PROJECT REPORT

Group 23

Members and Contributors

Deeksha Behara
B00784704
Dk488492@dal.ca

Shipra Malik
B00783066
Sh809102@dal.ca

Faculty of Computer Science Dalhousie University

6 August, 2018

CONTENTS

1. INTRODUCTION 1
2. PROJECT OVERVIEW 1
   - 2.1 Project Goals 1
   - 2.2 Project Objective 1
   - 2.3 Functionalities 2
   - 2.4 Tentative Users 2
   - 2.5 Project Scope 3
   - 2.6 Project Closure 3
   - 2.7 Lessons Learned 3
3. USER MANUAL 3
   - 3.1 Application URL 3
   - 3.2 User Instructions 4
4. DESIGN 7
   - 4.1 Architecture Design 7
   - 4.2 Database Design 8
5. APPLICATION API 8
6. FRAMEWORKS USED 8
7. INFRASTRUCTURE SETUP 9
8. REFERENCES 10

1. INTRODUCTION

The project aims to build a web application that provides sentiment analysis on a topic by fetching the tweets from Twitter about that topic. Social media websites have emerged as one of the platforms to raise users' opinions and influence the way any business is commercialized. Sentiment analysis of the tweets determine the polarity and inclination of vast population towards specific topic, item or entity. Our project targets to provide general people's opinion fetched from Twitter on a topic in form of a visual graph or pie chart. We are aiming to classify people's sentiments in the form of positive sentiment, negative or neutral. This project could be used by individual users to know the trend of a topic, or by businesses to know their customer's opinions about their products.

2. PROJECT OVERVIEW

This project is a web application to analyze the sentiments of people on a particular topic and present the result to the end users in the form of a pie chart. The results in the form of pie chart or graph will show positive, negative or neutral sentiment of public for a topic. The goal of this project is to provide a platform to users or businesses to get people's opinions in a visual form. In today's customer centric society, people's opinion is everything that matters and our project is trying to resolve the gap between a business and its customer's opinions or to solve the curiosity of an individual to know public opinion about a topic. Sentiment analysis can help businesses or individuals to improve their marketing strategy, promote sales, get product reviews, generate leads or improve customer service. Using our platform, users would be able to visualize the sentiments.

2.1 Project Goals

The idea of the project is to provide a platform to end users to mine the opinions of general public about a product or a topic. Sentiment analyses is the computational study of people's opinions, sentiments, attitudes and emotions expressed in a written language. Twitter has given people a very good platform to express themselves. Now people can tweet their opinions and emotions on any topic on Twitter. It not realistic to read all the tweets on Twitter for any user or business Our idea is to use the opinions of people to generate useful information. Our project focusses on analyzing the sentiments of people on a topic to show the analyses in a visual form which could be used further by individuals or businesses.

This project has been converted to 50% by the professor.

2.2 Project Objective

Our project targets to visualize the sentiments of people expressed on Twitter for a topic. Sentiment analyses popularity is mainly due to two reasons. First, opinions are central to almost all human activities and are key influencer to our behaviors. Second, it provides an area of research not only in businesses but also in terms of management, social sciences and health. The customers of our project are businesses, government, politicians or

individual people. It not realistic to read all the tweets on Twitter for any user or business. So, our platform classifies the sentiments in positive, negative or neutral and visualize it to the customers in the form of a graph.

This application will be used by individuals to know people's opinion, as we humans need to know the opinion of people before taking any decision. Our project aims to target businesses as well as reading the tweets for a topic is costly affair to them, so they can use our application to know about the opinions of people on their products, or to improve their customer service and for introduction of new products. This application could be useful for Politicians as well because public opinions matter most to them to shape their campaigns. Another user of this project could be governments as they can reshape their policies according to the people's sentiments.

