas, Scikit-learn, Matplotlib, Seaborn, Scipy

Tools : VScode editor, Jupyter Notebook, Google Colab

Python offers concise and readable code. Moreover there are many built-in predefined in python which can be very useful in sentiment analysis. It allows programmers to focus on business logic rather than on focusing on memory management and all that lower level stuff. Python code is readable by humans, which makes it easier to build models for Artificial Intelligence.

Flask is a light weight framework which uses python. It is used to connect jupyter notebook to frontend. Flask is used as it uses the same language as jupyter notebook.

NumPy is a very popular python library for large multi-dimensional array and, with the help of a large collection of complex mathematical functions. It is very useful for scientific computations in Machine Learning. High-end libraries like TensorFlow uses NumPy internally for manipulation of Tensors.Pandas is a popular Python library for data analysis.

Pandas was developed specifically for data extraction and machine learning

VScode is the most configurable and powerful editor with debugger. It is backed by Microsoft and moreover it is open source. Frontend is developed on this editor.

It helps a data scientist to document the process while developing the analysis process. One can also capture the result as the part of the notebook. With the help of jupyter notebooks, we can share our work with a Internet also. Google Colab is Jupyter Notebook running on more powerful server of google. It provides powerful GPU and CPU for high resource taking programs. It can be very useful if PC performance is not sufficient to run the code. Both jupyter

**Limitation of Existing System:**

- The size of our data will allow us to perform cross validation experiments and check variance in performance of model
- Usability of project:
  - It can be useful to business for analysing how to market product and what is the sentiment of people, how they will respond to product. It can be useful for predicting the tweets for a certain group of people.
  - It can provide a competitive advantage for business opportunities and product possibilities.

**Innovative ideas incorporated in system:**

- Tweet Sentiments: Returns the Sentiments of tweets.
- Text Processing: Performs sentiment analysis, parts of speech tagging and chunking, phrase extraction and named entity recognition.
- ML Analyzer: Preforms text classification, article summarization, sentiment analysis, extracts stock symbol, extracts person names, language detection, and extracts locations.
- It can be useful to business for analysing how to market product and what is the sentiment of people, how they will respond to product. It can be useful for predicting the tweets for a certain group of people.
- It can provide a competitive advantage for business opportunities and product possibilities.

It is supported in all web browsers viz google chrome, Microsoft edge, apple safari, mozilla firefox. It is written in Python3, HTML5, CSS3 and flask framework.

**Non-functional Requirements**

We will keep on detecting if our system hanged or an operating system error occurred. Also detecting the performance of the system in terms of the efficiency of integration of the different components

**Safety Requirements:** For the safety requirements nothing but an operation of weekly backups for the data base should take place.

**Security and Privacy Requirements:** There are no specific security requirements, anyone can access and use the portal but only authorized persons who are allowed to use and access the database, web pages and the product engine.

**Software Quality Attributes:**

The solution is to reach the level of accuracy. But also keeping in mind that this prototype version is for proving the concept of the project

The system we designed is used to find the opinion of the people based on twitter data.

We have used Tweepy to extract tweet from twitter. This API is useful to track live tweets and make live twitter sentiment analysis.

We have used Streamlit for web App as it is very popular among machine learning projects. In streamlit, we can focus on Algorithm and all the html, css tools are provided by streamlit.

For hosting, we have used Heroku, it is very easy to understand for us and hosting from github is very easy. We have used github repository for storing the project.

Sometimes algorithm fails to identify hate speech and violent tweets. Neutral tweets can sometimes become positive or negative based on the number of likes.

Live Twitter Sentiment Analysis is not possible with static dataset.

**Future Enhancement**

- Potential improvement can be made to our data collection and analysis method.
- Integrate APIs for live twitter tracking.

| Sources   | Similarity |
|---|------------|
| Jupyter Notebooks Machine Learning - 02/2021<br><br>One can also capture the result as the part of the notebook. With the help of jupyter notebooks, we can share our work with...<br><br><a href="https://www.coursef.com/jupyter-notebooks-machine-learning">https://www.coursef.com/jupyter-notebooks-machine-learning</a>   | 5%         |
| Assignment 01.docx - Shaheed Zulfikar Ali Bhutto Institute of ...<br><br>Also detecting the performance of the system in terms of the efficiency of integration of the different components Software Engineering BSCS 5 SZABIST-ISB Shaheed Zulfikar Ali Bhutto Institute of Science & Technology COMPUTER SCIENCE DEPARTMENT o Reliability The solution should provide reliability to the user that the product will run with all ...<br><br><a href="https://www.coursehero.com/file/86037423/Assignment-01docx/">https://www.coursehero.com/file/86037423/Assignment-01docx/</a> | 3%         |
| For the data gathering twitter is the only source and using Streaming ...<br><br>Safety Requirements For the safety requirements nothing but an operation of weekly backups for the data base should take place. 5.3. Security and Privacy ...  | 5%         |

<https://www.coursehero.com/file/p0rbgoj/Expected-Outcome-What-are-the-suitable-types-of-n-grams-model-to-use-in/>

**Stock Market Analysis and Prediction - SlideShare**

• Future scope of improvement Potential improvement can be made to our data collection and analysis method. Future research can be done with possible improvement such as more refined data and more accurate algorithm. Implementation of discussion forums and economic news portal including other sector apart from hydropower and going in national level.

<https://www.slideshare.net/anilsth01/stock-market-analysis-and-prediction>

## **UNDERTAKING ABOUT ORIGINALITY OF WORK**

We hereby certify that we are the sole authors of this UDP project report and that neither any part of this UDP project report nor the whole of the UDP Project report has been submitted for a degree by other student(s) to any other University or Institution.

We certify that, to the best of our knowledge, the current UDP Project report does not infringe upon anyone's copyright nor violate any proprietary rights and that any ideas, techniques, quotations or any other material from the work of other people included in our UDP Project report, published or otherwise, are fully acknowledged in accordance with the standard referencing practices. Furthermore, to the extent that we have included copyrighted material that surpasses the boundary of fair dealing within the meaning of the Indian Copyright (Amendment) Act 2012, we certify that we have obtained a written permission from the copyright owner(s) to include such material(s) in the current UDP Project report and have included copies of such copyright clearances to our appendix.

We have checked the write up of the present UDP Project report using anti-plagiarism database and it is in the allowable limit. In case of any complaints pertaining to plagiarism, we certify that we shall be solely responsible for the same and we understand that as per norms, University can even revoke BE degree conferred upon the student(s) submitting this UDP Project report, in case it is found to be plagiarized.

**Team:**

| Enrollment Number   | Name               | Signature |
|---------------------|--------------------|-----------|
| <b>170020107041</b> | <b>Tirth Patel</b> |           |
| <b>170020107049</b> | <b>Ravi Sahani</b> |           |

Place: Ahmedabad, Gujarat

Date: 30/04/2021

Guide Name: Asst.Prof. Bansari Thakkar

Signature of Guide: \_\_\_\_\_

## **ACKNOWLEDGEMENT**

No task can be accomplished without proper support, guidance and appraisal. We are highly thankful to many people who contributed either directly or indirectly for this project and provided their invaluable cooperation to us to complete it.

We would like to thank our **Head of Department, Dr. Ajay.N.Upadhyaya** and our faculty members of **Computer Engineering** for constantly guiding and showing us the correct way to reach towards the desired goal.

Then we would like to thank our **Internal guide Prof. Bansari Thakkar**, Asst. Prof, CE Department, who has helped us out in each and every phase of the project and without their support and guidance the project would not have been completed successfully.

The wholehearted help and co-operation by our friends is gratefully acknowledged.

Tirth Patel (170020107041),

Ravi Sahani (170020107049)

## **Abstract**

Sentiment Analysis is deep mining of text which extracts subjective information in source material and it also helps business to build product or service while monitoring online conversations.

In the age of artificial intelligence and machine learning, competition is between best and best. So, inorder to gain control over market, it is essential to understand market condition especially during covid-19 situation. For that sentiment of market is very important and sentiment of market is what consumer think of certain product.

Sentiment analysis is the interpretation and classification of emotions (positive, negative and neutral) within text data using text analysis techniques. Sentiment analysis tools allow businesses to identify customer sentiment toward products, brands or services in online feedback. The algorithm proposed works on Twitter Data, primarily it collects the tweets and then study it with the help of different statistical computing procedures.

Apply the following Machine Learning algorithms to arrive at the best result:

- K-Means Clustering
- Support Vector Machine
- Logistic Regression
- K Nearest Neighbours

## **List of Figures**

| Serial Number | Figure                                |
|---------------|---------------------------------------|
| 1             | Architectural diagram for the product |
| 2             | Use Case diagram                      |
| 3             | System Flow diagram                   |
| 4             | System diagram                        |
| 5             | Flow Chart                            |
| 6             | System Activity diagram               |
| 7             | Data Flow Diagram                     |

## **List of Tables**

| Serial Number | Table     |
|---------------|-----------|
| 1             | Tweet     |
| 2             | User      |
| 3             | Place     |
| 4             | Retweet   |
| 5             | Retweeted |

## TABLE OF CONTENTS

|                   |   |           |
|-------------------|---|-----------|
|                   | <b>College Certificate</b>                      | <b>2</b>  |
|                   | <b>Completion of all Activities Certificate</b> | <b>3</b>  |
|                   | <b>Plagiarism Certificate</b>                   | <b>5</b>  |
|                   | <b>Undertaking about Originality of Work</b>    | <b>7</b>  |
|                   | <b>Acknowledgement</b>                          | <b>8</b>  |
|                   | <b>Abstract</b>                                 | <b>9</b>  |
|                   | <b>List of Figures</b>                          | <b>10</b> |
|                   | <b>List of Tables</b>                           | <b>10</b> |
| <b>Chapter: 1</b> | <b>Introduction</b>                             | <b>12</b> |
|                   | 1.1 Project Summary                             |           |
|                   | 1.2 Project Purpose                             |           |
|                   | 1.3 Project Scope                               |           |
|                   | 1.4 Project Objectives                          |           |
|                   | 1.5 Technology Review                           |           |
|                   | 1.6 Literature Review                           |           |
| <b>Chapter: 2</b> | <b>Project Management</b>                       | <b>17</b> |
|                   | 2.1 Project Planning                            |           |
|                   | 2.2 Project Scheduling                          |           |
|                   | 2.3 Risk Management                             |           |
|                   | 2.4 Canvas Designing                            |           |
|                   | 2.4.1 AEIOU Summary Canvas                      |           |
|                   | 2.4.1.1 Activities                              |           |
|                   | 2.4.1.2 Environment                             |           |

|                   |   |           |
|-------------------|---|-----------|
|                   | <p>2.4.1.3 Interactions</p> <p>2.4.1.4 Objects</p> <p>2.4.1.5 Users</p> <p>2.4.2 Empathy Summary Canvas</p> <p>2.4.3 Ideation Canvas</p> <p>2.4.4 Product Development Canvas</p>  |           |
| <b>Chapter: 3</b> | <b>System Requirements Study</b>  | <b>22</b> |
|                   | <p>3.1 User Characteristics</p> <p>3.2 Hardware and Software Requirements</p> <p>3.3 Specific Requirements</p> <ul style="list-style-type: none"> <li>3.3.1 Functionality</li> <li>3.3.2 Usability</li> <li>3.3.3 Reliability</li> <li>3.3.4 Performance</li> <li>3.3.5 Supportability</li> <li>3.3.6 Design Constraints</li> </ul> <p>3.4 Assumptions and Dependencies</p> |           |
| <b>Chapter: 4</b> | <b>System Analysis</b>  | <b>26</b> |
|                   | <p>4.1 Requirements</p> <p>4.2 Diagrams</p> <ul style="list-style-type: none"> <li>1) Use Case Diagram</li> <li>2) System Flow Diagram</li> <li>3) System Design</li> <li>4) Flow Chart</li> <li>5) System Activity Diagram</li> </ul>  |           |

|                   |   |           |
|-------------------|---|-----------|
|                   | 6) Data Flow Diagram  |           |
| <b>Chapter: 5</b> | <b>System Design</b><br><br>5.1 Database Design<br><br>5.2 Input-Output Design  | <b>34</b> |
| <b>Chapter: 6</b> | <b>Limitation and Future Enhancement</b><br><br>6.1 Limitation<br><br>6.2 Future Enhancement  | <b>37</b> |
| <b>Chapter: 7</b> | <b>Conclusion and Discussion</b><br><br>7.1 Self Analysis of Project Viabilities<br><br>7.2 Problem Encountered and Possible Solutions<br><br>7.3 Summary of Project work | <b>38</b> |
| <b>Chapter: 8</b> | <b>References</b>   | <b>39</b> |
|                   | <b>Appendix 1: Periodic Progress Report</b>   | <b>40</b> |
|                   | <b>Appendix 2: Design Engineering Canvas</b>  | <b>42</b> |
|                   | <b>Appendix 3: PSAR Report</b>  | <b>46</b> |
|                   | <b>Appendix 4: Business Model Canvas and Report</b>   | <b>56</b> |
|                   | <b>Appendix 5: Patent Drafting Exercise</b>   | <b>65</b> |

## **Chapter 1: Introduction**

### **1.1 Project Summary**

Sentiment analysis in the domain of micro-blogging is a relatively new research topic so there is still a lot of room for further research in this area. Decent amount of related prior work has been done on sentiment analysis of user reviews , documents, web blogs/articles and general phrase level sentiment analysis . These differ from twitter mainly because of the limit of 140 characters per tweet which forces the user to express opinion compressed in very short text.

The best results reached in sentiment classification use supervised learning techniques such as Naive Bayes and Support Vector Machines, but the manual labelling required for the supervised approach is very expensive. Some work has been done on unsupervised and semi-supervised approaches, and there is a lot of room of improvement.

Various researchers testing new features and classification techniques often just compare their results to base-line performance. There is a need of proper and formal comparisons between these results arrived through different features and classification techniques in order to select the best features and most efficient classification techniques for particular applications.

### **1.2 Project Purpose**

The purpose of the project is to empower companies, individual, government, tech giants to identify their customer or in laymen's term find their target customer without wasting resources on expensive surveys and more important is that it is done in complete ethical way.

The main aim is that to develop a tool that can allow users to use a simple search bar to search for any services, products or any political topics and the engine of that tool is to crawl over the internet collecting all comments, reviews, tweets or even notes in blogs related to the user's search keyword. Then perform an intelligent processing technique to extract the true meanings of the people's comments and to decide and classify them in terms of positive, negative or neutral thus to know the majority of people like or dislike the desired topic. More specifically providing people's feelings regarding certain topics with high accuracy will lead to a better decision making.

The purpose of the prototype is to demonstrate the concept and to deliver operational and functional services for testing purposes. As for initial Twitter will be the only source of data for the prototype and then integration will be needed to include more sources like facebook, news websites and blogs.

### 1.3 Project Scope

The scope of the project is to provide a user friendly web based product that extracts people's sentiment feelings toward certain services, products, organizations, political or non-political topics and any influential people on social media. In this project phase which aims at developing a filed prototype, emphasis will be put on English tweets from Twitter in the political domain.

The project aims to:

- Provide an accurate sentiment analysis results.
- Smooth, fast, efficient, reliable and easy to use web-based tool.
- Providing a user friendly menu and good entertainment visualization capabilities.
- Having a plenty of options in term of filtering and viewing information according to user's needs.

### 1.4 Project Objectives

The objectives of this project are:

- To implement an algorithm for automatic classification of text into positive and negative.
- Sentiment Analysis to determine the attitude of the mass is positive, negative or neutral towards the subject of interest.
- Graphical representation of the sentiment in form of Pie-Chart and Scatter Plot.
- Real time data analysis makes it possible for business organizations to keep track of their services and generates opportunities to promote, advertise and improve from time to time. Through comprehensive analysis, businesses gain valuable insights towards their customers.
- It can provide a competitive advantage- A trove of business opportunities and product possibilities.
- Gives predictions of consumer trends so companies can develop strategies to gain an advantage.
- It can help revitalize a business' brand. Companies can quantify people's perception about their products or services, marketing strategies and their customer experience. When used right, it can help companies develop engaging marketing strategies to improve people's perception about their brands.
- It can help enhance customer experience towards brands and business.

## 1.5 Technology Review

Frontend Technology: Streamlit, HTML

Backend Technology : Python

Libraries: Numpy, Pandas, Scikit-learn, Matplotlib, Seaborn, tweepy, openpyxl

Tools : VScode editor, Jupyter Notebook, Google Colab

Python offers concise and readable code. While complex algorithms and versatile workflows stand behind machine learning and AI, Python's simplicity allows developers to write reliable systems. Developers get to put all their effort into solving an ML problem instead of focusing on the technical nuances of the language.

Moreover there are many built-in predefined libraries in python which can be very useful in sentiment analysis and machine learning. It allows programmers to focus on business logic rather than on focusing on memory management and all that lower level stuff. Additionally, Python is appealing to many developers as it's easy to learn. Python code is understandable by humans, which makes it easier to build models for machine learning.

Flask is a light weight framework which uses python. It is used to connect jupyter notebook to frontend. Flask is used as it uses the same language as jupyter notebook.

NumPy is a very popular python library for large multi-dimensional array and matrix processing, with the help of a large collection of high-level mathematical functions. It is very useful for fundamental scientific computations in Machine Learning. High-end libraries like TensorFlow uses NumPy internally for manipulation of Tensors. SciPy is a very popular library among Machine Learning enthusiasts as it contains different modules for optimization, linear algebra, integration and statistics. SciPy is also very useful for image manipulation. Scikit-learn is one of the most popular ML libraries for classical ML algorithms. Scikit-learn supports most of the supervised and unsupervised learning algorithms. Scikit-learn can also be used for data-mining and data-analysis, which makes it a great tool who is starting out with ML. Pandas is a popular Python library for data analysis. Pandas comes handy as it was developed specifically for data extraction and preparation. It provides high-level data structures and wide variety tools for data analysis. It provides many inbuilt methods for groping, combining and filtering data.

