Mini Project Report on

Twitter Sentiment Analysis

Submitted in partial fulfilment
for the degree of Bachelor of Technology in
Data Science

Submitted by
Komal Rote
Shruti Rahate

Under the guidance of
Ms Merrin Soloman

Usha Mittal Institute of Technology
S.N.D.T Women's University
Juhu Tara Road,Sir Vitthaldas Vidyavihar,Santacruz(West),
Mumbai - 40049

2022-2023

Approval Sheet

This is to certify that **Komal Rote and Shruti Rahate** has completed the mini project report on the topic " **Twitter Sentiment Analysis**" satisfactorily in partial fulfillment for the Bachelor's Degree in Data Science under the guidance of Ms Merrin Soloman during the year 2023-24 as prescribed by Usha mittal Institute of Technology

Guide

Ms Merrin Soloman

Head Of Department

Mr Rajesh Kolte

Principal
Name of Principal

Examiner 1

Examiner 2

Declaration

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

(Signature)

Komal Rote (2015060) / Shruti Rahate (2015057)

Date

Abstract

With the rise of social networking epoch and its growth, Internet has become a promising platform for online learning, exchanging ideas and sharing opinions. Social media contains huge amount of sentiment data in the form of tweet, blogs, and updates on the status, posts, etc. In this paper, the most popular micro blogging platform twitter is used. Twitter sentiments analysis is an application of sentiment analysis on data from Twitter (tweets). To extract user's opinions and sentiments. The main goals is to explore how text analysis techniques can be used to dig into some of the data in the series of posts focusing on different trends of tweets languages, tweets volumes on twitter. Experimental evaluations show that the proposed machine learning classifiers are efficient and performs better in terms of accuracy. The proposed algorithm is implemented in Machine learning.

Contents

Abstract

1 Introduction
1.1 Objectives of the Study
1.2 Problem Statement

2 Literature Survey
2.1 Technical Papers

3 Existing System

4 Proposed System

5 Architecture

6 Algorithm

7 Hardware and Software Requirement
7.1 Software Requirements
7.2 Hardware Requirement

8 Implementation 9 Future Scope

References

Chapter 1

Introduction

A Twitter sentiment analysis is the process of determining the emotional tone behind a series of words, specifically on Twitter. A sentiment analysis tool is an automated technique that extracts meaningful customer information related to their attitudes, emotions, and opinions.A Twitter sentiment analysis identifies negative, positive, or neutral emotions within the text of a tweet. It is a text analysis using natural language processing (NLP) and machine learning. It identifies and extracts subjective information from original data, providing a company with a better understanding of the social sentiment of its brand, product, or ser- vice. Sentiment analysis is frequently used to analyze customer feedback, survey responses, and product reviews.

1.1 Objectives of the Study

The objective of this task is to detect hate speech in tweets. For the sake of simplicity, we say a tweet contains hate speech if it has a racist or sexist sentiment associated with it. So, the task is to classify racist or sexist tweets from other tweets. Formally, given a training sample of tweets and labels, where label '1' denotes the tweet is racist/sexist and label '0' denotes the tweet is not racist/sexist, your objective is to predict the labels on the given test dataset

1.2 Problem Statement

A basic task in sentiment analysis is classifying the polarity of a given text at the document, sentence, or feature/aspect level — whether the expressed opinion in a document, a sentence or an entity feature/aspect is positive, negative, or neutral. Advanced, "beyond polarity" sentiment classification looks, for instance, at emotional states such as "angry", "sad", and "happy". The problem in sentiment analysis is classifying the polarity of a given text at the document, sentence, or feature/aspect level whether the expressed opinion in a document, a sentence or an entity feature/aspect is positive, negative, or neutral .

Chapter 2

Literature Survey

2.1 Technical Papers

1) **Twitter Sentiment Analysis: The Good the Bad and the OMG!" by Saif, Fern´andez, He and Alani (2013)** - This paper offers a comprehensive review of different approaches to sentiment analysis on Twitter, including lexicon- based, machine learning-based, and hybrid methods. It also discusses challenges and limitations, such as sarcasm, ambiguity, and context-dependency.