2.3 Functionalities

Functionalities	Description
Home Page	**Home Page which allows the users to enter the topic and number of tweets to fetch.**
Sentiment categorization	**Fetches the tweets, analyses the sentiments and categorizes into positive, neutral and negative sentiment in terms of percentage.**
Database Storage	**Inserts the topic and its sentiment category percentages into a database.**
Sentiment Analysis	**Visualize the sentiment categories (positive, negative, neutral) in the form of pie chart for the topic searched.**
Word Cloud Analysis	**Displays the word cloud of most popular hashtags on twitter for the topic searched.**
Trend Analysis	**Visualizes and compares the trend of people's sentiments over the past for the same topic searched by fetching the values from database.**

2.4 Tentative Users

The users of the application would be consumers of any service, applications or products. The application built would be useful for the consumers who would like to get opinion about the product they would like to buy or the services they would like to choose. It is also applicable to vendors who are running a business. The proposed application can help the vendors identify the problems in the services they provide and help promote their sales.

2.5 Project Scope

The scope of the project is defined in the previous section. However, the project is not personalized for individual users. It is generic and meant for public use. The searches of twitter topics are limited to the trending topics in the platform. Though there are wide range of topics being discussed on social media, there are certain topics which are hardly or never discussed. This project is limited to the scope of the trending topics on social media.

Also, Twitter API keeps a cap on the number of tweets extracted from the platform given a period of time. In order to get better results, the tweets extracted for the given search within a certain period can be increased by implementing the twitter streaming or twitter firehose. The results would be generic and accurate with more data points however, it would be challenging to contain the tweets streamed form the platform on limited cloud resources such as EC2 instance provided from Amazon.

This project is built on the assumption that the functionalities and the data retrieval and storage capacities are within the scope and the capabilities of the cloud services provided by the service providers.

2.6 Project Closure

Our project has been closed. We have submitted Project Proposal and have shown the demo to the business, that is, our professor. We have released the resources, that is, human resources and logistics for the project. We have also done a retrospective meeting to learn the good and bad things about the process we followed in the project.

2.7 Lessons Learned

We have successfully delivered the project to our client, that is, Dalhousie University. We have learned various new technologies while implementing the project. We have worked on Agile methodology, so now we know this process. We have also gone through manual testing phase. We need to improve upon project estimation and deadlines. The project was estimated to be complete early, but it took more time for completion.

3. USER MANUAL

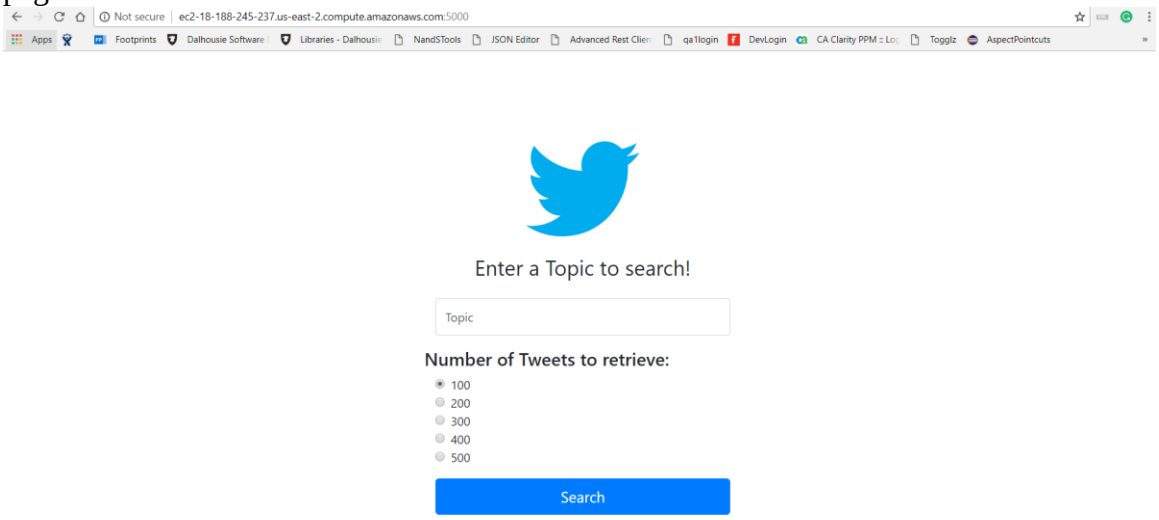
3.1 Application URL

The application has been deployed to EC2 instance. The URL is :