VScode is the most configurable and powerful editor with debugger. It is backed by Microsoft and moreover it is open source. Frontend is developed on this editor.

Jupyter Notebooks can illustrate the analysis process step by step by arranging the stuff like code, images, text, output etc. in a step by step manner. It helps a data scientist to document the thought process while developing the analysis process. One can also capture the result as the part of the notebook. With the help of jupyter notebooks, we can share our work with a peer also. Google Colab is Jupyter Notebook running on more powerful server of google. It provides powerful GPU and CPU for high resource taking programs. It can be very useful if PC performance is not sufficient to run the code. Both jupyter notebook and google colab are browser-based application.

## 1.6 Literature Review

This section summarizes some of the scholarly and research works in the field of Machine Learning and data mining to analyse sentiments on the Twitter and preparing prediction model for various applications. As the available social platforms are shooting up, the information is becoming vast and can be extracted to turn into business objectives, social campaigns, marketing and other promotional strategies as explained in [4]. The benefit of social media to know public opinions and extract their emotions are considered by authors in [2] and explained how twitter gives advantage politically during elections. Further, the concept of the hashtag is used for text classification as it conveys emotion in few words.

They suggested how previous research work suffered from lack of training set and misses some features of target data. They opted two stage approach for their framework- first preparing training data from twitter using mining conveying relevant features and then propounding the Supervised Learning Model to predict the results of Elections held in USA in 2016. After collecting and pre-processing the tweets, training data set was created first by manual labelling of hashtags and forming clusters, next by using online Sentimental Analyzer VADER which outputs the polarity in percentage.

This approach reduced the number of tweets or training set and further they applied Support Vector Machine and Naive Bayes classification algorithm to determine the polarity of tweets. Multistage Classification approach was used where an entity classifier receives general class of tweets and categorise them with respect to individual candidates for comparison. The metric they used to determine the winner was the “PvT ratio” which is Positive number of tweets to total count of tweets for respective candidate.

Sentiment Analysis by researchers Imran et al. [1] exploited the technology 'Apache Spark' for fast streaming of tweets and presented the approach Stream Sensing to handle real time data in unstructured and noisy form. They conducted the approach on twitter data to find some useful and interesting trends which further can be generalized to any real-time text stream. Unsupervised learning approach is used to locate interesting patterns and trends from tweets processed on Apache Spark.

### Study of current Existing System:

- Sentiment analysis has been handled as a Natural Language Processing task at many levels of granularity.
- Starting from being a document level classification, it has been handled at the sentence level and more recently at the phrase level.
- Microblog data like Twitter, on which users post real time reactions to and opinions about “everything”, poses newer and different challenges.

**Limitation of Existing System:**

- They build models using Naive Bayes, Max Ent and Support Vector Machines (SVM), and they report SVM outperforms other classifiers. In terms of feature space, they try a Unigram, Bigram model in conjunction with parts-of-speech (POS) features.
- The data they use for training and testing is collected by search queries and is therefore biased.
- Our data will be a random sample of streaming tweets unlike data collected by using specific queries.
- The size of our hand-labeled data will allow us to perform cross validation experiments and check forth variance in performance of the classifier across folds.

**Usability of project:**

- It can be useful to business for analysing how to market product and what is the sentiment of people, how they will respond to product. It can be useful for predicting the tweets for a certain group of people.
- It can provide a competitive advantage for business opportunities and product possibilities.

**Innovative ideas incorporated in system:**

- Tweet Sentiments: Returns the Sentiments of tweets.
- Text Processing: Performs sentiment analysis, parts of speech tagging and chunking, phrase extraction and named entity recognition.
- ML Analyzer: Preforms text classification, article summarization, sentiment analysis, extracts stock symbol, extracts person names, language detection, and extracts locations.

## **Chapter 2: Project Management**

### **2.1 Project Planning**

- Project planning is part of project management, which relates to the use of schedules such as Gantt charts to plan and subsequently report progress within the project environment.
- Initially, the project scope is defined and the appropriate methods for completing the project are determined. Following this step, the durations for the various tasks necessary to complete the work are listed and grouped into a work breakdown structure.
- Project planning is often used to organize different areas of a project, including project plans, workloads and the management of teams and individuals.

### **2.2 Project Scheduling**

- Project Scheduling is the culmination of a planning activity that is primary component of software project management.
- When combined with estimation methods and risk analysis, scheduling, establishes a road map for the project management.
- Scheduling begins with the process composition. The characteristics of the project are used to adapt an appropriate task set for the work to be done.
- The task network is used to compute the critical project path, a time line chart and a variety of project information.

### **2.3 Risk Management**

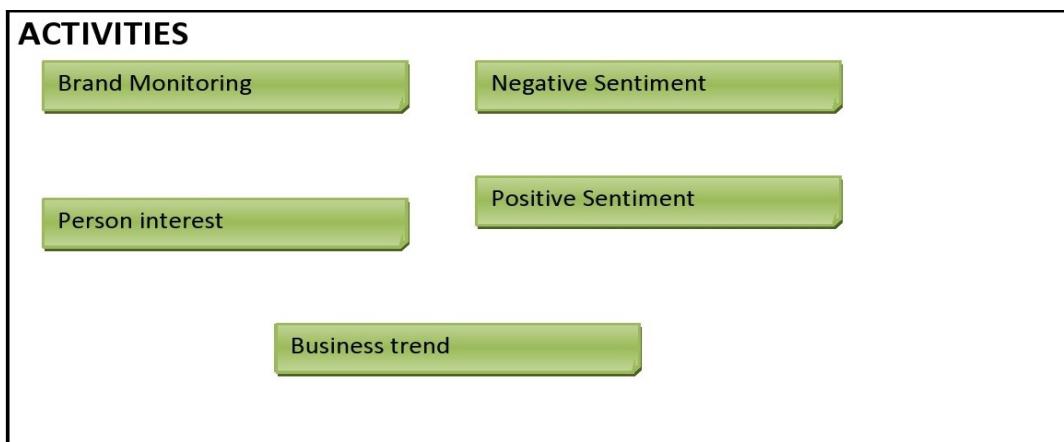
- Project Risk: Project risks concern differ forms of budgetary, schedule, personnel, resource, and customer-related problems. Sometimes budget increases than the original budget, it can hinder future development scope.
- Technical Risk: Technical risks concern potential method, implementation, interfacing, testing, and maintenance issue. It also consists of an ambiguous specification, incomplete specification, changing specification, technical uncertainty, and technical obsolescence. Ambiguous database can lead to failure of backend processing and queries can be complicated.
- Business Risk: This type of risks contain risks of building an excellent product that no one need, losing budgetary or personnel commitments, etc. Sometimes wrong analysis of market can lead to development of product which is not needed. Wrong sentiment analysis can lead to business failure or loss of millions of assets.

## 2.4 Canvas Designing

### 2.4.1 AEIOU Summary Canvas

| AEIOU Summary     |                          |                        | Team No:99626 |
|-------------------|--------------------------|------------------------|---------------|
| ENVIRONMENT       | INTERACTION              | OBJECTS                |               |
| MNC               | Monitoring               | Jupyter Notebooks      |               |
| News Channels     | Marketing                | TextBlob               |               |
| Political Parties | People opinions          | Scikit - Learn         |               |
| Industries        | E - commerce advertising | Binder                 |               |
| ACTIVITIES        |                          | USERS                  |               |
| Brand Monitoring  | Negative Sentiment       | Business Analyst       |               |
| Person interest   | Positive Sentiment       | Political Campaigners  |               |
| Business trend    |                          | News Reporter          |               |
|                   |                          | For monitoring someone |               |

#### 2.4.1.1 Activities



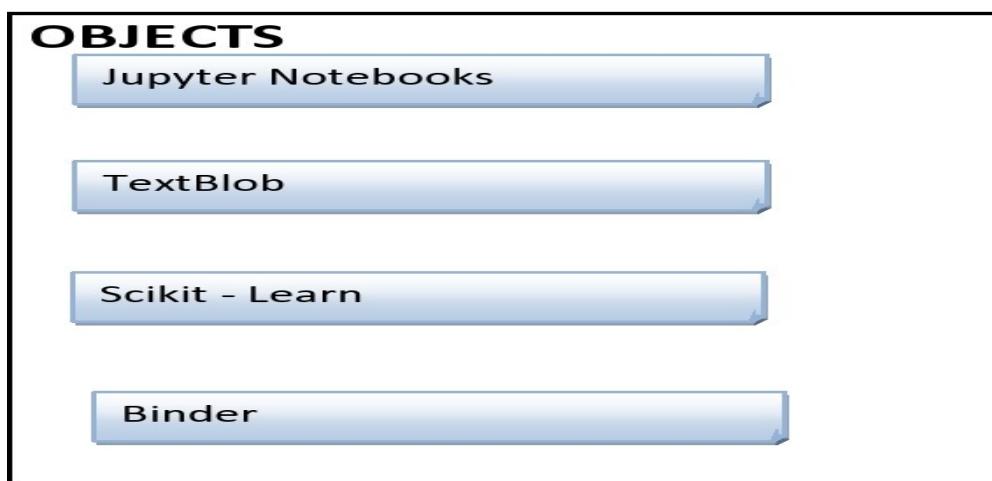
#### 2.4.1.2 Environment



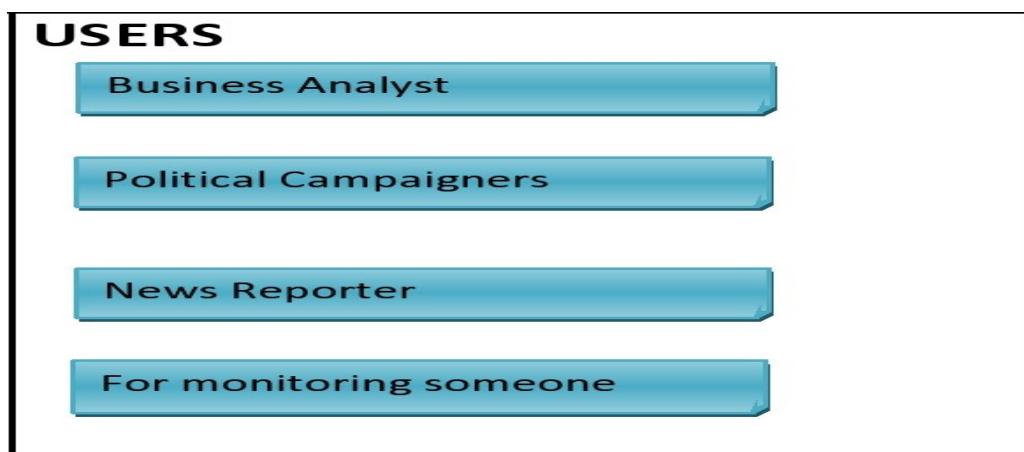
#### 2.4.1.3 Interactions



#### 2.4.1.4 Objects



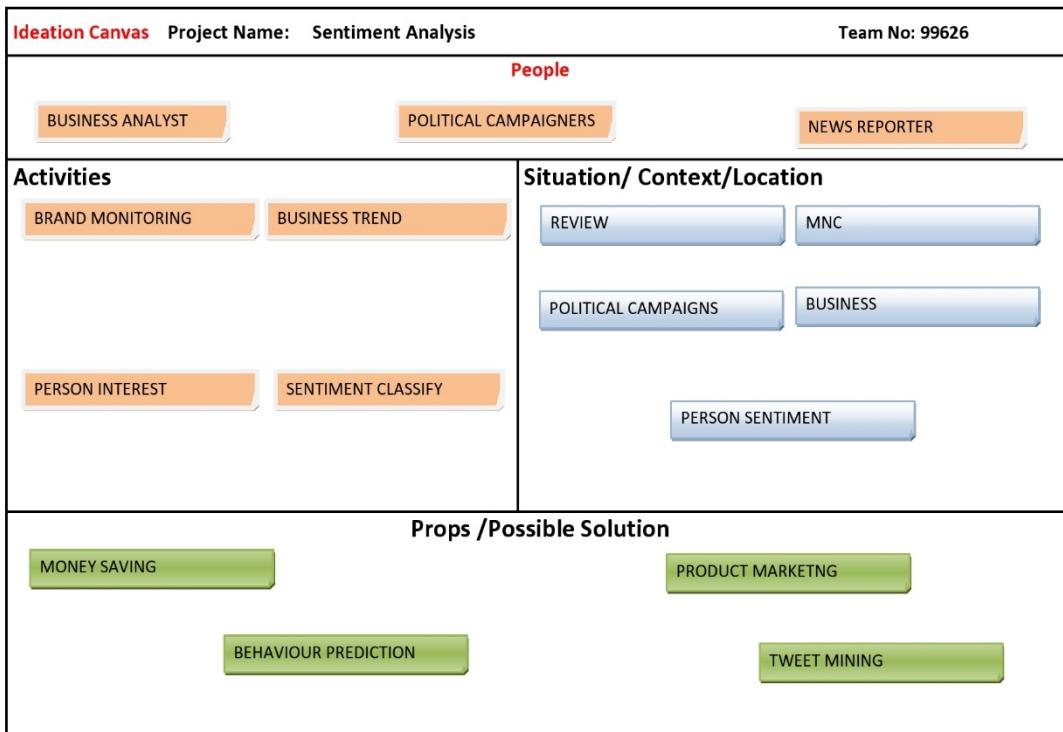
#### 2.4.1.5 Users



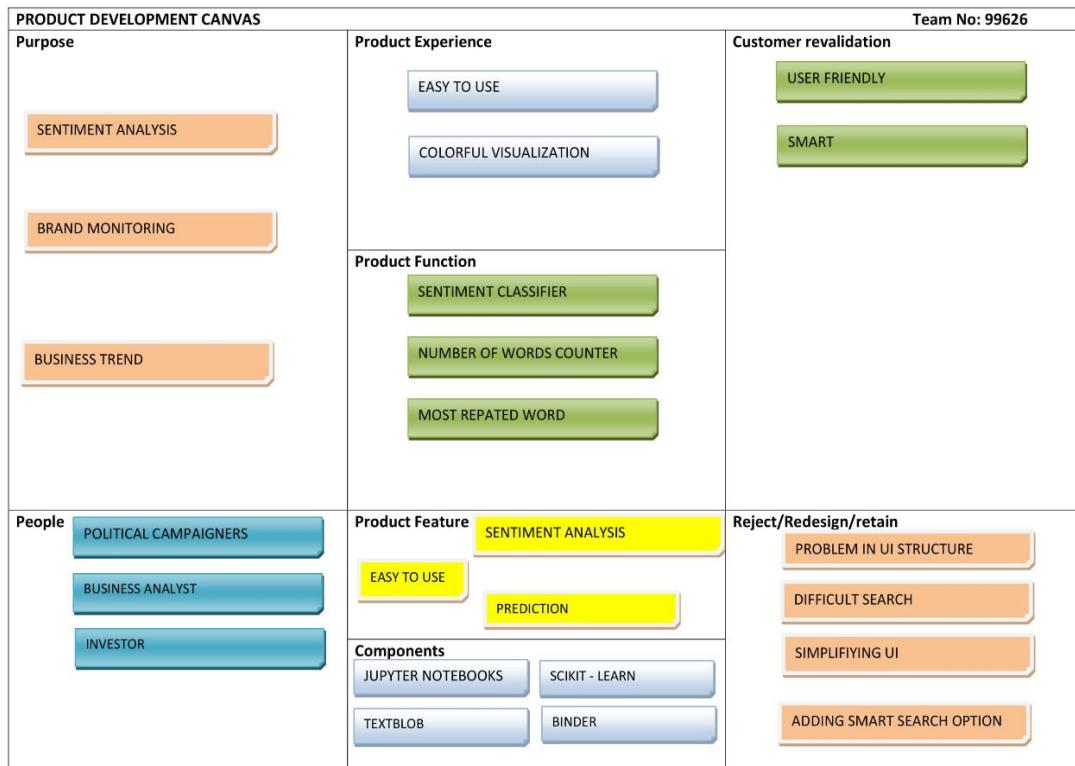
#### 2.4.2 Empathy Canvas

|  |            |   |                          |
|--|------------|---|--------------------------|
| Design For   | VIER       | Design By   | Tirth Patel, Ravi Sahani |
| Date   | 06/10/2020 | Version   | 1/1                      |
| <b>USER</b> <div style="background-color: yellow; padding: 5px;">POLITICAL CAMPAIGNERS</div> <div style="background-color: yellow; padding: 5px;">BUSINESS ANALYST</div>   |            | <b>STAKEHOLDERS</b> <div style="background-color: #ADD8E6; padding: 5px;">ENTREPRENEUR</div> <div style="background-color: #ADD8E6; padding: 5px;">INVESTOR</div> <div style="background-color: #ADD8E6; padding: 5px;">BUSINESSMAN</div> |                          |
| <b>ACTIVITIES</b> <div style="display: flex; justify-content: space-around;"> <div style="background-color: #FFB6C1; padding: 5px;">BRAND MONITORING</div> <div style="background-color: #FFB6C1; padding: 5px;">BUSINESS TREND</div> </div> <div style="display: flex; justify-content: space-around;"> <div style="background-color: #FFB6C1; padding: 5px;">PERSON INTEREST</div> <div style="background-color: #FFB6C1; padding: 5px;">SENTIMENTS FINDING</div> </div> <div style="text-align: center; margin-top: 10px;">PEOPLE OPINIONS</div>  |            |   |                          |
| <b>STORY BOARDING</b> <p><b>HAPPY</b></p> <div style="border: 1px solid black; padding: 5px; background-color: #F0F8FF;">Imagine a new mobile phone is released. A sentiment analysis of social media posts about the phone may give a company valuable, real-time insight into how it's performing.</div> <p><b>HAPPY</b></p> <div style="border: 1px solid black; padding: 5px; background-color: #F0F8FF;">The overall sentiment of a phrase can then be scored by looking at all the words in the post. For example, the average score for the post "My momma always said 'life is like a box of chocolates'" is an above-average 6.02, according to this dictionary, suggesting it expresses a positive feeling.</div> <p><b>SAD</b></p> <div style="border: 1px solid black; padding: 5px; background-color: #F0F8FF;">Given that sentiment analysis often relies on mining social media posts, it raises major ethical concerns, and this debate only beginning. Yet the complex nature of language and meaning makes it prone to error.</div> <p><b>SAD</b></p> <div style="border: 1px solid black; padding: 5px; background-color: #F0F8FF;">Take the phrase, "May the force be with you", which scores 5.35 using our dictionary's analysis. For any Star Wars fan, it is of course a hugely positive phrase. But it scored modestly in our test because the word "force" is rated a below-average 4.0.</div> |            |   |                          |

### 2.4.3 Ideation Canvas



### 2.4.4 Product Development Canvas



## **Chapter 3: System Requirements Study**

### **3.1 User Characteristics**

This part is to identify various user classes that we anticipate will use the web application. User classes will be differentiated based on the use, product functions and features, technical expertise, security and privilege levels and educational level. The solution is intended to be used by three main different user classes; system administrators, system operators and customers or regular users. No special knowledge or skills should be assumed for the part of the regular users. Users are not expected to learn or remember a set of commands in order to start using the application. The prototype application will be only a web based and then for the product versions there will be a desktop versions, smart phones and smart Tablets.