2) **Sentiment Analysis and Opinion Mining" by Pang and Lee (2008)** - This seminal paper provides an overview of sentiment analysis and opinion min- ing, including its history, applications, and challenges. It also proposes a frame- work for classification tasks and discusses different approaches, such as machine learning, lexicon-based, and rule-based methods.

3) **Deep Learning for Sentiment Analysis: A Survey" by Zhang, Zhao, and LeCun (2018)** - This survey paper focuses on the use of deep learning techniques, such as neural networks and convolutional neural networks (CNNs), for sentiment analysis on Twitter. It discusses different models and architectures, such as LSTM, GRU, and attention mechanisms, and their performance compared to traditional methods.

4) **Twitter Sentiment Analysis: A Comparative Study of Pre- processing Techniques" by Malik and Singh (2020)** - This paper compares the performance of different pre-processing techniques, such as stemming, stop word removal, and feature extraction, on Twitter sentiment analysis using ma- chine learning algorithms. The authors also discuss the impact of data imbalance and selection bias on the results.

5) **Sentiment Analysis of Twitter Data: A Comparative Study of Feature Selection Methods" by Kumar, Pandey, and Pal (2018)** - This paper compares the performance of different feature selection methods, such as TF-IDF, Chi-squared, and Mutual Information, on sentiment analysis of Twitter data using machine learning algorithms. The authors also discuss the impact of different feature dimensions and classifiers on the results.

Chapter 3

Existing System

1) **IBM Watson:** IBM Watson is a powerful artificial intelligence system that offers a range of cognitive services, including sentiment analysis for Twitter data. The system uses machine learning algorithms to detect emotions, opinions, and attitudes from Twitter feeds and produces a score from -1 to +1 for each tweet.

2) **Google Cloud Natural Language API:** Google Cloud Natural Language API provides a range of sentiment analysis features, including identifying positive, negative, neutral, and mixed sentiments in text. The system uses machine learning models to analyze Twitter data for sentiment analysis.

3) **Sentiment140:** Sentiment140 is a popular open-source tool for Twitter sentiment analysis that provides a user-friendly interface to analyze Twitter data. The system uses machine learning algorithms to automatically classify tweets as positive, negative, or neutral based on the presence of certain keywords.

4) **NLP Cloud:** NLP Cloud provides a range of natural language processing services, including sentiment analysis for Twitter data. The system uses a deep learning model to analyze tweets and produce a sentiment score from -1 to +1 for each tweet.

5) **Rapid Miner:** Rapid Miner is a data science platform that offers a range of tools for sentiment analysis and text analytics, including for Twitter data. The system uses machine learning models to analyze tweets and produce sentiment analysis results based on the presence of certain keywords and patterns.

Chapter 4

Proposed System

The Proposed solution is an Web Application, “Machine Learning Analytics With Twitter Data On Sentiment Analysis” Which is used to performing Sentiment Analysis on Twitter is trickier than doing it for large reviews.Overall, this pro- posed system for Twitter sentiment analysis can help businesses and organizations to gain insights into the public opinion about their brand, products or services on Twitter.

They can use this information to improve their customer experience, product de- sign, and marketing strategy. This is because the tweets are very short (only about 140 characters) and usually contain slangs, emoticons, hash tags and other twitter specific jargon. It is developed on web based platform so it can be used easily and anywhere with the use of internet on any your laptop, personal computer, or on your cell phone. The user interface is very easy to understand and uses that it could be used by individuals of all ages.

Naive Bayes is a probabilistic algorithm that works well for text classification problems like sentiment analysis. It assumes independence between features, and uses the Bayes Theorem to calculate the probability of a particular sentiment given the occurrence of words in a tweet.