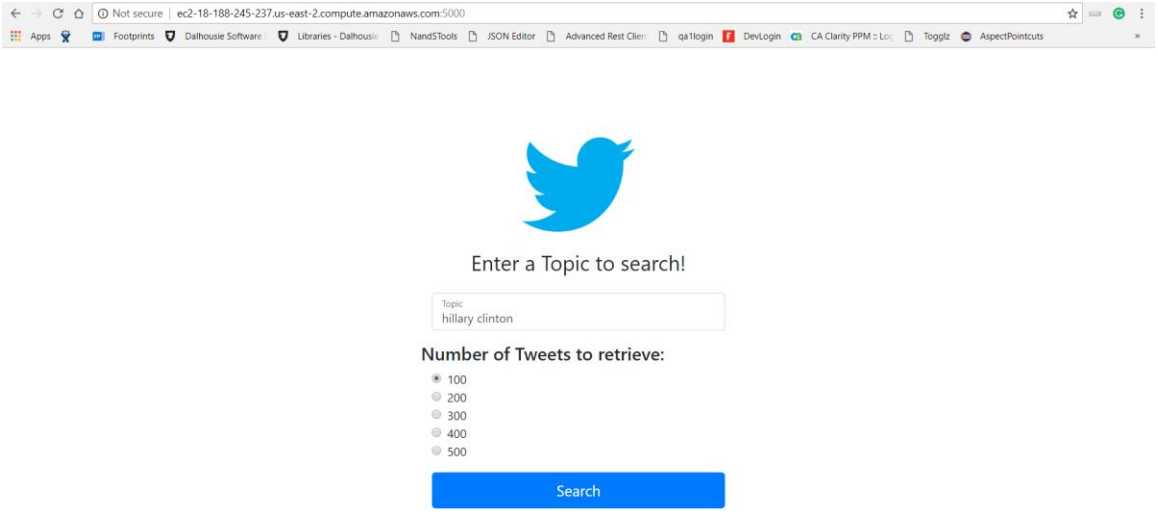
http://ec2-18-188-245-237.us-east-2.compute.amazonaws.com:5000

3.2 User Instructions

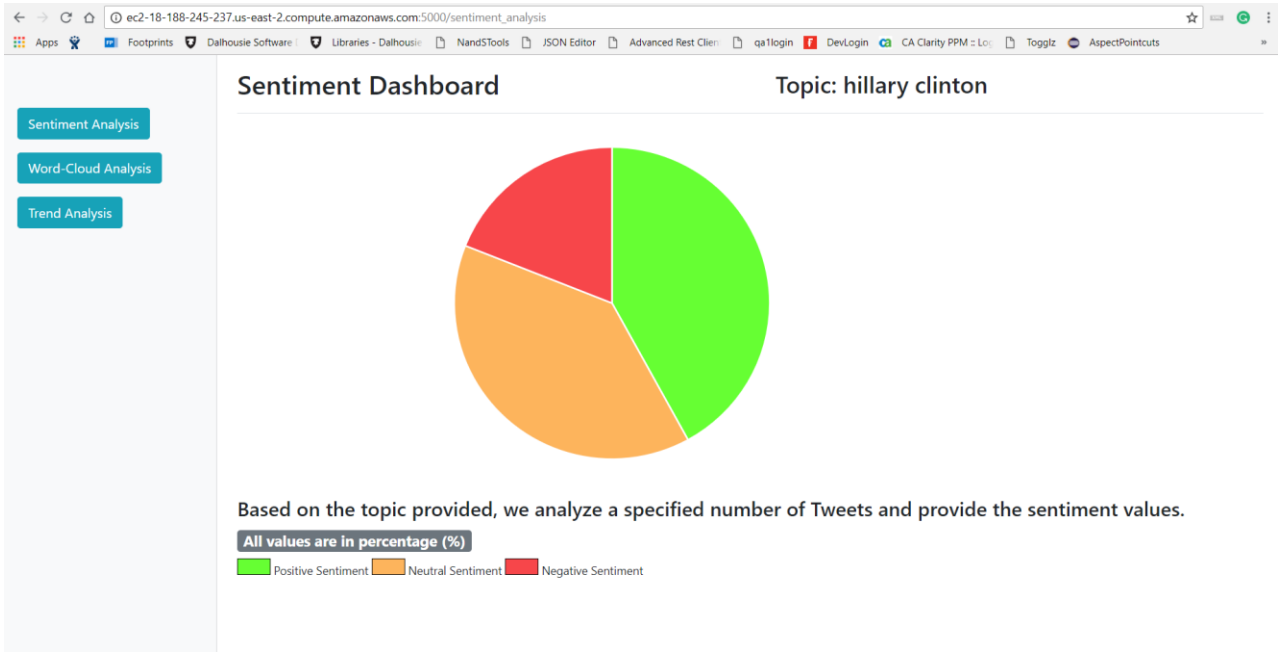
a. The user needs to go the above-mentioned application URL, he will land up to home page.