The following clearly describes a visionary role for each participant.

- **Users:** users with no particular knowledge needed, users who are interested to use the tool looking for knowing people's thoughts about a desired topic.
- **Advanced end users:** advanced users are those who have valuable input and feedbacks. Users who are more familiar with informative sites and can use our features efficiently. These valuable feeds will lead to enhancement of users' satisfaction.
- **System Operators:** Maintains for the functional interface of the application and troubleshooting issues, Suggest possible updates and identifying renewal application needs, Coordinate with service providers and infrastructure vendors, Coordinate and communicate with system administrators.
- **System Administrators:** Develop and maintain installation and configuration procedures and operational requirements, Perform weekly/monthly backup operations, ensuring all required files and data are successfully backed up, Repair and recover from hardware or software failures, Coordinate and communicate with system operators.

### **3.2 Hardware and Software Requirements**

**Software Requirements:** For the prototype we will launch the portal over the internet and other than the hardware specified in the hardware interface section, the software requirements are to support windows operating system with support to MySQL, apache and PHP servers. For the data gathering twitter is the only source and using Streaming API that offers high throughput. Using this API is perfect because we can retrieve real time information and also this continuous stream will be retrieved with no end and capturing all the messages in the stream without missing any information. The information retrieved in JSON format.

**Hardware Requirements:** The solution makes extensive use of several hardware devices. These devices include:

- MySQL database server with intensive use of memory space.
- PHP server with high performance and intensive use for CPU usage.
- Windows and Linux users' computers.
- PC with minimum 8 GB RAM or more and processor having clock speed of more than 2.7 GHz or more.

### 3.3 Specific Requirements

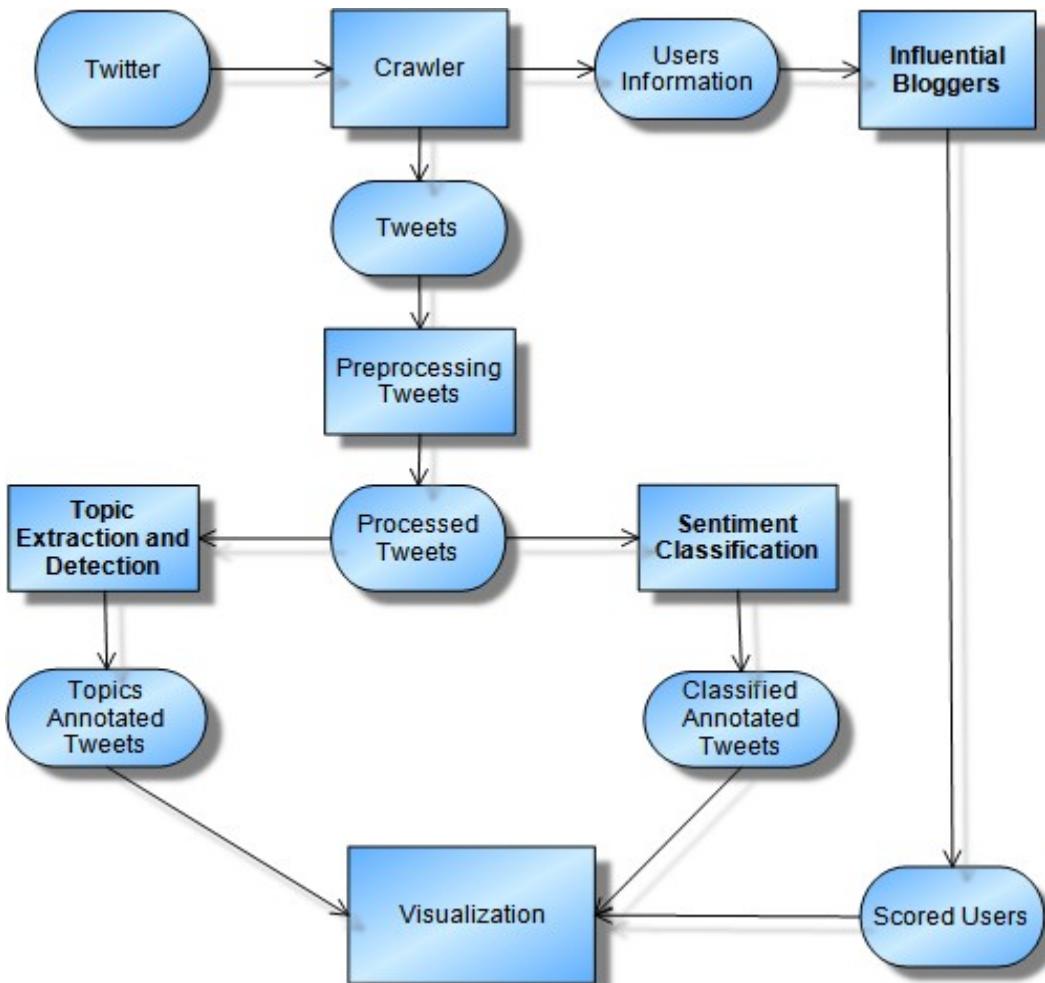
#### 3.3.1 Functionality

The architecture diagram of the tool is shown in figure 1. This tool will provide the following functions:

**Topic Extraction:** This part is considered a key stone in the project as it detects and extracts topics titles from the tweets. Using hash tags is not informative enough about the topic of sentiment the author mentions in his/her tweet. Our approach goes as follows, first we do pre-processing which includes removal of stop words that occur frequently in the tweets but have no relevant meaning, then generate the feature vector. The features used are n-grams, unigram, bigram, and trigrams, and some named entities that are extracted from the crawled tweets. The main step is to cluster related tweets together using similarity measures so we can have multiple clusters each has one topic. Afterwards key-phrase extraction is used on each cluster to extract the key-phrases that are candidates to be title topics. Clusters that result in similar key phrase are merged together and this key phrase has higher weight to be the topic title.

**Sentiment Classification:** Sentiment classification is the primary module of the product. The objective of this part is to provide as much as possible an accurate classification for opinions embedded in certain sentences like tweets or micro-blogs written in Egyptian dialect as positive, negative or neutral. In addition to counting the total numbers of positive, negative and neutral tweets found in the data source with regards to specified topic. 5

**Determining Influential Bloggers:** Since influential members in a social network can be responsible for starting a buzz or getting the community to notice a new trend, product, or even adopt an opinion, we are interested in the problem of identifying which users are leaders. For companies, organizations and governments, it is of great importance to learn about opinions in order to assess chances and risks. A manual analysis is only possible on a very limited scale. An automated computer supported analysis is necessary given the large number of virtual communities with huge amounts of postings.



Architectural diagram for the product

### 3.3.2 Usability

- It can be useful to business for analysing how to market product and what is the sentiment of people, how they will respond to product. It can be useful for predicting the tweets for a certain group of people.
- It can provide a competitive advantage for business opportunities and product possibilities

### 3.3.3 Reliability

The solution should provide reliability to the user that the product will run with all the features mentioned in this document are available and executing perfectly. It should be tested and debugged completely. All exceptions should be well handled.

### 3.3.4 Performance

As for this prototype version we will keep on detecting if the system crashed, hanged or an operating system error occurred. Also detecting the performance of the system in terms of the efficiency of integration of the different components

### 3.3.5 Supportability

It is supported in all web browsers viz google chrome, Microsoft edge, apple safari, mozilla firefox. It is written in Python3, HTML5, Streamlit framework.

### 3.3.6 Design Constraints

As we are planning to launch a prototype for testing purposes then a beta version for more advanced validation process then launching the final version. The following constraints will apply to for both the prototype and the different live service solution versions.

**Processing Power:** SATA requires high speed machine for data capturing from various sources, classifying the sentiment polarity of large data and extracting topics.

**Deployment Point:** SATA is built to be deployed as internet services. High bandwidth of the portal is required to fulfill the large number of concurrent users.

**Operating Platform:** SATA may work for several distributions of Linux and Windows PCs, also smart phones and smart tablets.

## 3.4 Assumptions and Dependencies

There are four types of project planning dependencies. They establish the relationships among the tasks. They are listed in the order most often used.

- Finish To Start (FS). The first task must complete before the second task can start. For example, the task "Write code module 1" must finish before the task "test code module 1" can begin.
- Finish To Finish (FF). The second task cannot finish before the first task finished. The task "all code tested" cannot finish before the task "test code module x" finishes.
- Start To Start (SS). The second task doesn't start until the first task starts. The task "write training manual" must start before the task "write chapter 1 of training manual" can start.
- Start To Finish (SF). The first task must start before the second task can finish. The task "assign coder for module 3" must start before the task "all work assigned" can finish.

## **Chapter 4: System Analysis**

### **4.1 Requirements**

#### **Functional Requirements**

|                         |   |
|-------------------------|---|
| System Feature          | Sentiment Classification  |
| Priority                | high  |
| Description             | Identifying the sentiment polarity (positive, negative or neutral) of tweets on certain topics from twitter.  |
| Action                  | This module is activated after the user provides a query (topic, service or a product) or following the activation of the hot topic module.   |
| Result                  | The system shows the results of the search of a query or the output of the hot topic module associated with the sentiment polarity of each item retrieved together with the percentage of Positive, Negative and Neutral sentiment of the whole result. |
| Functional requirements | A focused crawler, pre-processing module, sentiment classifier module, hot topic module and sentiment visualization module.   |

|                         |   |
|-------------------------|---|
| System Feature          | User Feedback   |
| Priority                | Medium  |
| Description             | The user can give feedback by correcting the polarity of the classified retrieved tweets, and save the results  |
| Action                  | The user selects a result and suggests a better annotation for it.  |
| Result                  | The suggested correction by the user is stored in a system database to be handled by an administrator, and it is applied for future training and modifications to the system. |
| Functional requirements | A feedback interaction module   |

|                         |   |
|-------------------------|---|
| System Feature          | Influential Bloggers Identification   |
| Priority                | medium  |
| Description             | Identifying the influential users on social media in certain topics.  |
| Action                  | This module is activated after the user provides a query (topic, service or a product) or following the activation of the hot topic module. |
| Result                  | The system shows a list of all influential users on Twitter platform in certain topic, with indications on the level of influence.          |
| Functional requirements | A focused crawler, Influential bloggers identification module, hot topic module, and influential blogger visualization module.              |

|                         |  |
|-------------------------|--|
| System Feature          | Hot Topics Identification  |
| Priority                | High   |
| Description             | Identifying the Hot topics and Trending topics in Twitter according time period.   |
| Action                  | This module is activated after the user provides a date interval. The default interval is the last week using the system date.   |
| Result                  | The system shows the hot and trending topics, putting them in order from high trending topics to lower and the user can browse the tweets related to any of the topic. |
| Functional requirements | A focused Crawling, and the topic extraction module  |

|                         |  |
|-------------------------|--|
| System Feature          | Results Visualization of the SATA components   |
| Priority                | medium   |
| Description             | Visualizing the results of sentiment classification, influential blogger and topic extraction modules into clear and interesting form. |
| Action                  | The proper modules will be activated by the user using a bottom included in the output screen of each of SATA modules.                 |
| Result                  | The system shows the results in the visualization form selected.   |
| Functional requirements | Sentiment classification, influential blogger and topic extraction Visualization modules.  |

|                         |  |
|-------------------------|--|
| System Feature          | Statistics and info-graphics   |
| Priority                | Low  |
| Description             | Viewing different collected statistics about retrieved classified tweets, hot topics tweets, and influential bloggers in a good visualized form such as info-graphics. |
| Action                  | The proper modules will be activated by the user using a bottom included in the output screen of each of SATA modules.   |
| Result                  | Reports and Info-graphics that shows the statistics required   |
| Functional requirements | Sentiment classification, influential blogger and topic extraction Statistics modules.   |

## Non-functional Requirements

**Performance Requirements:** As for this prototype version we will keep on detecting if the system crashed, hanged or an operating system error occurred. Also detecting the performance of the system in terms of the efficiency of integration of the different components

**Safety Requirements:** For the safety requirements nothing but an operation of weekly backups for the data base should take place.

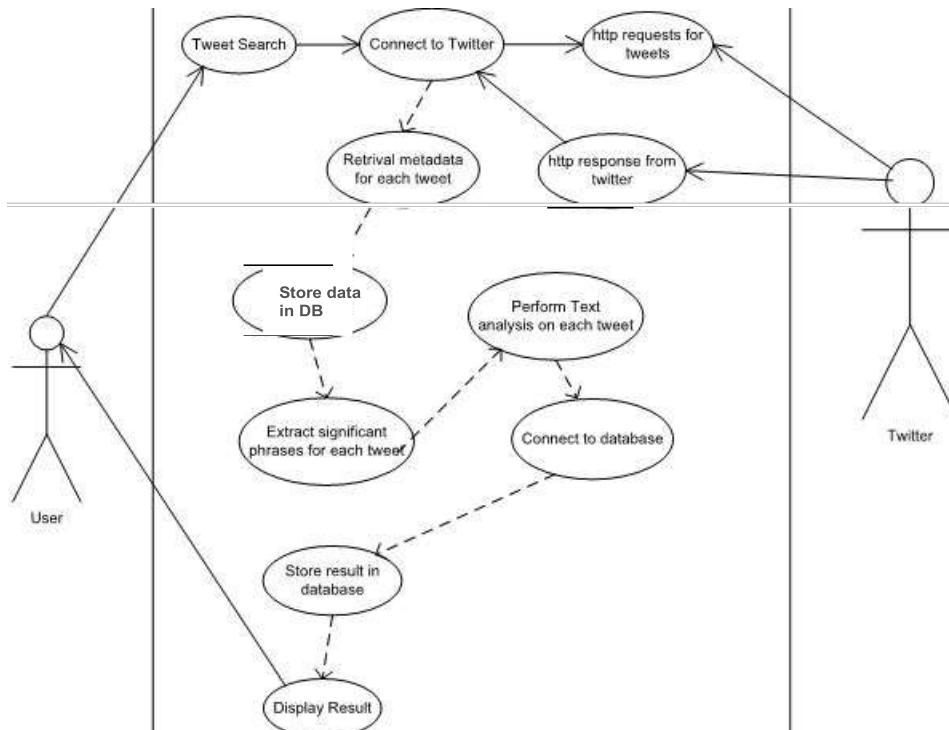
**Security and Privacy Requirements:** There are no specific security requirements, anyone can access and use the portal but only authorized persons who are allowed to use and access the database, web pages and the product engine.