Chapter 5

Architecture

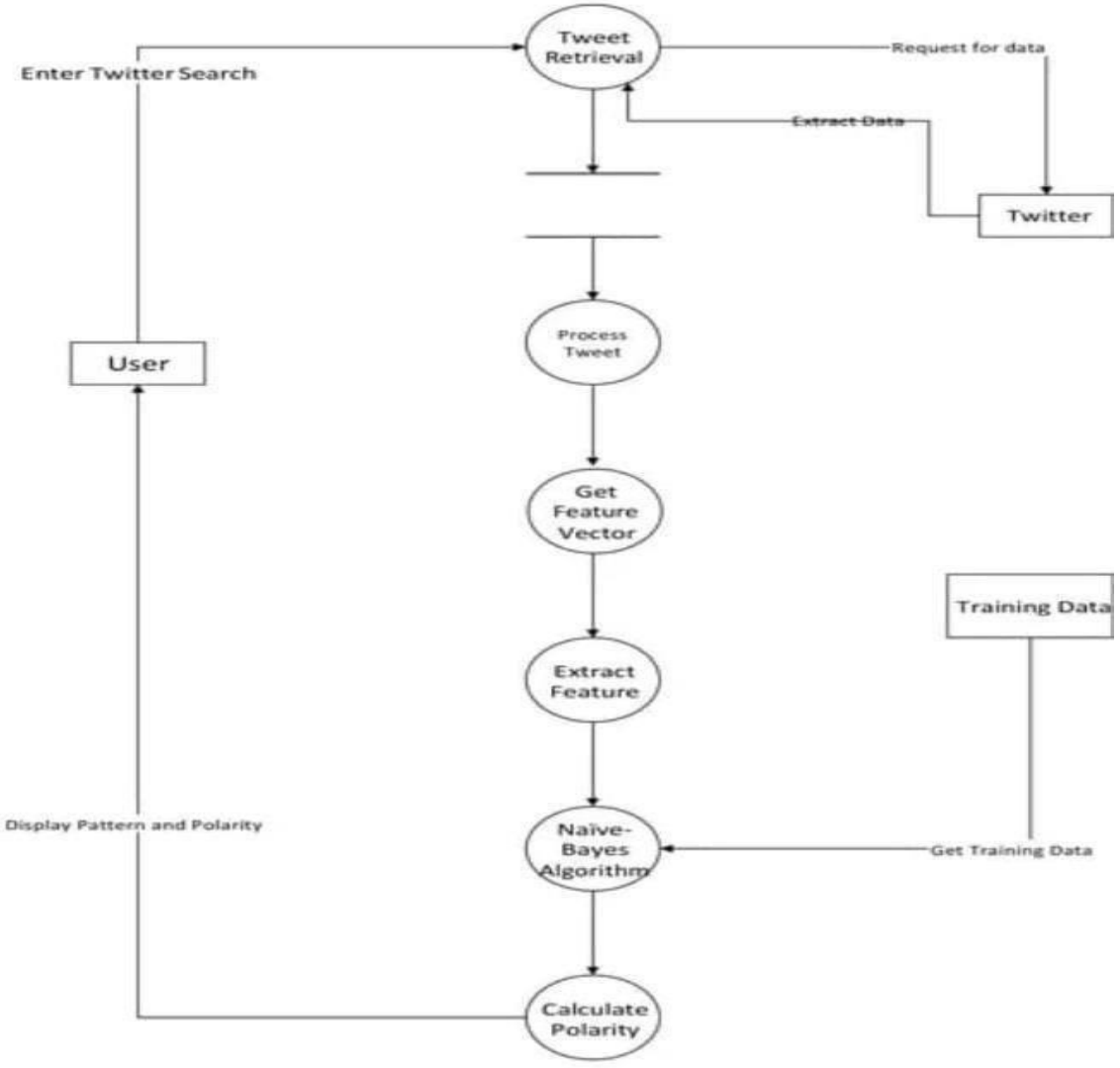


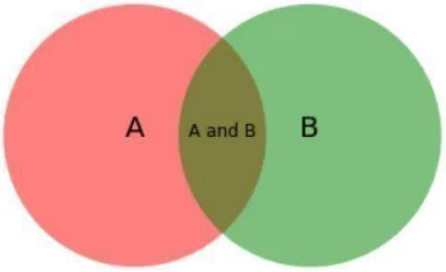
Figure 5.1: Architecture of twitter sentiment analysis project

Chapter 6

Algorithm

Millions of tweets are posted every second. It helps us know how the public is responding to a particular event. To get the sentiments of tweets, We can use the Naive Bayes classification algorithm, which is simply the application of Bayes rule.

Bayes Rule is merely describing the probability of an event on prior knowledge of the occurrence of another event related to it.