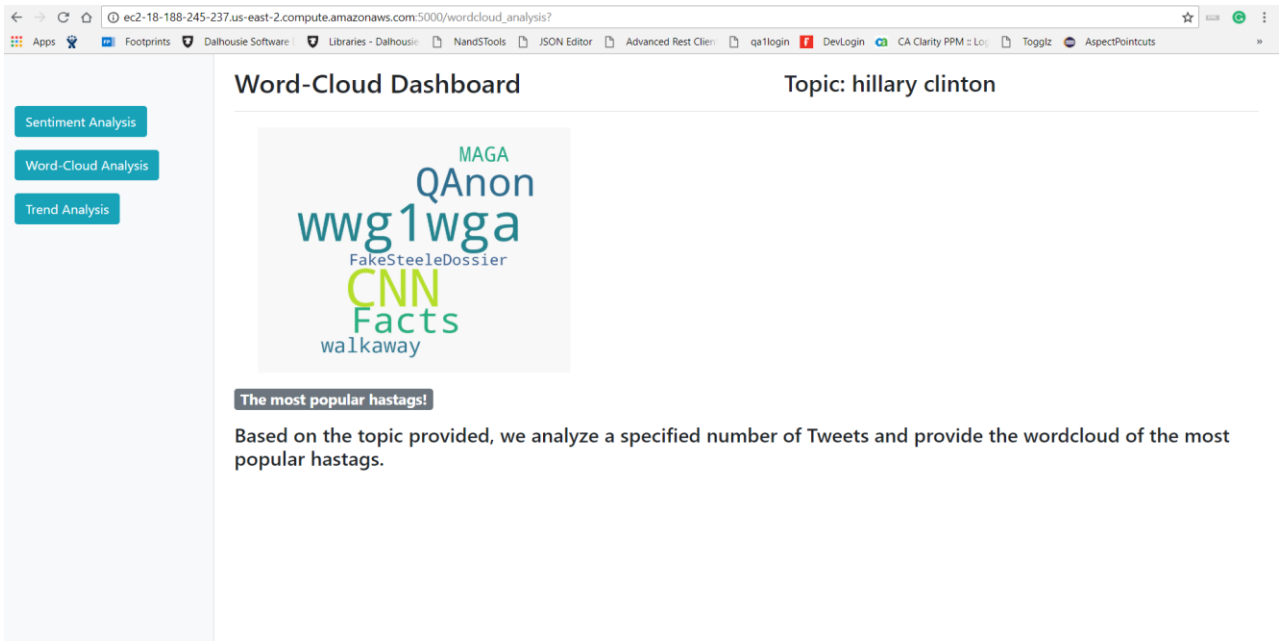
b. The user will input the topic such as Trump and select the number of tweets to fetch on home page.



c. The user will be shown sentiment analysis for the topic he searches in the form of positive, neutral and negative sentiment percentages in the form of a pie chart. Parallelly, the topic and corresponding sentiments are inserted into database.