### Software Quality Attributes:

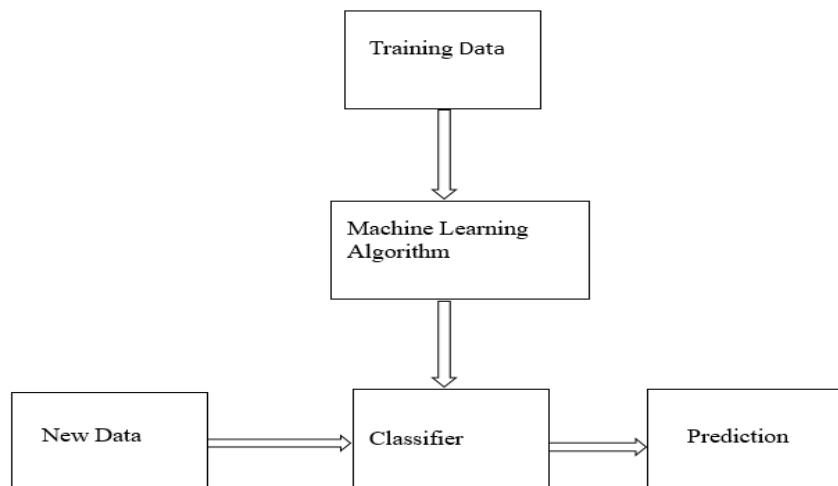
Accuracy: The solution should be able to reach the desired level of accuracy. But also keeping in mind that this prototype version is for proving the concept of the project

## 4.2 Diagrams

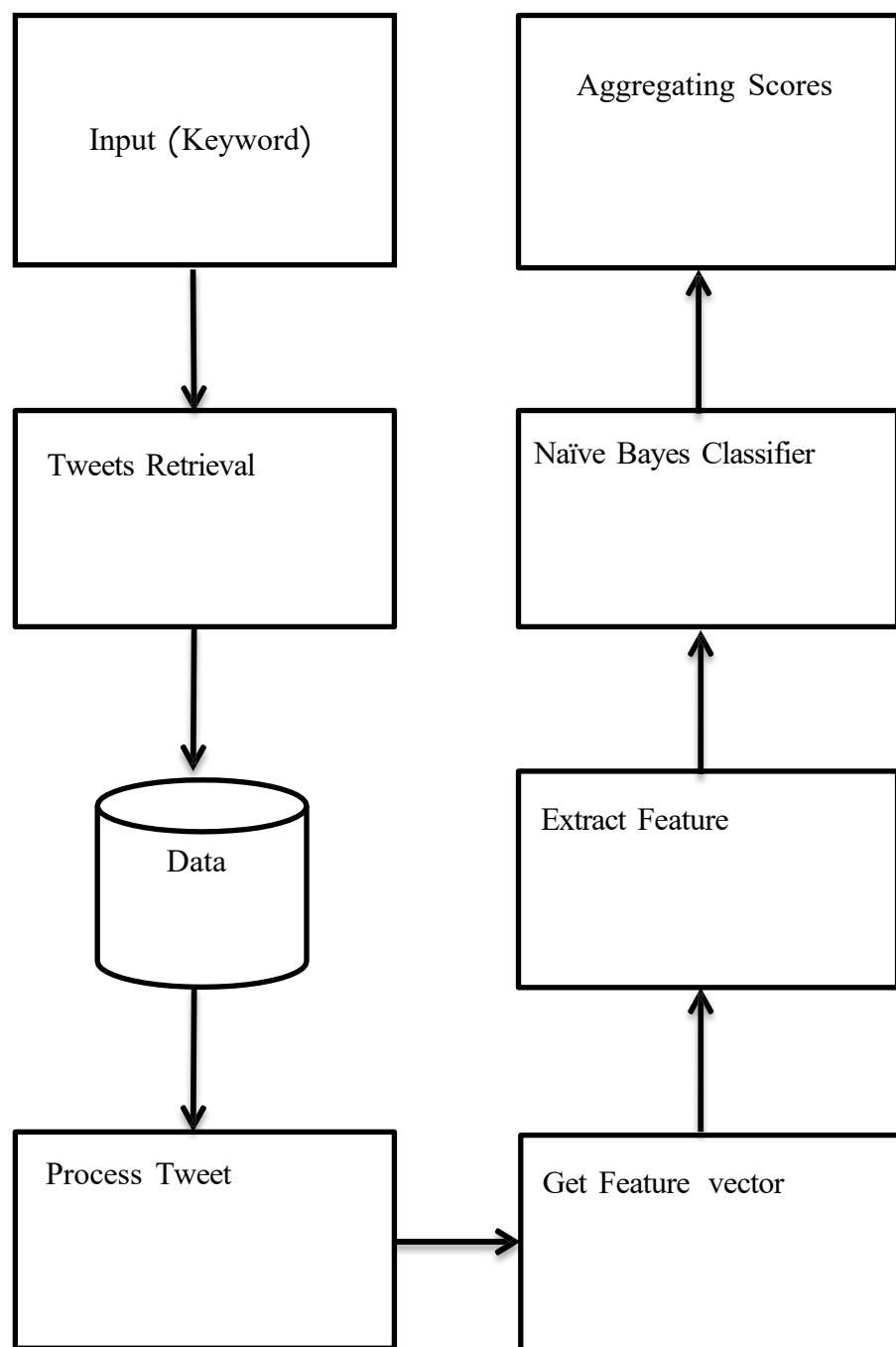
### 1) Use Case Diagram



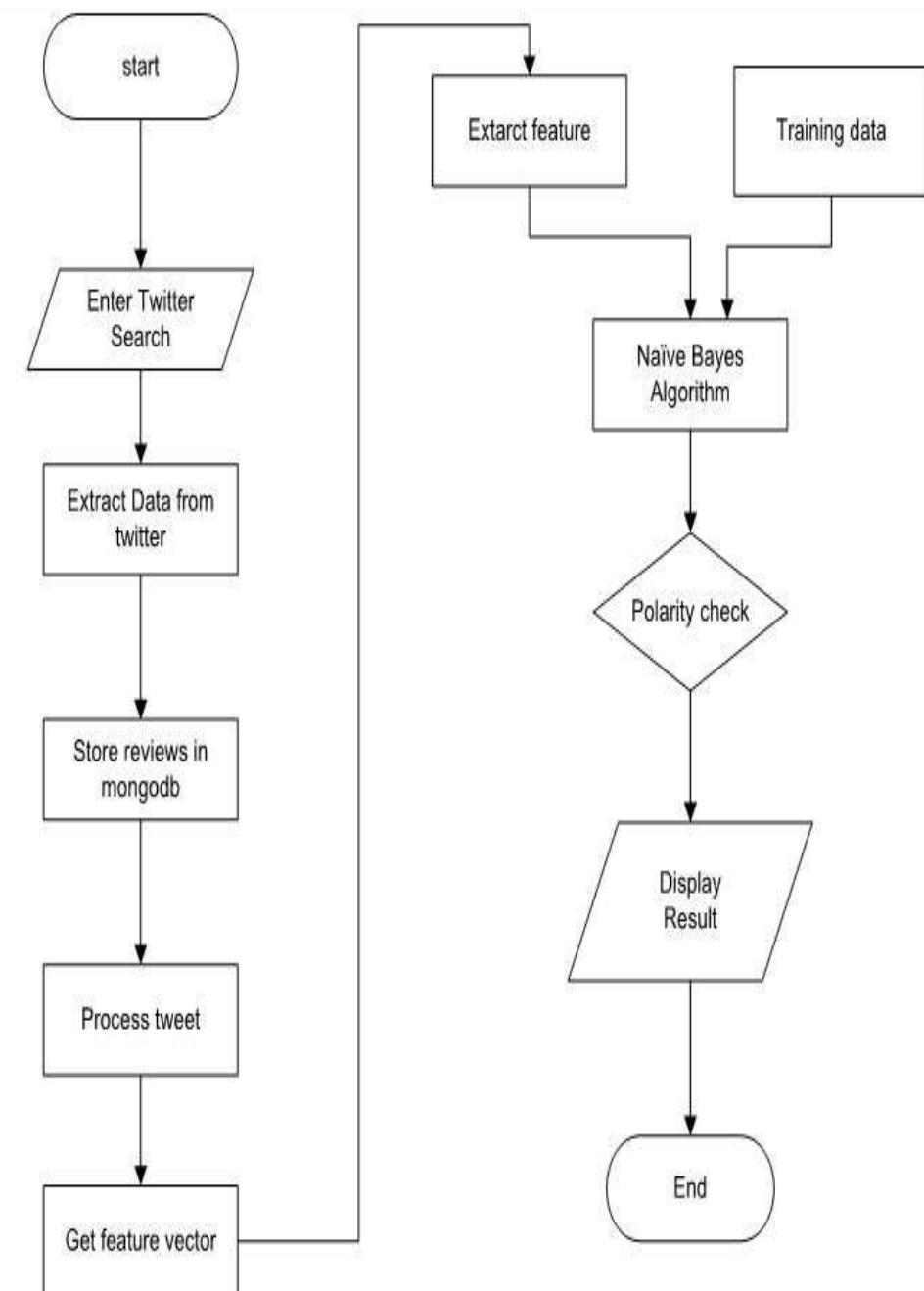
### 2) System Flow Diagram



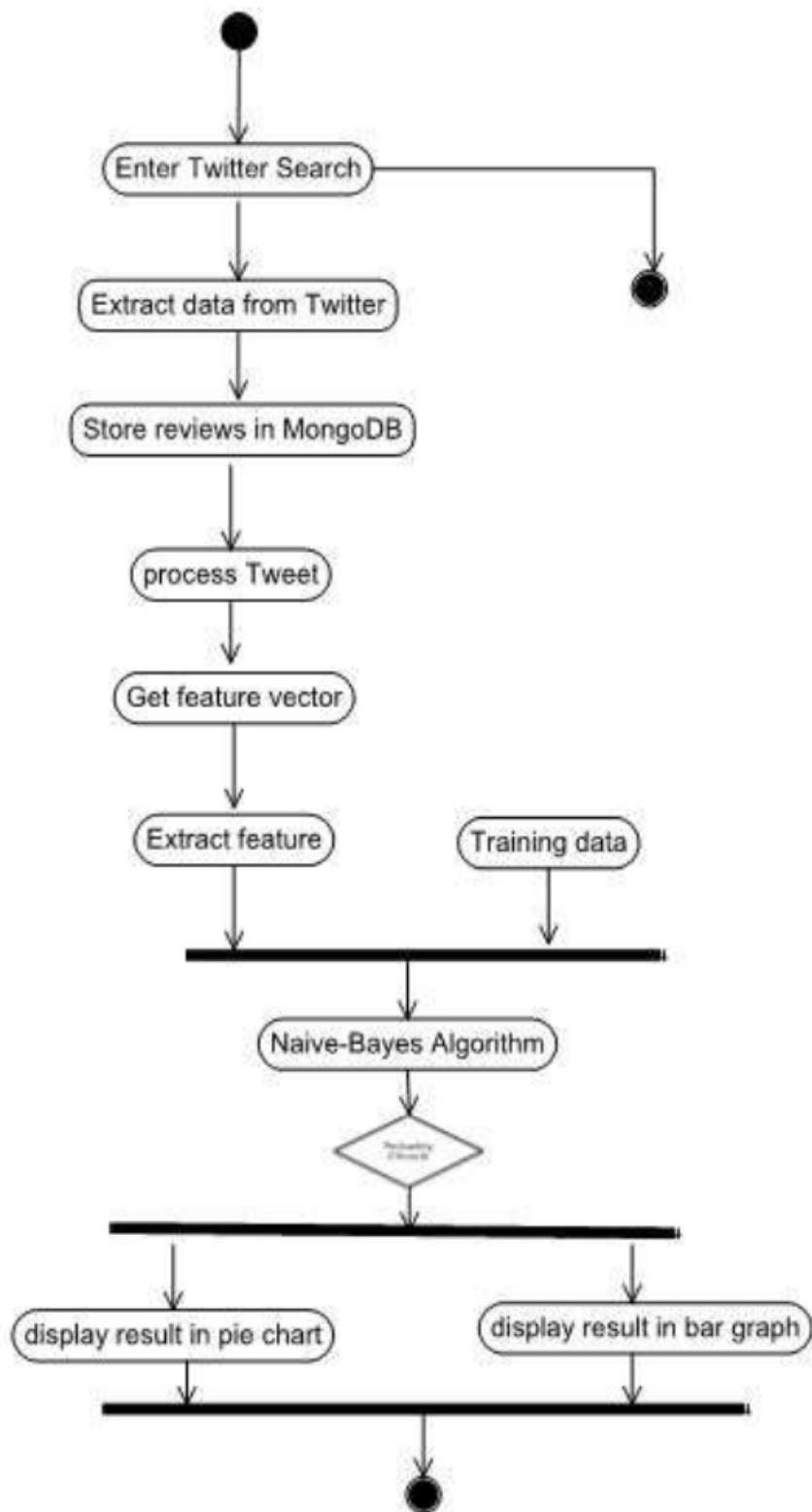
## 3) System Design



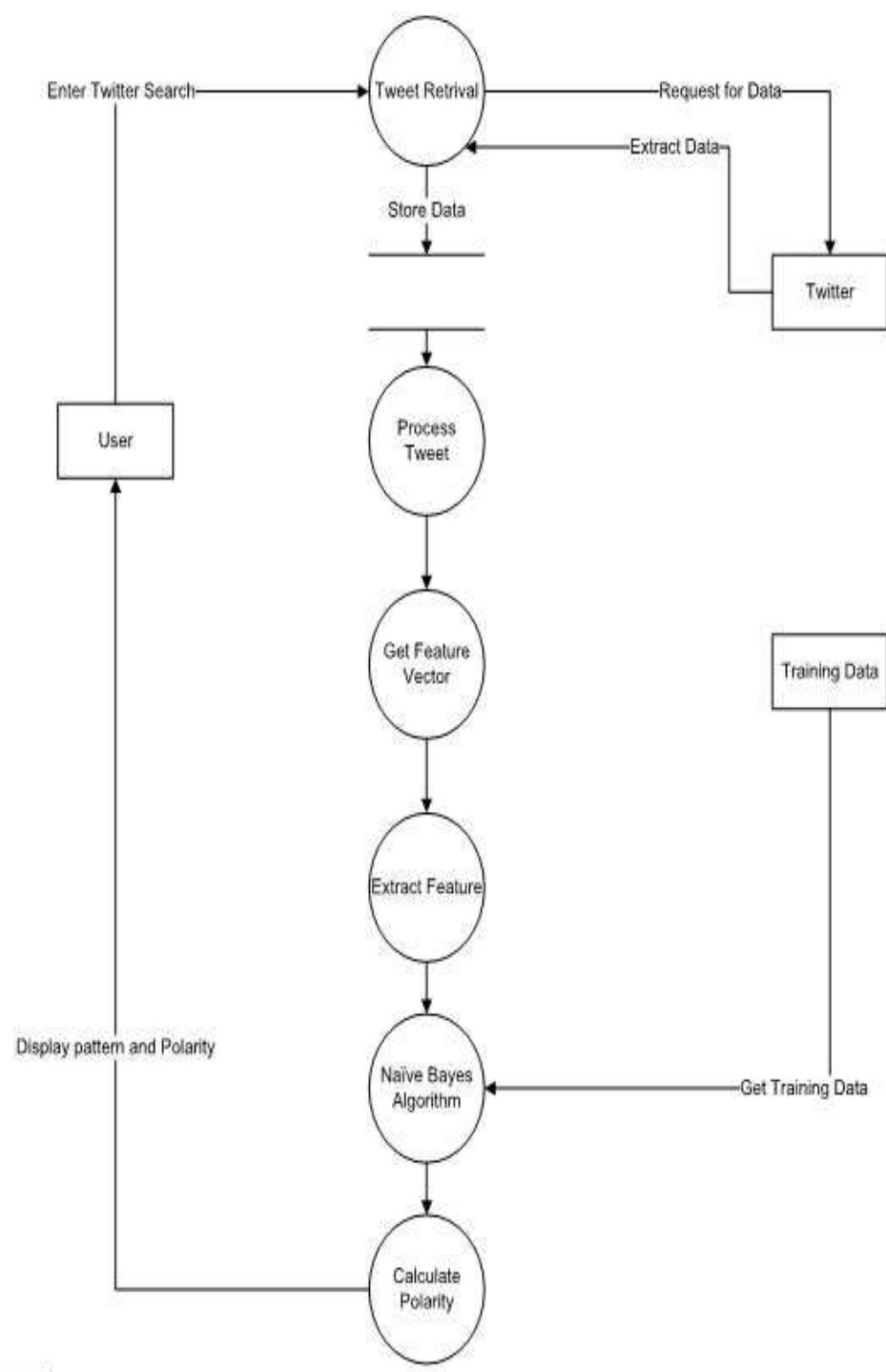
## 4) Flow Chart



## 5) System Activity Diagram



## 6) Data Flow Diagram



## Chapter 5: System Design

### 5.1 Database Design

**Table Name: Tweet**

| No. | Name   | Type            | Contents   |
|-----|--|-----------------|--|
| 1   | <code>twitter.domains</code>                 | Array of string | List of domains from links mentioned in this Tweet.                          |
| 2   | <code>twitter.geo</code>                     | Geo             | The location from which this Tweet was sent.                                 |
| 3   | <code>twitter.in_reply_to_screen_name</code> | String          | The Twitter username of the user this Tweet is replying to if it is a reply. |
| 4   | <code>twitter.links</code>                   | Array of string | List of links mentioned in Tweet.  |
| 5   | <code>twitter.mentions</code>                | Array of string | List of Twitter usernames mentioned in this tweet.                           |
| 6   | <code>twitter.source</code>                  | String          | The source of the Tweet. For example, "web" or "TweetMeme".                  |
| 7   | <code>twitter.text</code>                    | String          | The text of the Tweet.   |

**Table Name: User**

| No. | Name                                      | Type    | Contents  |
|-----|---|---------|---|
| 1   | <code>twitter.user.description</code>     | String  | The Twitter user's biographical description.              |
| 2   | <code>twitter.user.followers_count</code> | Integer | The number of followers the user has.                     |
| 3   | <code>twitter.user.follower_ratio</code>  | Float   | Ratio of followers to following users.                    |
| 4   | <code>twitter.user.friends_count</code>   | Integer | The number of people the user follows.                    |
| 5   | <code>twitter.user.id</code>              | Integer | Unique ID of the Twitter user.                            |
| 6   | <code>twitter.user.lang</code>            | String  | Two-character language code that the User set in Twitter. |
| 7   | <code>twitter.user.listed_count</code>    | Integer | Number of lists in which the user appears.                |
| 8   | <code>twitter.user.location</code>        | String  | The string description of the Twitter user's location.    |
| 9   | <code>twitter.user.name</code>            | String  | The "real name" the user has assigned to themselves.      |
| 10  | <code>twitter.user.profile_age</code>     | Integer | The number of days since this user joined Twitter.        |

**Table Name: Place**

| No. | Name                                    | Type            | Contents   |
|-----|---|-----------------|--|
| 1   | <code>twitter.place.attributes</code>   | Array of string | Additional information about the Twitterer's location. |
| 2   | <code>twitter.place.country</code>      | String          | The country from which this Tweet was sent.            |
| 3   | <code>twitter.place.country_code</code> | String          | Country code for the country this Tweet was sent from. |

|   |                                 |        |  |
|---|---------------------------------|--------|--|
| 4 | <b>twitter.place.full_name</b>  | String | Full name of the place from which this Tweet was sent.   |
| 5 | <b>twitter.place.name</b>       | String | Short name of the place from which this Tweet was sent.  |
| 6 | <b>twitter.place.place_type</b> | String | The type of place from which this Tweet was; for example:<br>city, neighborhood, point of interest.                                    |
| 7 | <b>twitter.place.url</b>        | String | For a Tweet with place information, this string contains a link to the Twitter API to retrieve further information about the location. |