Then the probability of occurrence of event A given that event B has already occurred is

P(A — B) = P(A and B) P(B)
P(Event 'A')= Probability of occurrence of event A

And for the probability of occurrence of event B given that event A has already occurred is

$$P(B \mid A) = \frac{P(A \text{ and } B)}{P(A)}$$

Using both these equations, we can rewrite them collectively as

$$P(B|A) = \frac{P(A|B)*P(B)}{P(A)}$$

A,B= Events
P(A|B)=Probability of 'A' given is True

P(A),P(B)=The Independent Probabilities of 'A' and 'B'

Let's take a look at tweets and how we are going to extract features from them We will be having two corpora of tweets, positive and negative tweets.

Positive tweets:'I am happy because I am learning NLP,' 'I am happy, not sad.'

Negative tweets:'I am sad, I am not learning NLP,' 'I am sad, not happy.'

TextBlob –

TextBlob it uses NLTK (Natural Language ToolKit) and the input contains a single sentence, The output of TextBlob is **polarity** and **subjectivity**. Polarity score lies between (-1 to 1) where -1 identifies the most negative words such as ***'disgusting', 'awful', 'pathetic',*** and 1 identifies the most positive words like ***'excellent', 'best'.*** Subjectivity score lies between (0 and 1), It shows the amount of personal opinion, If a sentence has high subjectivity i.e. close to 1, It resembles that the text contains more personal opinion than factual information. I was more concerned about the Polarity score as my objective was not to identify factual information, so I skipped the subjectivity score in my project.

As TextBlob is a Lexicon-based sentiment analyzer. It has some predefined rules or we can say word and weight dictionary, where it has some scores that help to calculate a sentence's polarity. That's why the Lexicon-based sentiment analyzers are also called "Rule-based sentiment analyzers".

Chapter 7

Hardware and Software Requirement

7.1 Software Requirements -

- Operating System Windows 10
- Visual Studio code
- Twitter Developer Account
- Python
- TextBlob
- HTML and CSS

7.2 Hardware Requirement -

- Processor: Intel I5
- Motherboard: Intel Chipset Motherboard
- RAM : 8GB
- Hard Disk : 256 GB
- Speed : 2.7GHZ and more

Chapter 8

Implementation

1) **Data Collection:** Collect the data from Twitter by using the Twitter API or a public dataset.

2) **Pre-processing:** Pre-process the data to remove unwanted elements suchas stop words, punctuations, and URLs. Convert the words to lowercase toreduce the number of unique words in the dataset.

3) **Feature Extraction:** Extract features from the pre-processed text data byusing techniques such as bag-of-words, n-grams, or word embeddings.

4) **Train/Test Spli**t: Split the dataset into training and test sets to evaluatethe performance of the model.

5) **Training:** Use the Naive Bayes algorithm to train a model on the training set. The algorithm calculates the probability of each feature given each class label, and the prior probability of each class label.

6) **Validation:** Validate the model by applying it to the test set and calculate the accuracy, precision, recall, and F1 score.

7) **Real-time Prediction:** Apply the trained model to new tweets in real- time to predict their sentiment.

8) **Deployment**: Deploy the model as an API for real-time predictions or integrate it into a web application.