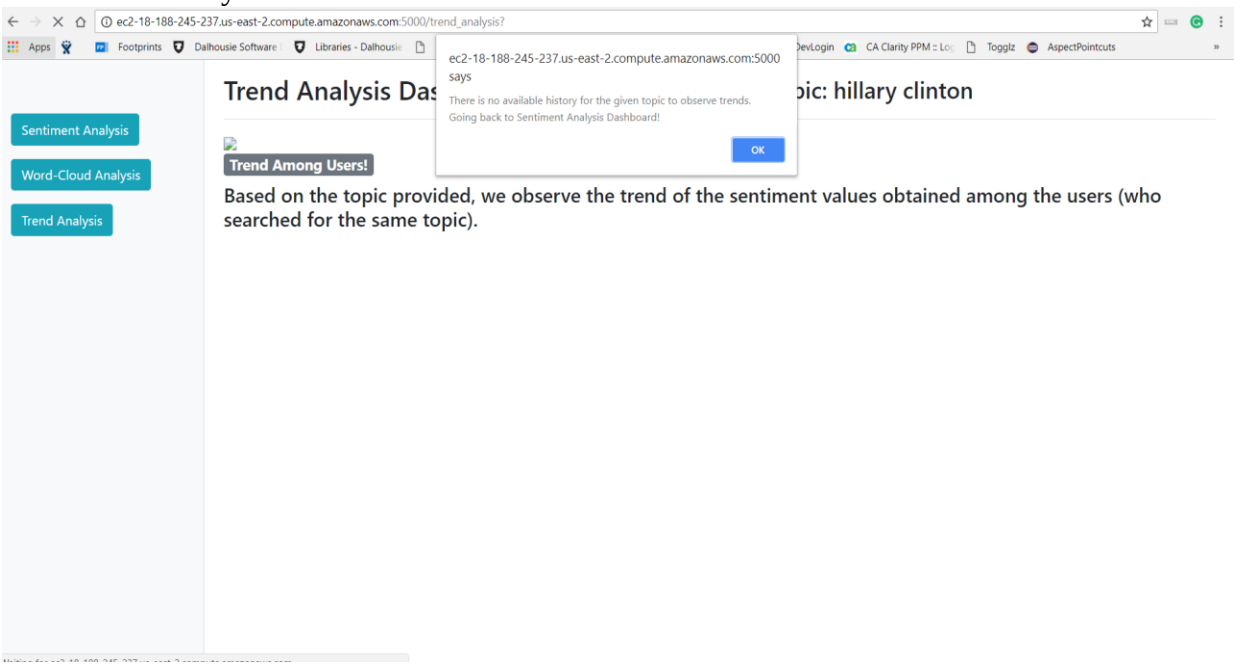


d. Click on Word count analysis from the left bar , the user will be shown the most popular hash tags for the topic searched in the form of a word cloud.