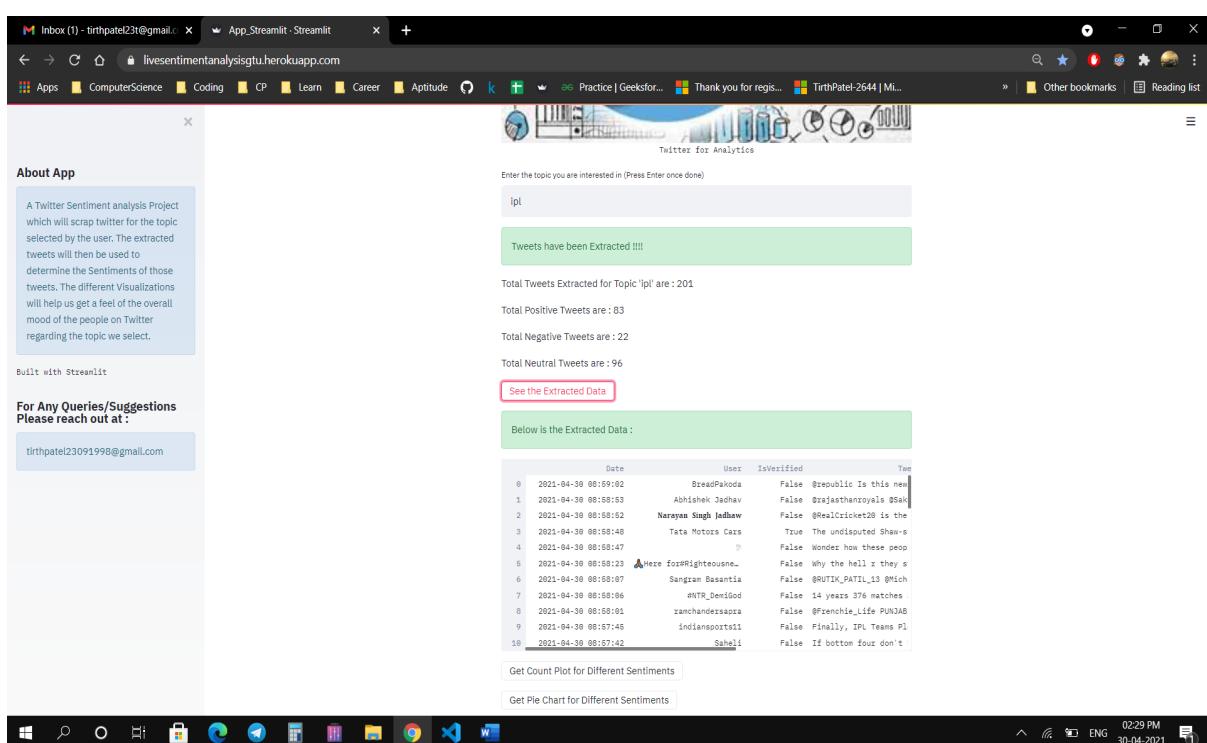
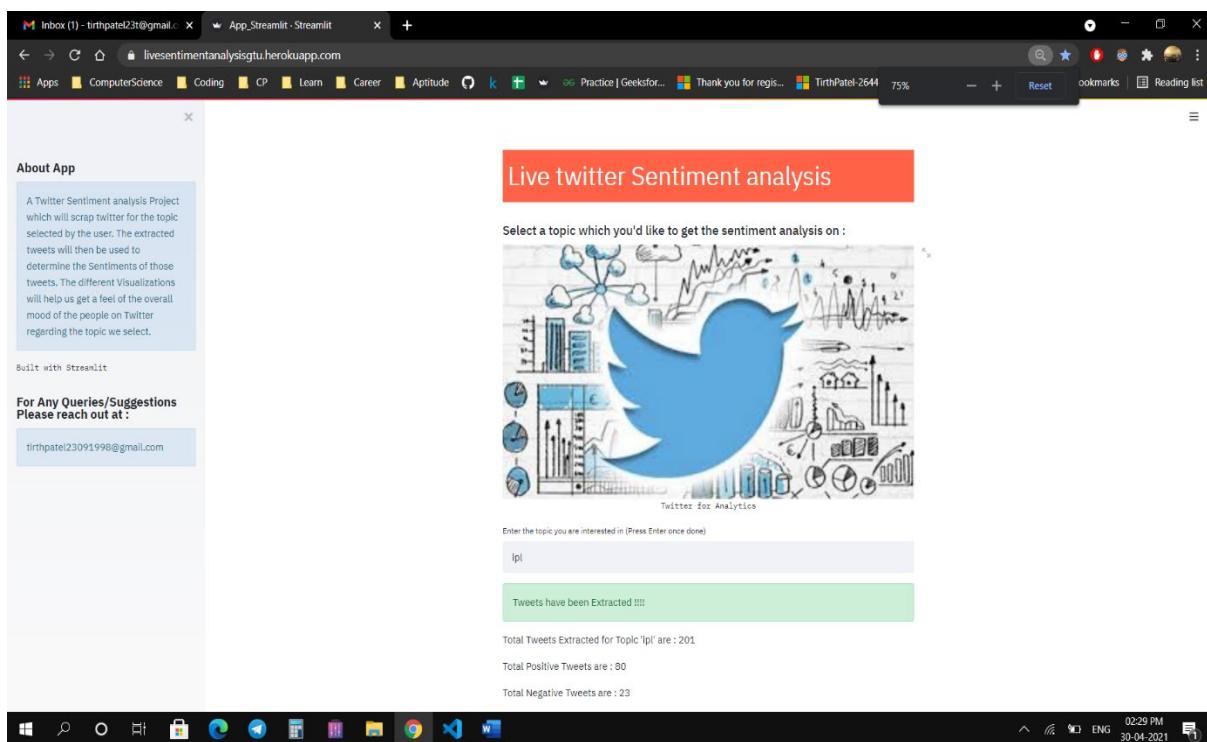
**Table Name: Retweet**

| No. | Name  | Type            | Contents  |
|-----|---|-----------------|---|
| 1   | <b>twitter.retweet.count</b>                | Integer         | The total number of Retweets for this Tweet.                                |
| 2   | <b>twitter.retweet.domains</b>              | Array of string | List of domains from links mentioned in the tweet that was Retweeted.       |
| 3   | <b>twitter.retweet.elapsed</b>              | Integer         | In seconds how long between this retweet and the Tweet are they retweeting. |
| 4   | <b>twitter.retweet.links</b>                | Array of string | List of links mentioned in the Tweet that was Retweeted.                    |
| 5   | <b>twitter.retweet.source</b>               | String          | The string source of the Retweet; for example: "web" or "Tweetdeck".        |
| 6   | <b>twitter.retweet.text</b>                 | String          | The Retweet text.   |
| 7   | <b>twitter.retweet.user.description</b>     | String          | The biography information for the Twitter user who Retweeted this Tweet.    |
| 8   | <b>twitter.retweet.user.followers_count</b> | Integer         | The number of followers the user has.                                       |
| 9   | <b>twitter.retweet.user.followers_ratio</b> | Float           | Ratio of followers to following users.                                      |
| 10  | <b>twitter.retweet.user.friends_count</b>   | Integer         | The number of people the Retweeting user follows.                           |

**Table Name: Retweeted**

| No. | Name                                   | Type            | Contents  |
|-----|--|-----------------|---|
| 1   | twitter.retweeted.id                   | String          | The unique ID of the Tweet that was Retweeted.                        |
| 2   | twitter.retweeted.mentions             | Array of string | List of Twitter usernames mentioned in the Tweet that was Retweeted.  |
| 3   | twitter.retweeted.place.country        | String          | The Retweeted Country from which this tweet was made.                 |
| 4   | twitter.retweeted.place.country_code   | String          | Country Code for the country this Retweeted Tweet was made from.      |
| 5   | twitter.retweeted.place.full_name      | String          | Full name of the Place from which this Retweeted Tweet was made.      |
| 6   | twitter.retweeted.place.name           | String          | Short name of the Place from which this Retweeted Tweet was made.     |
| 7   | twitter.retweeted.place.place_type     | String          | The Retweeted type of place this tweet was made from.                 |
| 8   | twitter.retweeted.source               | String          | The source of the Retweeted Tweet. For example: "web" or "TweetDeck". |
| 9   | twitter.retweeted.user.description     | String          | The Retweeted Twitter user's description.                             |
| 10  | twitter.retweeted.user.followers_count | Integer         | The number of followers the Retweeted author has.                     |
| 11  | twitter.retweeted.user.foll            | Float           | Ratio of followers to following Retweeted users.                      |

## 5.3 Input-Output Design



Inbox (1) - tirthpatel23t@gmail.com App\_Streamlit - Streamlit +

See the Extracted Data

Get Count Plot for Different Sentiments

Generating A Count Plot

Total Neutral Tweets are : 97

About App

A Twitter Sentiment analysis Project which will scrap twitter for the topic selected by the user. The extracted tweets will then be used to determine the Sentiments of those tweets. The different Visualizations will help us get a feel of the overall mood of the people on Twitter regarding the topic we select.

Built with Streamlit

For Any Queries/Suggestions Please reach out at :

tirthpatel23091998@gmail.com

Count Plot for Different Sentiments

AxesSubplot(0.125, 0.11; 0.775x0.77)

| Sentiment | Count |
|-----------|-------|
| Neutral   | 97    |
| Negative  | 23    |
| Positive  | 83    |

Get Pie Chart for Different Sentiments

Get Count Plot Based on Verified and unverified Users

02:30 PM 30-04-2021

Inbox (1) - tirthpatel23t@gmail.com App\_Streamlit - Streamlit +

See the Extracted Data

Get Count Plot for Different Sentiments

Generating A Pie Chart

About App

A Twitter Sentiment analysis Project which will scrap twitter for the topic selected by the user. The extracted tweets will then be used to determine the Sentiments of those tweets. The different Visualizations will help us get a feel of the overall mood of the people on Twitter regarding the topic we select.

Built with Streamlit

For Any Queries/Suggestions Please reach out at :

tirthpatel23091998@gmail.com

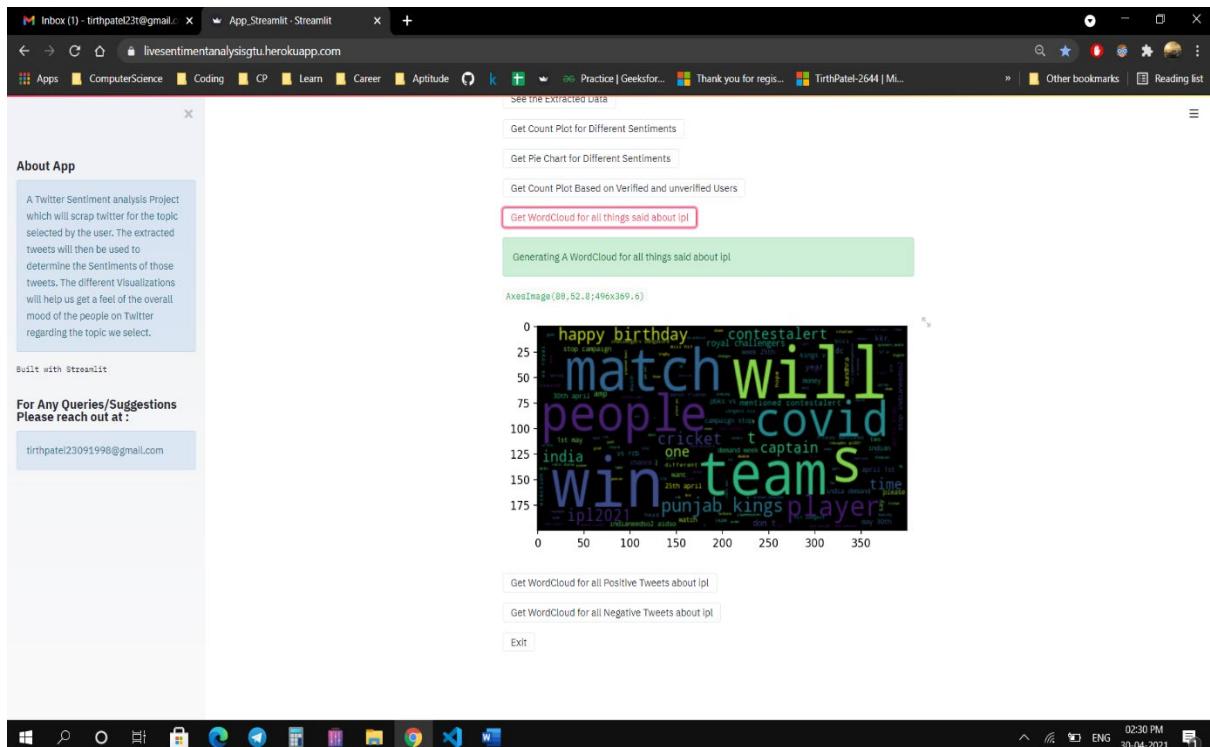
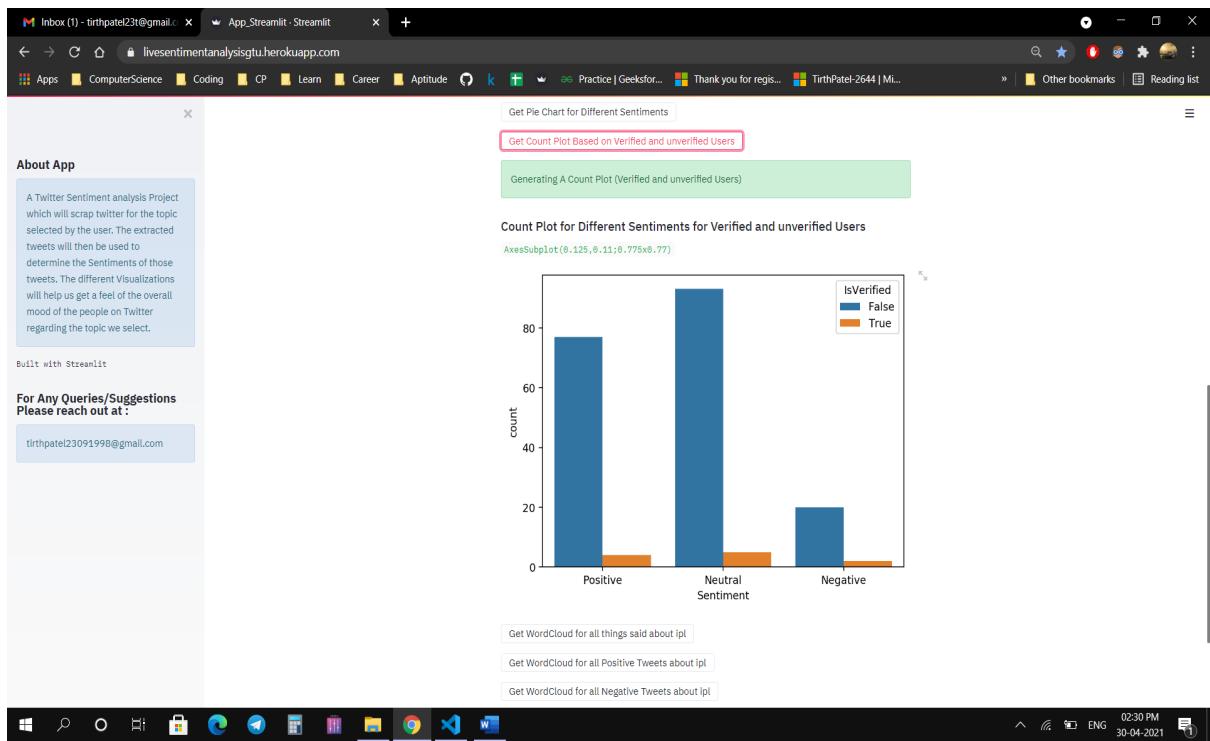
Get Pie Chart for Different Sentiments

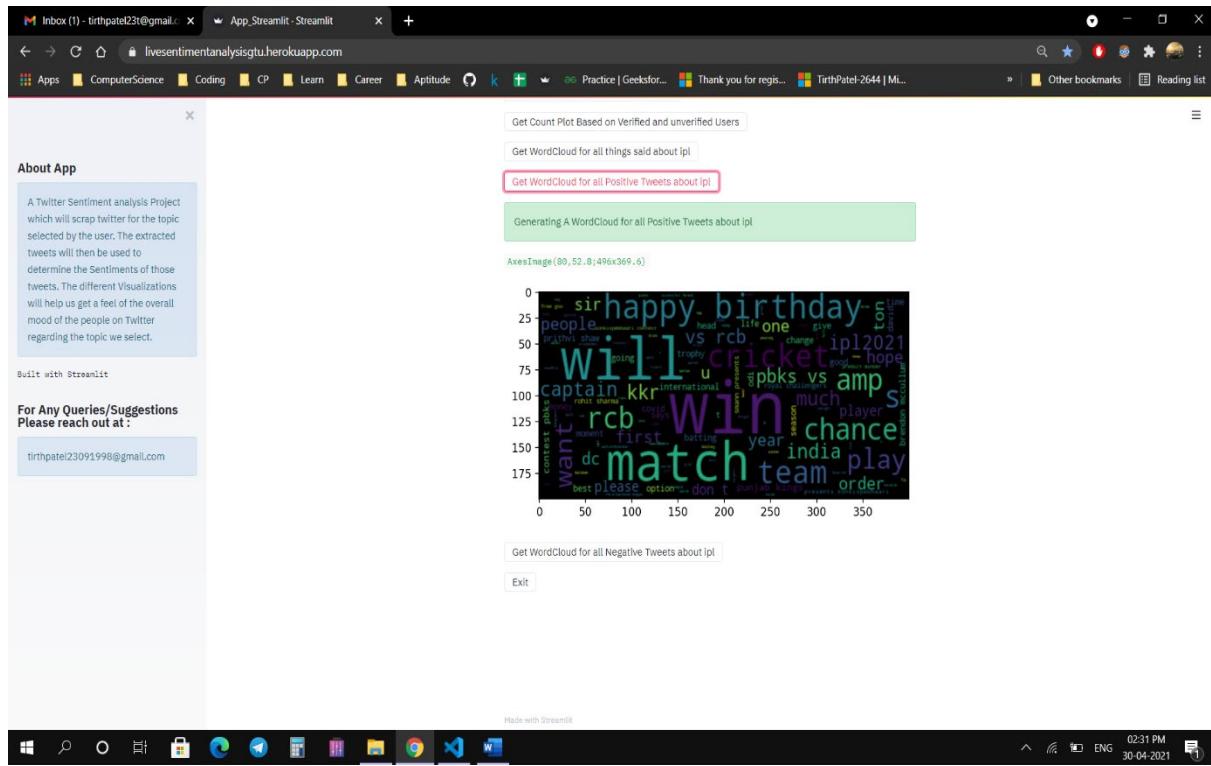
Generating A Pie Chart

```
[(<matplotlib.patches.Wedge object at 0x7efddde2acf0>, <matplotlib.patches.Wedge object at 0x7efddde08518>, <matplotlib.patches.Wedge object at 0x7efddde08518>], [Text(0.3421699868652487, 1.1561833795771452, 'Positive'), Text(-1.0699072699423284, -0.26553583324188905, 'Negative'), Text(0.06561257813392594, -1.1982849830872526, 'Neutral')], [Text(0.19959736335308284, 0.6709403047533345, '40.80%'), Text(0.139381817683091, '10.95%'), Text(0.03827400391145679, -0.69895286887564, '48.26%')])
```

| Sentiment | Percentage |
|-----------|------------|
| Positive  | 40.80%     |
| Negative  | 10.95%     |
| Neutral   | 48.26%     |

02:30 PM 30-04-2021





## **Chapter 6: Limitation and Future Enhancement**

### **6.1 Limitation**

The system we designed is used to determine the opinion of the people based on twitter data. Limitations of not having live tweets is removed and with the help of tweepy, we are able to do live analysis of tweet.