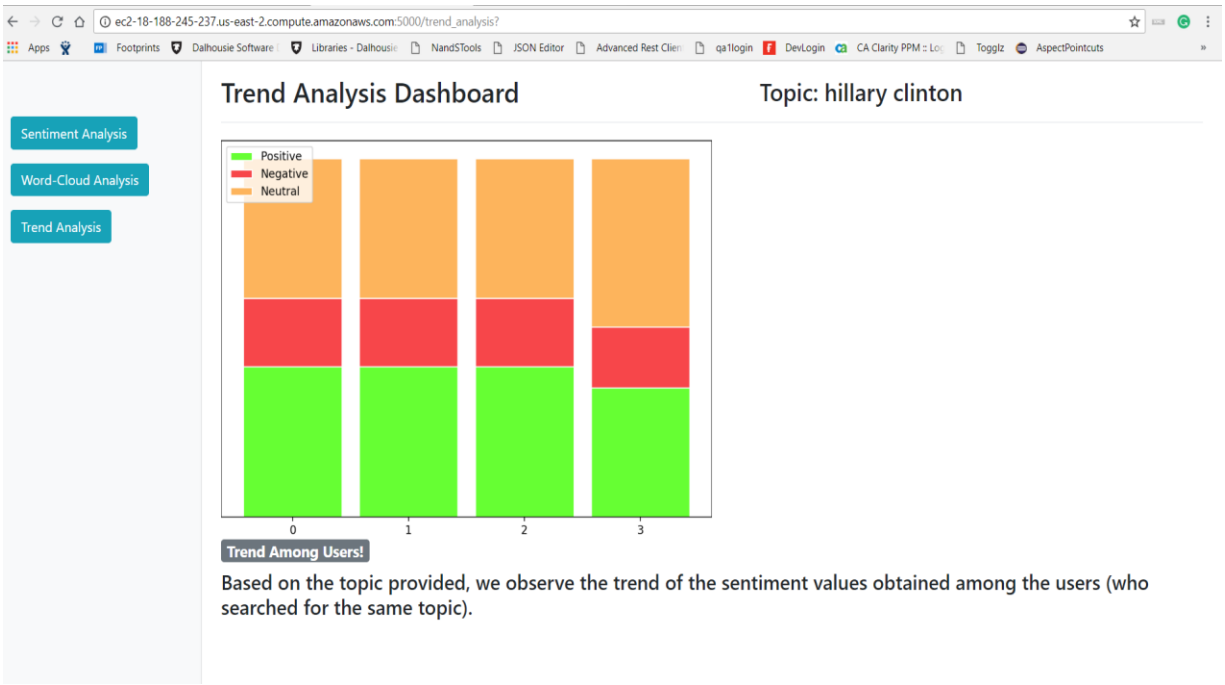


e. Click on Trend Analysis from the left bar, the user will be shown the trend of sentiments over the past for the same topic in the form of a bar graph.

i If no user has already searched for that topic, then an alert message is displayed that no user has searches for this topic yet. After we click OK, page is redirected to sentiment analysis.



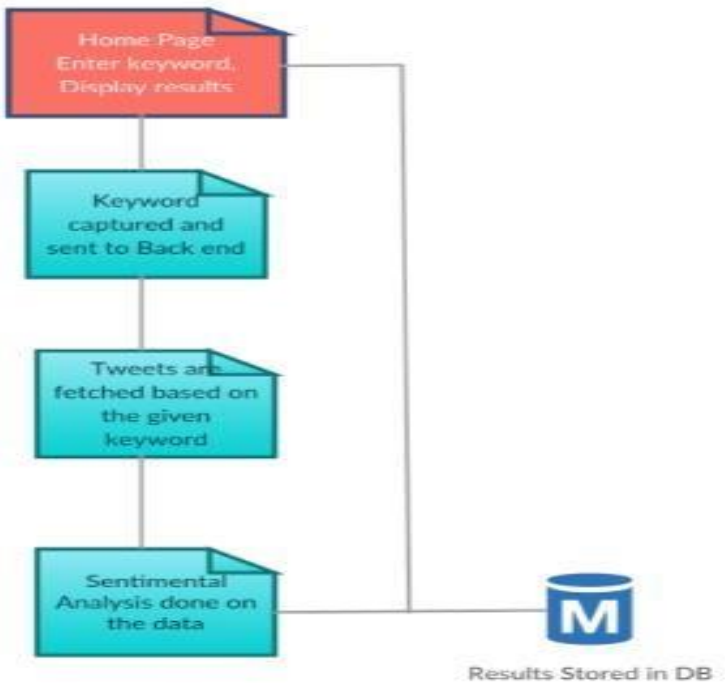
ii If a user has already searched for that topic, then trend analysis of that topic is displayed compared to the past searches.



4. DESIGN

4.1 Architecture Design

a. User enters topic, select number of tweets on Home Page.
b. Home Page interacts with Python web service using Flask.
c. Python web service interacts with twitter through python back end to fetch the tweets.
d. Web service does the sentiment analysis and send the response to HTML front end.
e. Home page redirects the results to sentiment analysis page, where a pie chart showing sentiments for that topic is displayed and a record for the topic and sentiment is inserted into database.
f. User clicks Word cloud analysis, same process happens as described above, HTML interacts with Python web service to respond with most popular hash tags for that topic.
g. The response goes to word_analysis html page where word cloud of tags is displayed.
h. User clicks Trend Analysis, html calls a python web service, which retrieves the sentiments for a topic in database, and sends the response to html which in turn shows the bar chart.



4.2 Database Design

. The database contains the search topic and the corresponding sentiment percentages.

Attributes	DataType
ID	Int
Search (Keyword)	Varchar
POSTIVE_VALUE	Long
NEGATIVE_VALUE	Long
NEUTRAL_VALUE	Long

ID	SEARCH	POSITIVE_VALUE	NEGATIVE_VALUE	NEUTRAL_VALUE
1	Twitter	13.00	7.50	79.50
2	Roonev	19.33	19.33	61.33
3	Keke	38.50	16.50	45.00
4	Drake	34.00	14.00	52.00
5	Keke	38.50	17.25	44.25
6	Pupper	19.80	17.20	63.00
7	Ronaldo	24.00	3.67	72.33
8	Salt	28.00	15.67	56.33
9	Football	37.00	10.00	53.00
10	Computer	84.67	4.00	11.33
11	Motorola	17.67	3.33	79.00
12	ManUtd	51.33	2.67	46.00
13	League o...	30.75	7.75	61.50
14	Pizza	24.00	13.33	62.67
15	India	36.00	13.33	50.67
16	javascript	35.50	10.00	54.50
17	Suits	44.00	13.67	42.33
18	ManUtd	52.33	2.00	45.67

5. APPLICATION API

The URL of our Application is:
http://ec2-18-188-245-237.us-east-2.compute.amazonaws.com:5000

6. FRAMEWORKS USED

We have used various third party frameworks while developing application.

1. Front End Development:
    - HTML for building the web pages.
    - CSS/Bootstrap for Styling.
    - Chart.min js for plotting charts.

2. Back End Technologies:
    - Python for fetching tweets and performing sentimental analysis.
    - Libraries used for sentiment, wordcloud and trend analysis are Tweepy, textblob, wordcloud, numpy, scipy, csv, pandas etc.
    - Libraries used for image are PIL.
    - Libraries used for plotting graphs and charts are matplotlib.
    - Libraries for database connectivity are pymysql.
3. Data exchange from front end to back end:
    - Flask for establishing the connectivity between HTML and python program.
    - Libraries used are flask_sqlalchemy, requests and sqlalchemy.
4. Database:
    - Amazon RDS to store topic and its related sentiments.
5. Visualization:
    - Charts.js with Flask to visualize the data.

7. INFRASTRUCTURE SETUP

Development Infrastructure:

a. Create a Twitter account and generate OAuth credentials for fetching the tweets.
b. Create an account on aws.amazon.com.
c. Create an RDS database on aws.amazon.com.

Deployment Steps:

a. configure an EC2 instance for Ubuntu CLI for deployment.
b. Transfer the project to Ubuntu Home directory in EC2.
c. Login to EC2 Ubuntu CLI.
d. Install python 3 and pip3
e. Install all the required libraries using pip3 install 'libraryname'.
f. All third-party libraries include Textblob, Flask, sqlAlchemy, wordcloud, pandas, tweepy, Scipy and flask_sqlalchemy
g. Run the project using the following command: sudo python3 app.py
h. The project will run on EC2 public DNS url. Go to browser and hit the public DNS url of the EC2 instance. This will open the application home page.

8. GITHUB LINK

We have used GitHub for hosting and reviewing the code. Below is the GitHub link.
https://github.com/shipra1101/twitter-sentiment-visualization

9. REFERENCES

1. Docs - Twitter Developers. (n.d.). Retrieved from https://developer.twitter.com/en/docs.html
2. Natural Language Toolkit¶. (n.d.). Retrieved from https://www.nltk.org/
3. Welcome. (n.d.). Retrieved from http://flask.pocoo.org/
4. JS Charting. (n.d.). Retrieved from http://www.jscharting.com/
5. Anto, M. P., Antony, M., Muhsina, K. M., Johny, N., James, V., & Wilson, A. (2016, March). Product rating using sentiment analysis. In *Electrical, Electronics, and Optimization Techniques (ICEEOT), International Conference on* (pp. 3458-3462). IEEE.
6. Amazon Web Services (AWS) - Cloud Computing Services. (n.d.). Retrieved from https://aws.amazon.com/
7. AYLIEN. (2018). Why is Sentiment Analysis important from a business perspective. (n.d) Retrieved from http://blog.aylien.com/why-is-sentiment-analysis-important-from-a/
8. Sentiment Analysis Importance (n.d). Retrieved from https://www.linkedin.com/pulse/importancesentiment-analysis-social-media-christine-day/