Sometimes algorithm fails to identify hate speech and violent tweets. Neutral tweets can sometimes become positive or negative based on the number of likes.

Live Twitter Sentiment Analysis is not possible with static dataset.

### **6.2 Future Enhancement**

- Analysing sentiments on emo/smiley.
- Determining neutrality.
- Potential improvement can be made to our data collection and analysis method.
- Future research can be done with possible improvement such as more refined data and more accurate algorithm.
- Integrate APIs for live twitter tracking.

## **Chapter 7: Conclusion and Discussion**

The task of sentiment analysis, especially in the domain of micro-blogging, is still in the developing stage and far from complete. So we propose a couple of ideas which we feel are worth exploring in the future and may result in further improved performance.

The testing process is a part of broader subject referring to verification and validation. We have to acknowledge the system specifications and try to meet the customer's requirements and for this sole purpose, we have to verify and validate the product to make sure everything is in place. Verification and validation are two different things. One is performed to ensure that the software correctly implements a specific functionality and other is done to ensure if the customer requirements are properly met or not by the end product.

Verification is more like 'are we building the product right?' and validation is more like 'are we building the right product?'.

We have build project using Streamlit App which is most favourable for machine learning projects. In backend, we have used python for algorithm and sentiment analysis. For extracting tweets, we have used Tweepy which will fetch live tweets from twitter and help us to perform live twitter sentiment analysis.

In this research we are focussing on general sentiment analysis. There is potential of work in the field of sentiment analysis with partially known context. For example we noticed that users generally use our website for specific types of keywords which can be divided into a couple of distinct classes, namely: politics/politicians, celebrities, products/brands, sports/sportsmen, media/movies/music. So we can attempt to perform separate sentiment analysis on tweets that only belong to one of these classes (i.e. the training data would not be general but specific to one of these categories) and compare the results we get if we apply general sentiment analysis on it instead.

With various machine learning algorithm which are developing, this project has maximum chance of improvement and no algorithm can accurately differentiate between positive and negative sentiments. It is just that how much accurate and precise we can make our system to get best results.

## **Chapter 8: References**

developers.twitter.com

[www.kaggle.com](http://www.kaggle.com)

www.heroku.com

Twitter (2014) Twitter apis. <https://dev.twitter.com/start>.

LiuB(2014)The science of detecting fake reviews. <http://content26.com/blog/bingliu-the-science-of-detecting-fake-reviews/>

Efthymios Kouloumpis and Johanna Moore,IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 4, No 3, July 2012

S. Batra and D. Rao, "Entity Based Sentiment Analysis on Twitter", Stanford University,2010

[www.github.com](http://www.github.com)

[www.javatpoint.com](http://www.javatpoint.com)

## **Appendix 1: Periodic Progress Report**

### **PPR1:-**

#### **PPR Details**

Periodic Progess Report : First PPR

Project : Sentiment Analysis

Status : Reviewed

1. What Progress you have made in the Project ?

Firstly we both have figure which tools we are going to use for sentiment analysis and then after we both in teams started learning the libraries that we are going to use for sentiment analysis .

2. What challenge you have faced ?

For sentiment analysis we have to learn four libraries they are Numpy, Pandas, Matplotlib and seaborn this was the biggest challenge that we both have faced.

3. What support you need ?

We will need some for help learning machine learning

4. Which literature you have referred ?

We referred documentation of all these libraries and some books

### **PPR2:-**

#### **PPR Details**

Periodic Progess Report : Second PPR

Project : Sentiment Analysis

Status : Reviewed

1. What Progress you have made in the Project ?

We downloaded some bunch of datasets for twitter sentiment analysis and pre-processed the datasets that means we cleaned the data removed some common words, punctuation that had no meaning for twitter sentiment analysis.

2. What challenge you have faced ?

Pre processing the data was the challenge that we have faced this time.

3. What support you need ?

We would like to help for learning machine learning(ML) and natural language processing(NLP).

4. Which literature you have referred ?

We have referred some blogs on data set pre processing.

### **PPR3:-**

#### **PPR Details**

Periodic Progress Report : Third PPR

Project : Sentiment Analysis

Status : Reviewed

1. What Progress you have made in the Project ?

• Learnt various ML Algorithm like MLR, RFR, KNN-clustering. • Collected data dictionary of words from Kaggle. • Made weekly plan for developing Project. • In this period, all the necessary documentations are prepared like diagrams, data dictionary, literature review and SRS. • Diagrams include process flow diagram, use case diagram, data flow diagram, etc. which describes the project in better format. • SRS contains whole project behavior in the documentation format. • Data dictionary will be made on the basis of database.

2. What challenge you have faced ?

Dictionary of data is very hard to find but ultimately we finally find from kaggle. Learning how API work and how to use them is very challenging.

3. What support you need ?

Working how to fetch twitter tweets through API.

4. Which literature you have referred ?

<https://www.kaggle.com> <https://www.coursera.com>

### **PPR4:-**

#### **PPR Details**

Periodic Progress Report : Forth PPR

Project : Sentiment Analysis

Status : Reviewed

1. What Progress you have made in the Project ?

• In this period, we have started working on the implementation of our project. • Exploratory Data Analysis and Data Preprocessing. • Clean dataset by removing redundant data. • Remove numbers and special symbols from tweet. • Divide dataset in trainset and testset. • Train trainset by fitting into Machine learning Model. • Created Algorithms for Sentiment Analysis.

2. What challenge you have faced ?

Creating algorithm was very challenging. Cleaning dataset was very hard.

3. What support you need ?

Integrating backend with frontend.

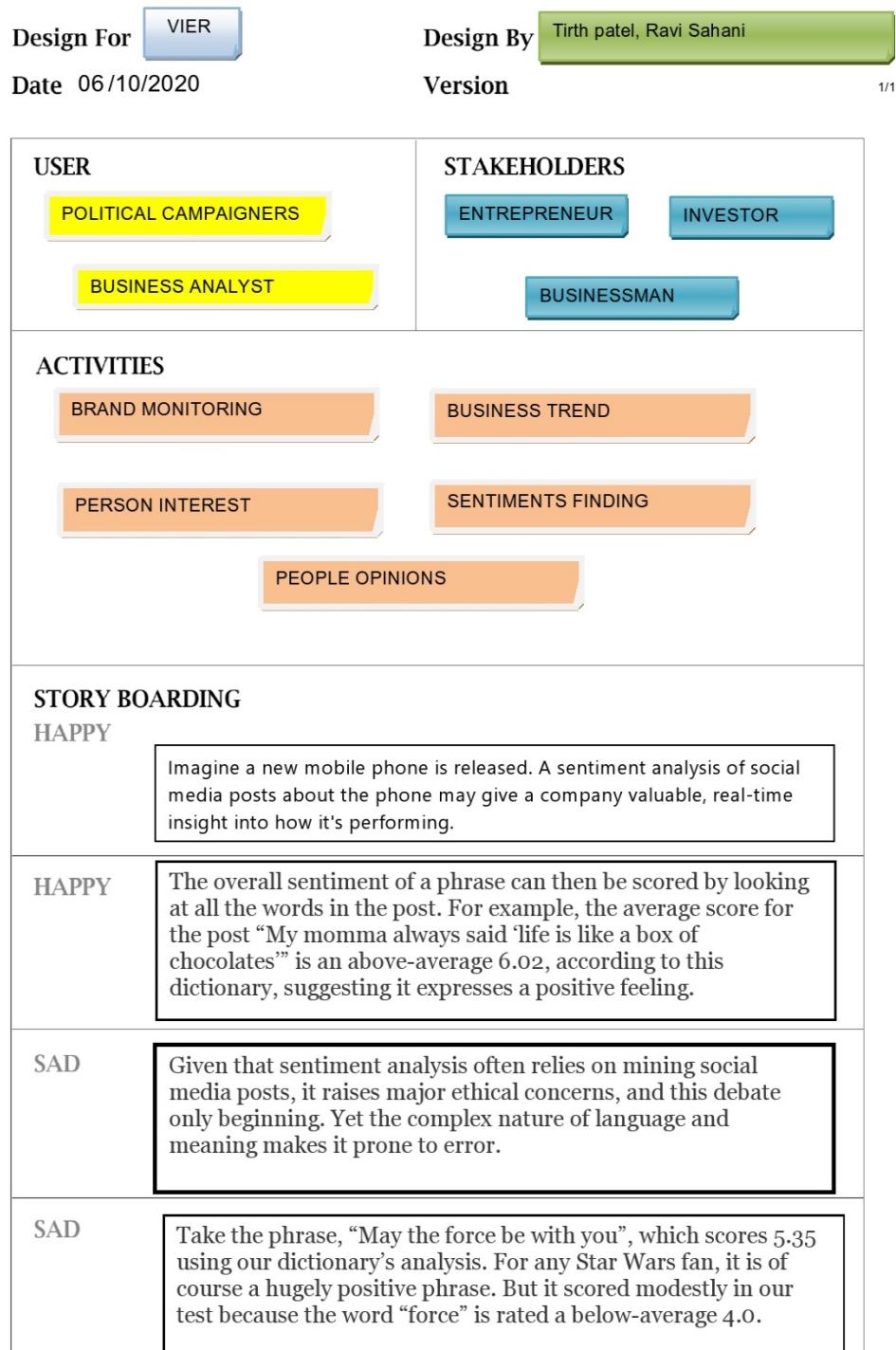
4. Which literature you have referred ?

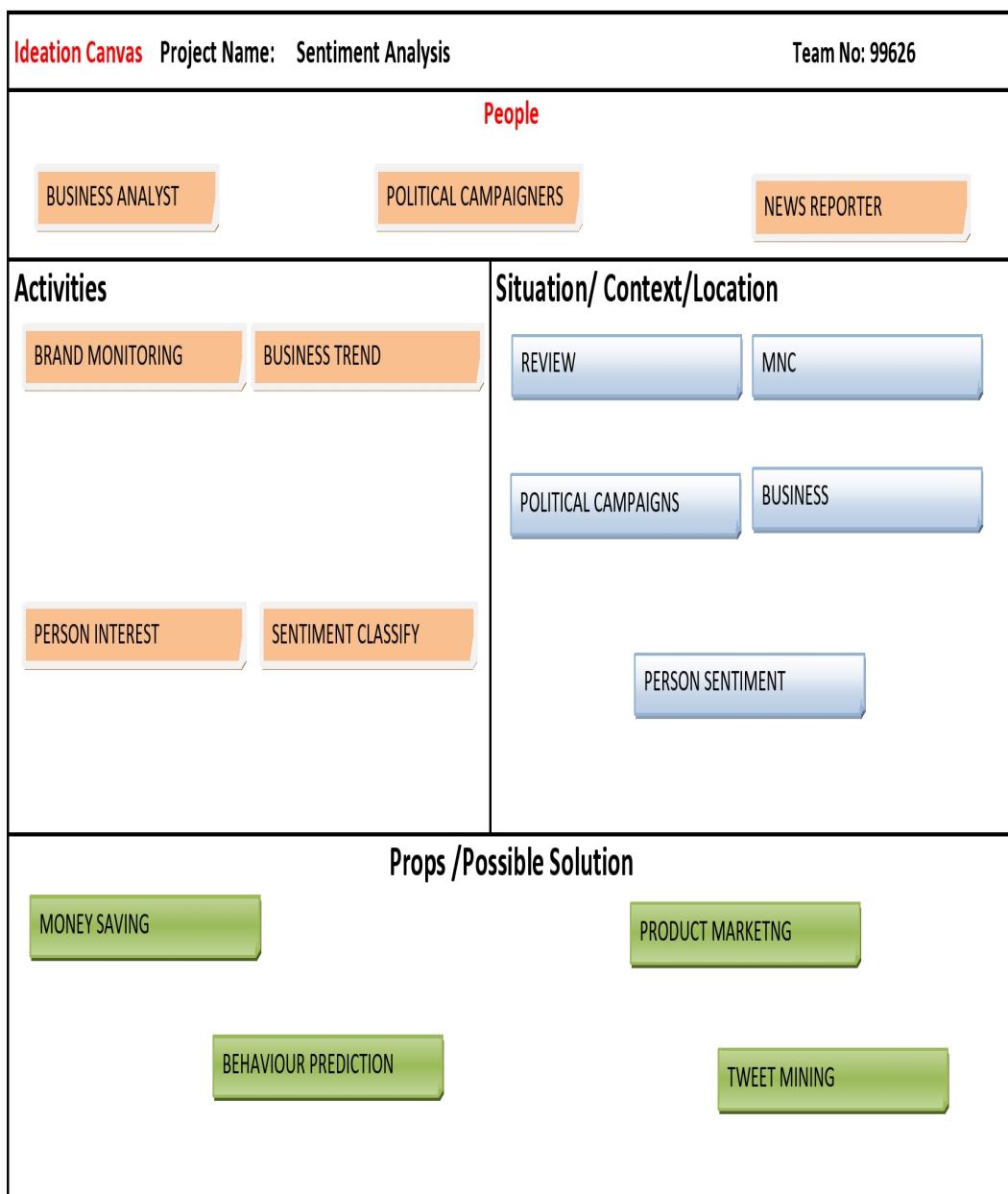
<https://www.kaggle.com/> <https://www.kaggle.com/andyxie/sentiment-analysis-dictionary> <https://www.udemy.com>

## Appendix 2: Design Engineering Canvas

### **AEIOU Summary Canvas:-**

| AEIOU Summary   |  |   | Team No:99626  |
|---|--|---|--|
| <b>ENVIRONMENT</b>                                      | <b>INTERACTION</b>   | <b>OBJECTS</b>  |  |
| MNC<br>News Channels<br>Political Parties<br>Industries | Monitoring<br>Marketing<br>People opinions<br>E-commerce advertising | Jupyter Notebooks<br>TextBlob<br>Scikit - Learn<br>Binder |  |
| <b>ACTIVITIES</b>                                       |  |   | <b>USERS</b>   |
| Brand Monitoring<br>Person interest<br>Business trend   |  |   | Business Analyst<br>Political Campaigners<br>News Reporter<br>For monitoring someone |

**Empathy Canvas:-**

**Ideation Canvas:-**

## Product Development Canvas:-

| PRODUCT DEVELOPMENT CANVAS                               |   |   | Team No: 99626 |
|--|---|---|----------------|
| Purpose  | Product Experience  | Customer revalidation   |                |
| SENTIMENT ANALYSIS<br>BRAND MONITORING<br>BUSINESS TREND | EASY TO USE<br>COLORFUL VISUALIZATION                                 | USER FRIENDLY<br>SMART  |                |
|  | Product Function  |   |                |
|  | SENTIMENT CLASSIFIER<br>NUMBER OF WORDS COUNTER<br>MOST REPEATED WORD |   |                |
| People   | Product Feature   | Reject/Redesign/retain  |                |
| POLITICAL CAMPAIGNERS<br>BUSINESS ANALYST<br>INVESTOR    | SENTIMENT ANALYSIS<br>EASY TO USE<br>PREDICTION                       | PROBLEM IN UI STRUCTURE<br>DIFFICULT SEARCH<br>SIMPLIFYING UI<br>ADDING SMART SEARCH OPTION |                |
|  | Components  |   |                |
|  | JUPYTER NOTEBOOKS<br>TEXTBLOB   | SCIKIT - LEARN<br>BINDER  |                |

## Appendix 3: PSAR Report

### **PSAR1:-**

PSAR No. : 20BE7\_170020107041\_1

#### **Part - I : PATENT SEARCH TECHNIQUE USED**

- |                                   |   |   |
|-----------------------------------|---|---|
| 1. Patent Search Database Used    | : | Google Patents  |
| Web Link of the Database          | : | <a href="https://patents.google.com/">https://patents.google.com/</a> |
| 2. Keywords Used for Search       | : | Detailed,Sentiment,Analysis   |
| 3. Search String Used             | : | Detailed Sentiment Analysis   |
| 4. Number of Results/Hits getting | : | 1351  |

#### **Part - II : BASIC DATA OF PATENTED INVENTION/BIBLIOGRAPHIC DATA**

- |  |   |   |
|--|---|---|
| 5. Category/Field of invention   | : |   |
| 6. Invention Is Related to/Class of Invention  | : | sentiment score based on a content source.  |
| 6a. IPC class of the studied patent  | : | G06Q30/02, G06F16/951, G06F40/30  |
| 7. Title of Invention  | : | Detailed sentiment analysis   |
| 8. Patent No.  | : | US 8,463,595 B1   |
| 9. Application No.   | : | 13/538,710  |
| 9a. Web Link of the studied patent   | : | <a href="https://patents.google.com/patent/US8463595B1/en?q=detailed+sentiment+analysis&amp;oq=detailed+sentiment+analysis">https://patents.google.com/patent/US8463595B1/en?q=detailed+sentiment+analysis&amp;oq=detailed+sentiment+analysis</a> |
| 10. Date of Filing/Application   | : | 06/28/2012  |
| 11. Priority Date  | : | 03/06/2012  |
| 12. Publication/Journal Number - (Issue No. of Journal in which Patent is published) | : | US8463595B1   |
| 13. Publication Date   | : | 06/11/2013  |
| 14. First Filed Country  | : | 284   |
| <b>15. Also Published as</b>   |   |   |

| Country       | Patent No |
|---------------|-----------|
| United States | 13/538    |

#### **16. Inventor**

| Name of Inventor      | Address/City/Country of Inventor |
|-----------------------|----------------------------------|
| John Andrew Rehling   | US                               |
| Thomas Gerardo Dignan | US                               |

#### **17. Applicant**

| Name of Applicant/Assignee | Address/City/Country of Applicant |
|----------------------------|-----------------------------------|
| Reputation.com Inc         | US                                |

18. Applicant for Patent Is : Company

### Part - III : TECHNICAL PART OF PATENTED INVENTION

**19. Limitation of Prior Technology/Art :**

One problem with existing sentiment analysis techniques is seen when the text being evaluated expresses two independent opinions, such as in the following: "Bob is a terrible trumpet player, but the cookies he bakes taste great!" Using existing techniques, the text would be assigned a neutral score—the negative and positive opinions expressed cancelling one another out. For this and other reasons, improvements in sentiment analysis techniques would be desirable.

**20. Specific Problem Solved/Objective of Invention :**

Feelings in statement can be solved.

**21. Brief about Invention :**

The invention can be implemented in numerous ways, including as a process; an apparatus; a system; a composition of matter; a computer program product embodied on a computer readable storage medium; and/or a processor, such as a processor configured to execute instructions stored on and/or provided by a memory coupled to the processor.

**22. Key Learning Points :**

Performing detailed sentiment analysis includes generating a first sentiment score for a first entity based on a content source. The first sentiment score is generated with respect to a first dimension. A second sentiment score for the first entity is generated based on the content source. The second sentiment score is generated with respect to a second dimension.

**23. Summary of Invention :**

Sentiment analysis techniques can be used to assign a piece of text a single value that represents opinion expressed in that text.

**24. Number of Claims** : 20

**25. Patent Status** : Published Application

**26. How much this invention is related with your IDP/UDP? :** 71 to 90%

**27. Do you have any idea to do anything around the said invention to improve it? :**

No

**PSAR2:-**

PSAR No. : 20BE7\_170020107041\_3

**Part - I : PATENT SEARCH TECHNIQUE USED**

- |  |   |   |
|--|---|---|
| <b>1. Patent Search Database Used</b>    | : | Google Patents  |
| <b>Web link of the Database</b>          | : | <a href="https://patents.google.com/">https://patents.google.com/</a> |
| <b>2. Keywords Used for Search</b>       | : | twitter ,sentiment,analysis   |
| <b>3. Search String Used</b>             | : | twitter sentiment analysis  |
| <b>4. Number of Results/Hits getting</b> | : | 5560  |

**Part - II : BASIC DATA OF PATENTED INVENTION/BIBLIOGRAPHIC DATA**

- |   |   |   |
|---|---|---|
| <b>5. Category/Field of Invention</b>   | : |   |
| <b>6. Invention Is Related to/Class of Invention</b>  | : | sentiment score based on a content source.  |
| <b>6a. IPC class of the studied patent</b>  | : | G06F40/10, G06F40/35  |
| <b>7. Title of Invention</b>  | : | Large-scale sentiment analysis  |
| <b>8. Patent No.</b>  | : |   |
| <b>9. Application No.</b>   | : | 13/163,636  |
| <b>9a. Web link of the studied patent</b>   | : | <a href="https://patents.google.com/patent/US8515739?oq=twitter+sentiment+analysis">https://patents.google.com/patent/US8515739?oq=twitter+sentiment+analysis</a> |
| <b>10. Date of Filing/Application</b>   | : | 06/17/2011  |
| <b>11. Priority Date</b>  | : |   |
| <b>12. Publication/Journal Number - (Issue No. of Journal in which Patent Is Published)</b> | : |   |
| <b>13. Publication Date</b>   | : |   |
| <b>14. First Filed Country</b>  | : |   |

**15. Also Published as**

| Country       | Patent No |
|---------------|-----------|
| United States | 13/163    |

**16. Inventor**

| Name of Inventor      | Address/City/Country of Inventor |
|-----------------------|----------------------------------|
| Namrata Godbole       | US                               |
| Steven Skiena         | US                               |
| Manjunath Srinivasiah | US                               |

**17. Applicant**

| Name of Applicant/Assignee                 | Address/City/Country of Applicant |
|--|-----------------------------------|
| Research Foundation of State University NY | US                                |

**18. Applicant for Patent Is** : University

### Part - III : TECHNICAL PART OF PATENTED INVENTION

**19. Limitation of Prior Technology/Art :**

The present invention relates generally to data mining and, more particularly, to a system and method for sentiment analysis.. Some methods for generating sentiment lexicons assume positive and negative sentiment using synonyms and antonyms. Such methods may not accurately capture the sentiment of a word. Other methods for generating sentiment lexicons using semantics, such as "and" and "but", or tone/orientation to determine a sentiment of a word. Such methods may have low accuracy.

**20. Specific Problem Solved/Objective of Invention :**

Methods for analyzing sentiment treat only single complete documents, for example, to determine if a movie review is good or bad or quantify opinion from a product review. Therefore a need exists for a method of generating an accurate sentiment lexicon and for determining a sentiment over a plurality of texts.

**21. Brief about Invention :**

Sentiment analysis of natural language texts is a large and growing field. The analysis includes both methods for automatically generate sentiment lexicons and analyzing sentiment for entire documents.

**22. Key Learning Points :**

An exemplary embodiment of lexicon expansion uses path analysis. Expanding seed lists into lexicons by recursively querying for synonyms using a computer dictionary, e.g., WordNet, is limited by the synonym set coherence weakening with distance. For example, FIG. 8 shows four separate ways to get from good to bad using chains of WordNet synonyms.

**23. Summary of Invention :**

According to an embodiment of the present disclosure, a method for determining a sentiment associated with an entity includes inputting a plurality of texts associated with the entity, labeling seed words in the plurality of texts as positive or negative, determining a score estimate for the plurality of words based on the labeling, re-enumerating paths of the plurality of words and determining a number of sentiment alternations, determining a final score for the plurality of words using only paths whose number of alternations is within a threshold, converting the final scores to corresponding z-scores for each of the plurality of words, and outputting the sentiment associated with the entity.

**24. Number of Claims** : 9

**25. Patent Status** : Granted Patent & In-force Patent

**26. How much this invention is related with your IDP/UDP? :** 71 to 90%

**27. Do you have any idea to do anything around the said invention to improve it? :**

No

**PSAR3:-**

PSAR No. : 20BE7\_170020107041\_2

**Part - I : PATENT SEARCH TECHNIQUE USED**

- 1. Patent Search Database Used : Google Patents
- 2. Web Link of the Database : <https://patents.google.com/>
- 3. Keywords Used for Search : Detailed,Sentiment,Analysis
- 4. Search String Used : Detailed Sentiment Analysis
- 5. Number of Results/Hits getting : 1351

**Part - II : BASIC DATA OF PATENTED INVENTION/BIBLIOGRAPHIC DATA**

- 6. Category/Field of Invention :
- 7. Invention Is Related to/Class of Invention : sentiment score based on a content source.
- 8. IPC class of the studied patent : G06F16/36, G06F17/30731
- 9. Title of Invention : NLP-based sentiment analysis
- 10. Patent No. : 13/208,278
- 11. Application No. : <https://patents.google.com/patent/US8838633B2/en?q=detailed+sentiment+analysis&oq=detailed+sentiment+analysis>
- 12. Date of Filing/Application : 08/11/2011
- 13. Priority Date :
- 14. Publication/Journal Number - (Issue No. of Journal in which Patent is published) :
- 15. Also Published as :

We do not find any published data.

**16. Inventor**

| Name of Inventor  | Address/City/Country of Inventor |
|-------------------|----------------------------------|
| Navdeep S Dhillon | US                               |
| Jisheng Liang     | US                               |

**17. Applicant**

| Name of Applicant/Assignee | Address/City/Country of Applicant |
|----------------------------|-----------------------------------|
| Fiver LLC                  | US                                |

- 18. Applicant for Patent is : Company

### Part - III : TECHNICAL PART OF PATENTED INVENTION

**19. Limitation of Prior Technology/Art :**

Determining the underlying sentiment of an article using a computing system may be difficult because of the variety of styles people employ in expressing sentiment—a comment may be an offhanded compliment in amongst an otherwise negative article, for example. Current techniques often involve traditional keyword searching for particular negative or positive words (verbs) such as "hate," "like," "distaste," etc. to guesstimate the underlying sentiment of an article.

**20. Specific Problem Solved/Objective of Invention :**

Find the percentage of positive and negative expressions of sentiment made by an entity, or about an entity. For example, one can find out what percentage of things being written about the iPhone are positive and which percent are negative. Discover who is criticizing and who is praising a particular person, place or thing. For example, see who is criticizing and praising IBM right now.

**21. Brief about Invention :**

Embodiments described herein provide enhanced computer- and network-based methods, techniques, and systems for providing sentiment analysis and for presenting the results of such analysis. Example embodiments provide a Sentiment Analysis System ("SAS"), which provides tools to enable authors, programmers, users, developers, and the like to incorporate sentiment analysis into their content, such as into their web pages, and other web blogs or textual content

**22. Key Learning Points :**

The present disclosure relates to methods, techniques, and systems for providing sentiment analysis and, in particular, to methods, techniques, and systems for providing sentiment analysis using natural language processing to determine sentiment of objects in a corpus.

**23. Summary of Invention :**

Methods, systems, and techniques for providing sentiment analysis and for presenting the results of such analysis. Example embodiments provide a Sentiment Analysis System ("SAS"), which provides tools to enable authors, programmers, users, developers, and the like to incorporate sentiment analysis into their content, such as into their web pages, and other web blogs or textual content. In one embodiment, the SAS provides a Sentiment Analysis Engine, an SAS API, and one or more user interface tools for presenting sentiment analysis.

**24. Number of Claims** : 16

**25. Patent Status** : Published Application

**26. How much this invention is related with your IDP/UDP? :** 71 to 90%

**27. Do you have any idea to do anything around the said invention to improve it? :**

No

**PSAR4:-**

**PSAR No.** : 20BE7\_170020107041\_5

**Part - I : PATENT SEARCH TECHNIQUE USED**

1. Patent Search Database Used : Google Patents
2. Web link of the Database : <https://patents.google.com/>
3. Keywords Used for Search : sentiment,analysis,using ,machine learning
4. Search String Used : sentiment analysis using machine learning
5. Number of Results/Hits getting : 5570

**Part - II : BASIC DATA OF PATENTED INVENTION/BIBLIOGRAPHIC DATA**

6. Category/Field of Invention :
7. Invention is Related to/Class of Invention : The invention discloses a kind of subjective texts sentiment analysis method based on deep learning
8. IPC class of the studied patent : G06F40/30
9. Title of Invention : Subjective texts sentiment analysis method based on deep learning
10. Patent No. :
11. Application No. : CN201710093687.1A
12. Web link of the studied patent : <https://patents.google.com/patent/CN106776581A/en?oq=sentiment+analysis+using+machine+learning>
13. Date of Filing/Application : 02/21/2017
14. Priority Date :
15. Publication/Journal Number - (Issue No. of Journal in which Patent is published) :
16. Publication Date :
17. First Filed Country :
18. Also Published as :

We do not find any published data.

**16. Inventor**

| Name of Inventor                   | Address/City/Country of Inventor |
|------------------------------------|----------------------------------|
| Shi Hanxiao Li Xiaojun Chen Nannan | China                            |

**17. Applicant**

| Name of Applicant/Assignee         | Address/City/Country of Applicant |
|------------------------------------|-----------------------------------|
| Shi Hanxiao Li Xiaojun Chen Nannan | China                             |

19. Applicant for Patent is : Individual

### Part - III : TECHNICAL PART OF PATENTED INVENTION

**19. Limitation of Prior Technology/Art :**

In recent years, the propulsion of community network (social network) application and user created the emerging of content (UGC) pattern Rise, common netizen increasingly becomes the important producer of the information content. By taking Financial Information comment as an example, as finance on internet. The user comment of one of UGC important forms in information, had both expressed individual viewpoint of the user to securities market, and user is covered again Relation between stock, user and user, has content and the feature for associating concurrently, the new heat of as sentiment analysis research Point.

**20. Specific Problem Solved/Objective of Invention :**

Current sentiment analysis method substantially has two classes. The first is rule-based method. It is main first according to sentiment dictionary Find out the emotion word occurred in text, then carry out simple feeling polarities statistics, according to final score with threshold value is previously set Compare and draw feeling polarities conclusion, be generally used for the sentiment analysis of chapter rank. Second is the method based on machine learning. It is logical. The training to a large amount of mark language materials is crossed, emotion classifiers is generated, for classifying to test text.

**21. Brief about Invention :**

Sentiment analysis research for subjective texts exactly adapts to this demand, it is desirable to set up a user to the bridge of emotional information Beam, enables users to effectively obtain emotional information.

**22. Key Learning Points :**

Carrying out text emotion analysis using rule-based method, either sentence or chapter are divided Analysis, is all strongly depend on sentiment dictionary, and the quality that builds of sentiment dictionary directly influences the accuracy of final sentiment analysis.

**23. Summary of Invention :**

The invention discloses a kind of subjective texts sentiment analysis method based on deep learning, including: (1) on the basis of C&W models, build C&W SP models, by the affective tag of sentence and part of speech label for labelling in sentence, build the training set of C&W\_SPC&W SP models, and C&W\_SP models are trained using the training set, the term vector of each word in training set is obtained, term vector file is constituted; (2) according to the term vector file for obtaining, using LSTM model construction sentence vector sets; (3) neural network model is trained using sentence vector set, obtains sentiment classification model; (4) test comment sentence is pre-processed, be input to test sentence vector in sentiment classification model by test sentence vector, is calculated this section of Sentiment orientation of comment. The method Sentiment orientation information, part-of-speech information are added in word, improve the degree of accuracy of sentiment analysis.

**24. Number of Claims** : 7

**25. Patent Status** : Granted Patent & In-force Patent

**26. How much this invention is related with your IDP/UDP? :** 71 to 80%

**27. Do you have any idea to do anything around the said invention to improve it? :**

No

**PSAR5:-**

PSAR No. : 20BE7\_170020107041\_4

**Part - I : PATENT SEARCH TECHNIQUE USED**

- 1. Patent Search Database Used : Google Patents
- Web Link of the Database : <https://patents.google.com/>
- 2. Keywords Used for Search : twitter ,sentiment,analysis
- 3. Search String Used : twitter sentiment analysis
- 4. Number of Results/Hits getting : 5560

**Part - II : BASIC DATA OF PATENTED INVENTION/BIBLIOGRAPHIC DATA**

- 5. Category/Field of Invention :
- 6. Invention Is Related to Class of Invention : computer-implemented method of performing sentiment analysis
- 6a. IPC class of the studied patent : G06N20/00, G06F16/355, G06F40/30, G06Q30/02
- 7. Title of Invention : Performing sentiment analysis
- 8. Patent No. :
- 9. Application No. : 13/280,031
- 9a. Web Link of the studied patent : <https://patents.google.com/patent/US20130103385?q=twitter+sentiment+analysis>
- 10. Date of Filing/Application : 24/10/2011
- 11. Priority Date :
- 12. Publication/Journal Number - (Issue No. of Journal in which Patent is published) :
- 13. Publication Date :
- 14. First Filed Country :

**15. Also Published as**

We do not find any published data.

**16. Inventor**

| Name of Inventor | Address/City/Country of Inventor |
|------------------|----------------------------------|
| Riddhiman Ghosh  | US                               |
| Lei Zhang        | US                               |
| Mohamed E Dekhil | US                               |

**17. Applicant**

| Name of Applicant/Assignee                | Address/City/Country of Applicant |
|---|-----------------------------------|
| Hewlett Packard Enterprise Development LP | US                                |

- 18. Applicant for Patent is : Company

### Part - III : TECHNICAL PART OF PATENTED INVENTION

**19. Limitation of Prior Technology/Art :**

It may be difficult to identify and determine opinions for entities effectively using this class of data. Microblogs are typically short. For example, a tweet, which is a post on the TWITTER® network, is currently limited to a maximum of 140 characters. The data sources may also be noisy and prone to spam. Another issue is that opinions are commonly expressed in a manner specific to the domain, thereby incorporating emoticons, colloquial expressions, abbreviations, acronyms, neologisms, etc. These characteristics may increase the computational complexity of sentiment analysis on microblogs.

**20. Specific Problem Solved/Objective of Invention :**

One approach to perform sentiment analysis on entities in microblogging data is machine learning-based method. That is, we train a sentiment classifier to determine positive, negative and neutral sentiments. The method has been used for sentiment classification of documents or sentences. However, it is not easy to be applied for mining opinion on entities in microblogging data. Firstly, manual labeling of a large set of microblog training examples for a classifier is labor-intensive and time-consuming. Secondly, manual labeling often needs to be done for each application domain. It is not scalable for microblogging data which covers almost all domains as people can express opinions about anything on microblogs.

**21. Brief about Invention :**

Sentiment analysis is the computational study of people's opinions, appraisals, and emotions toward entities (e.g. products for sale, organizations, events and individuals). A growing online source of public opinion is the microblog. Microblogs are short web blogs, and are often posted on social media websites (e.g. FACEBOOK® and TWITTER®). People may provide, in microblogs, personal statuses, tweets, comments, etc. Opinions mined from microblogs may provide advantageous insights into public sentiments on various entities.

**22. Key Learning Points :**

In one embodiment, sentiment analysis on microblogs may be performed by applying a hybrid method by combining the opinion lexicon and sentiment classifier. In this way, sentiment orientations may be determined on entities discussed in microblogs.

**23. Summary of Invention :**

There is provided a computer-implemented method of performing sentiment analysis. An exemplary method comprises performing a first sentiment analysis on microblogging data based on a method using an opinion lexicon. The method also includes training a classifier using training data from the first sentiment analysis. Additionally, the method includes identifying a new opinion term in the microblogging data by performing a statistical test. The new opinion terms are not in the opinion lexicon. The method also includes identifying new microblogging data based on the new opinion term. Further, the method includes performing a second sentiment analysis on the new microblogging data using the classifier.

**24. Number of Claims** : 20

**25. Patent Status** : Granted Patent & In-force Patent

**26. How much this invention is related with your IDP/UDP? :** 71 to 90%

**27. Do you have any idea to do anything around the said invention to improve it? :**

No

## Appendix 4: Business Model Canvas and Report

### **GUJARAT TECHNOLOGICAL UNIVERSITY**

Chandkheda, Ahmedabad.



**Ahmedabad Institute of Technology,  
Ahmedabad**

### **BUSINESS MODEL CANVAS**

On  
**“Sentiment Analysis”**  
Under subject of  
Final Year Project  
B. E. Semester – VIII  
(Computer Engineering)

Group:- 125083

Submitted by:-

| Name of student | Enrollment No. |
|-----------------|----------------|
| Tirth Patel     | 170020107041   |
| Ravi Sahani     | 170020107049   |

Guided by:-  
**Prof. Bansari Thakkar**

Head of the Department  
**Prof. Ajay Upadhyaya**

Academic year  
(2020-2021)

1 | Page

**Ahmedabad Institute of TECHNOLOGY**

**Department of Computer Engineering**

**2020-2021**



**CERTIFICATE**

**Date: 28/04/2021**

This is to certify that the BUSINESS MODEL CANVAS report entitled “Sentiment Analysis” has been carried out by “Tirth Patel (170020107041), Ravi Sahani (170020107049)” under my guidance in fulfillment of the B.E. in Computer Engineering (8<sup>th</sup> Semester) of the Gujarat Technological University, Ahmadabad during the academic year 2020-21.

**Guided by:**

**Prof. Bansari Thakkar**

**Head of the Department**

**Prof. Ajay Upadhyaya**

➤ **INDEX**

| <b>TITLE</b>           | <b>PAGE NO.</b> |
|------------------------|-----------------|
| KEY PARTNERS           | 4               |
| KEY ACTIVITIES         | 4               |
| KEY RESOURCES          | 5               |
| VALUE PROPOSITIONS     | 5               |
| CUSTOMER RELATIONSHIPS | 6               |
| CHANNELS               | 6               |
| CUSTOMER SEGMENTS      | 7               |
| COST STRUCTURE         | 7               |
| REVENUE STREAM         | 8               |
| SNAPSHOT OF BMC        | 9               |

## ➤ KEY PARTNERS

- Investors
- Students
- News Reporter
- Tech Companies
- Fintech
- Government
- Secret Service
- NGOs
- Political Parties
- Influencers
- Social Media Companies

## ➤ KEY ACTIVITIES

- Positive Sentiment
- Negative Sentiment
- Neutral Sentiment
- Business Sentiment
- Mass Surveillance
- Product Review
- Behavioral Analysis

## ➤ KEY RESOURCES

- Internet Connection - As our project require much data so must have higher bandwidth networks, so we can operate with a stable 4G/ WIFI internet connectivity.
- Computing – Our project requires computation on data set like we have clean data, we have to merge data from multiple sources. So, a better computing device is must.
- Data - In our project data plays a crucial role because it is a sentiment analysis project.
- Web Hosting – The end user will interact with our website and will also be able to make use of the different functionalities that we will provide.

## ➤ VALUE PROPOSITIONS

- Sentiment of Company – Finding sentiment of company will be very helpful for Countries Security because it will help know the company's nature against a country.
- Sentiment of Person - Finding sentiment of a person will be very helpful to know the behavior of that person with other persons whether he/she is good or bad person.
- Sentiment of Politician - Finding sentiment of politician will be very helpful to know whether the politician is keeping his/hers promises that they made in the election and are they helping the poor peoples or not.

## ➤ CUSTOMER RELATIONSHIPS

- Bringing Innovative Technology to everyone.
- Cost Effective - Sentiment Analysis is very cost effective because we do not charge any money to our users. They only need their devices with fast internet connection.
- Security - Our aim to is provide sentiment analysis to everyone without tracking them or there is not search history or any other kind of data of that.
- It can be useful to appeal mass population about certain product.
- Can be helpful for political parties to influence certain group of people who are initially influenced by opposition parties.

## ➤ CHANNELS

- Mobile
- Computer
- Laptop
- Smart devices having screen with internet connection.
- Social Network
- Twitter, Instagram, Facebook

## ➤ CUSTOMERS SEGMENTS

- Social Worker
- Companies
- News Reporter
- Normal Person
- News Companies
- Broadcaster
- Entertainment Media
- Social Media Influencers

## ➤ COST STRUCTURE

- Web Hosting – Hosting our project is one the main cost for our project.
- Computation – For our project computation power of the device should high because we are having lots of computation on the data so less computation devices will perform this function quick.
- Internet – As our project totally depends on internet because we are getting our data for sentiment analysis from the twitter directly so higher bandwidth networks is required, so we must operate with a stable WIFI/4G internet connectivity.

## ➤ **REVENUE STREAMS**

- Low Effort – As our project will provide accurate data based on calculations done by the machine, so it will not require much effort.
- Cost Effective – In our project user only require to have smart device with internet connection.
- Easy to Access – Our project can be easily found on the website will be accessible by any user as it is designed in a user-friendly manner so that anyone can use it.
- Advertisement - The main source of income for our project is from ads on the website.

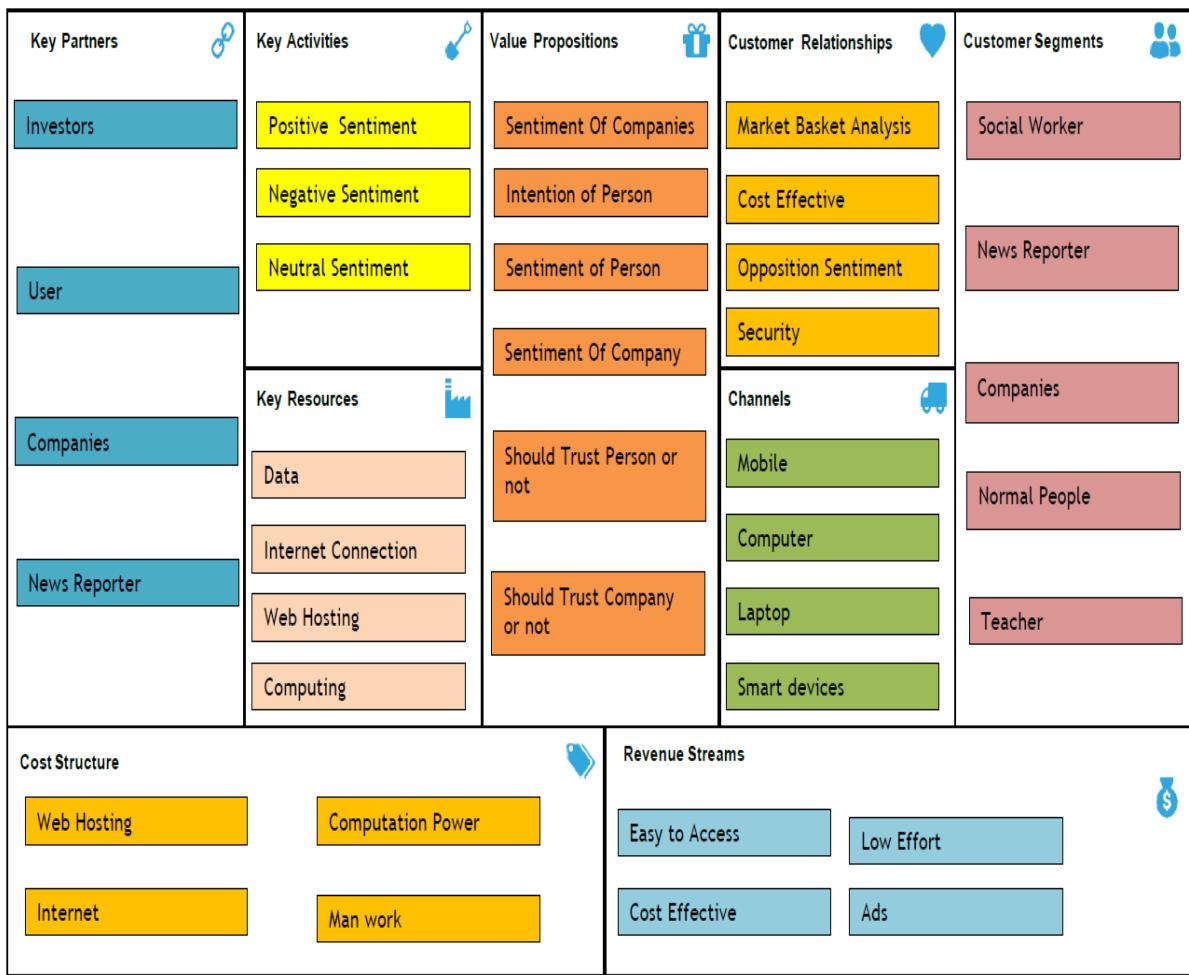
# BMC Canvas

Designed for: Sentiment Analysis

Designed by: Tirth Patel (170020107041),  
Ravi Sahani (170020107049)

Date: 28-04-2021

Version: 1



## Appendix 5: Patent Drafting Exercise

---

|                     |   |  |
|---------------------|---|--|
| <b>College</b>      | : | AHMEDABAD INSTITUTE OF TECHNOLOGY, GOTA, AHMEDABAD |
| <b>Department</b>   | : | Computer Engineering                               |
| <b>Discipline</b>   | : | BE   |
| <b>Semester</b>     | : | Semester 8   |
| <b>Project Name</b> | : | Sentiment Analysis                                 |
| <b>Team ID</b>      | : | 125083   |

### Form 1 – APPLICATION FOR GRANT OF PATENT

#### Applicants :

| Sr. No | Name                        | Nationality | Address  | Mobile No. | Email Id                |
|--------|-----------------------------|-------------|--|------------|-------------------------|
| 1      | Sahani Ravikumar Ramsajivan | Indian      | Computer Engineering ,<br>AHMEDABAD INSTITUTE OF<br>TECHNOLOGY, GOTA, AHMEDABAD ,<br>Gujarat Technological University. | 7405558622 | rsahani486@gmail.com    |
| 2      | Patel Tirth Pravinbhai      | Indian      | Computer Engineering ,<br>AHMEDABAD INSTITUTE OF<br>TECHNOLOGY, GOTA, AHMEDABAD ,<br>Gujarat Technological University. | 9978244308 | tirthpatel23t@gmail.com |

#### Inventors :

| Sr. No | Name                        | Nationality | Address  | Mobile No. | Email Id                |
|--------|-----------------------------|-------------|--|------------|-------------------------|
| 1      | Sahani Ravikumar Ramsajivan | Indian      | Computer Engineering ,<br>AHMEDABAD INSTITUTE OF<br>TECHNOLOGY, GOTA, AHMEDABAD<br>, | 7405558622 | rsahani486@gmail.com    |
| 2      | Patel Tirth Pravinbhai      | Indian      | Computer Engineering ,<br>AHMEDABAD INSTITUTE OF<br>TECHNOLOGY, GOTA, AHMEDABAD<br>, | 9978244308 | tirthpatel23t@gmail.com |

I/We, the applicant(s) hereby declare(s) that:

Following are the attachments with the applications :

## Form 2 - PROVISIONAL/COMPLETE SPECIFICATION

### 1. Title of the project/invention :

Sentiment Analysis

### 2. Preamble to the description :

Provisional

### 3. Description

#### a) Field of Project / Invention / Application :

Computer Engineering, Machine Learning

#### b) Prior Art / Background of the Project / Invention :

It can help revitalize a business' brand.

Companies can quantify people's perception about their products or services, marketing strategies and their customer experience.

When used right, it can help companies develop engaging marketing strategies to improve people's perception about their brands.

#### c) Summary of the Project / Invention :

Sentiment analysis in the domain of micro-blogging is a relatively new research topic so there is still a lot of room for further research in this area. Decent amount of related prior work has been done on sentiment analysis of user reviews , documents, web blogs/articles and general phrase level sentiment analysis . These differ from twitter mainly because of the limit of 140 characters per tweet which forces the user to express opinion compressed in very short text. The best results reached in sentiment classification use supervised learning techniques such as Naive Bayes and Support Vector Machines, but the manual labelling required for the supervised approach is very expensive. Some work has been done on unsupervised and semi-supervised approaches, and there is a lot of room of improvement.

#### d) Objects of Project / Invention :

The objectives of this project are:

- To implement an algorithm for automatic classification of text into positive and negative.
- Sentiment Analysis to determine the attitude of the mass is positive, negative or neutral towards the subject of interest.
- Graphical representation of the sentiment in form of Pie-Chart and Scatter Plot.
- Real time data analysis makes it possible for business organizations to keep track of their services and generates opportunities to promote, advertise and improve from time to time. Through comprehensive analysis, businesses gain valuable insights towards their customers.
- It can provide a competitive advantage- A trove of business opportunities and product possibilities.
- Gives predictions of consumer trends so companies can develop strategies to gain an advantage.
- It can help revitalize a business' brand. Companies can quantify people's perception about their products or services, marketing strategies and their customer experience. When used right, it can help companies develop engaging marketing strategies to improve people's perception about their brands.
- It can help enhance customer experience towards brands and business.

#### e) Drawings :

---

f) Description of Project / Invention : (full detail of project) :

The algorithm proposed works on Twitter Data, primarily it collects the tweets and then study it with the help of different statistical computing procedures.

In the age of artificial intelligence and machine learning, competition is between best and best.

For that sentiment of market is very important and sentiment of market is what consumer think of certain product.

The scope of the project is to provide a user friendly web based product that extracts people's sentiment feelings toward certain services, products, organizations, political or non-political topics and any influential people on social media. In this project phase which aims at developing a ~~filed~~ prototype, emphasis will be put on Arabic tweets from Twitter in the political domain.

The project aims to:

- Provide an accurate sentiment analysis results.
- Achieve a wide range of users in Egypt and the MENA region.
- Support Arabic Egyptian dialect in the first run and English will be considered later.
- Smooth, fast, efficient, reliable and easy to use web-based tool.
- Providing a user friendly menu and good entertainment visualization capabilities.
- Having a plenty of options in term of filtering and viewing information according to user's needs.

g) Examples :

h) Claims (Not required for Provisional Application) / Unique Features of Project

- It can be useful to business for analysing how to market product and what is the sentiment of people, how they will respond to product. It can be useful for predicting the tweets for a certain group of people.
- It can provide a competitive advantage for business opportunities and product possibilities.

4. Claims

5. Date and signature

6. Abstract of the project / invention :

Sentiment Analysis is contextual mining of text which identifies and extracts subjective information in source material, and helping a business to understand the social sentiment of their brand, product or service while monitoring online conversations.

Inorder to gain control over market, it is essential to understand market condition especially during covid-19 situation. For that sentiment of market is very important and sentiment of market is what consumer think of certain product.

Sentiment analysis is the interpretation and classification of emotions (positive, negative and neutral) within text data using text analysis techniques. Sentiment analysis tools allow businesses to identify customer sentiment toward products, brands or services in online feedback. The algorithm proposed works on Twitter Data, primarily it collects the tweets and then study it with the help of different statistical computing procedures.

## Form 3 – STATEMENT AND UNDERTAKING UNDER SECTION 8

Name of the applicant(s) : I/We, Sahani Ravikumar Ramsajivan ,Patel Tirth Pravinbhai

Name,Address and Nationality of the joint  
applicant : Hereby declare :

(i) that I/We have not made any application for the same/substantially the same victim invention outside India.

(ii) that the rights in the application(s) has/have been assigned to

| Name of the Country | Date of Application | Application Number | Status of the Application | Date of Publication | Date of Grant |
|---------------------|---------------------|--------------------|---------------------------|---------------------|---------------|
| N/A                 | N/A                 | N/A                | N/A                       | N/A                 | N/A           |

(iii)That I/We undertake that upto the date of grant of the patent by the Controller, I/We would keep him informed in writing the details regarding corresponding applications for patents filed outside India within three months from the date of filing of such application.

Dated this 30 day of April 2021

To be signed by the applicant or his authorised      Signature.....

registered patent agent :

Name of the Natural Person who has signed : Sahani Ravikumar Ramsajivan ,Patel Tirth Pravinbhai

To,  
The Controller of Patents,  
The Patent Office,  
At Mumbai

---