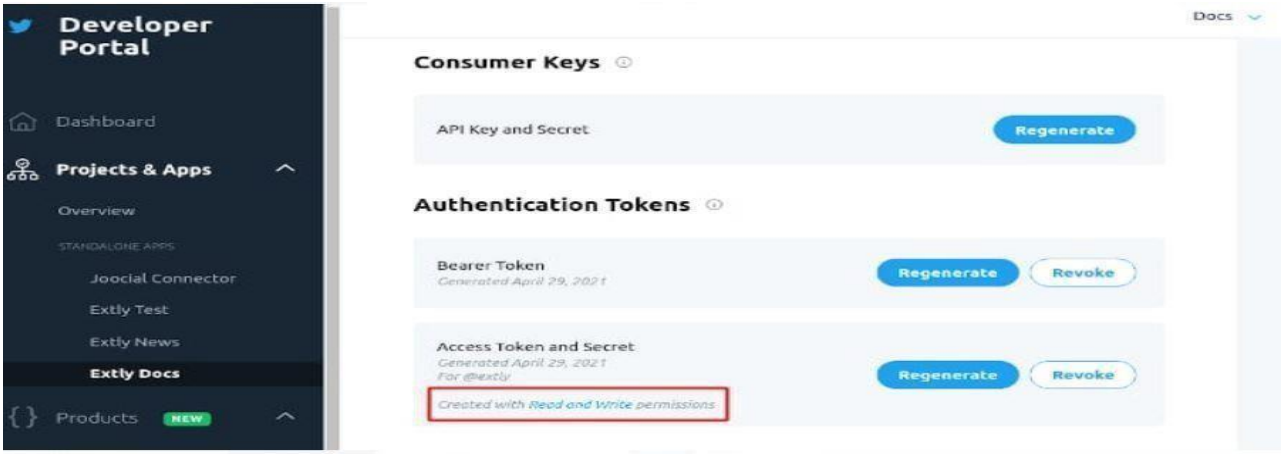


Figure 8.1: Creating twitter developer account to fetch twitter data from API

Your Twitter developer account application has been approved!

Thanks for applying for access. We've completed our review of your application, and are excited to share that your request has been approved.

Sign in to your developer account to get started.

Thanks for building on Twitter!

```
templates > index.html
<!DOCTYPE html>
<html>
<head>
    <title>Sentiment analysis</title>
    <link rel="stylesheet" type="text/css" href="{{ url_for('static', filename='css/Indexstyle.css'
    <link rel="stylesheet" href="https://stackpath.bootstrapcdn.com/bootstrap/4.4.1/css/bootstrap.m

</head>
<body>
    <header class="title">
      <div class="title-text">
        <h1 class="big-text">Twitter Sentiment Analysis</h1>
      </div>
    </header>

    <div class="mainsection">
        <div class="sub1">
            <div class="container">
            <div class="row">
                <div class="col-sm-12">
                  <div class="card">
                    <div class="card-body">
                      <h5 class="card-title">Phrase level sentiment analysis</h5>
                      <form action="/predict" method="post">
                        <div class="form-group">
                            <input type="text" class="form-control" name="query" placeholder="Enter
                        </div>
                        <div class="form-group">
```

Figure 8.2: Code of html

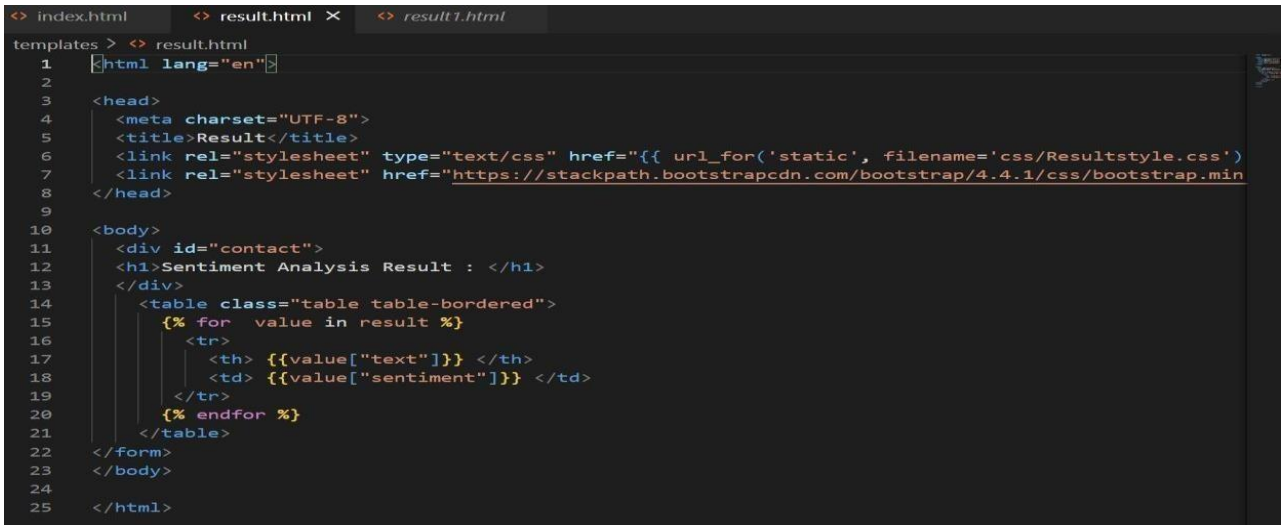


Figure 8.3: Code of html and CSS

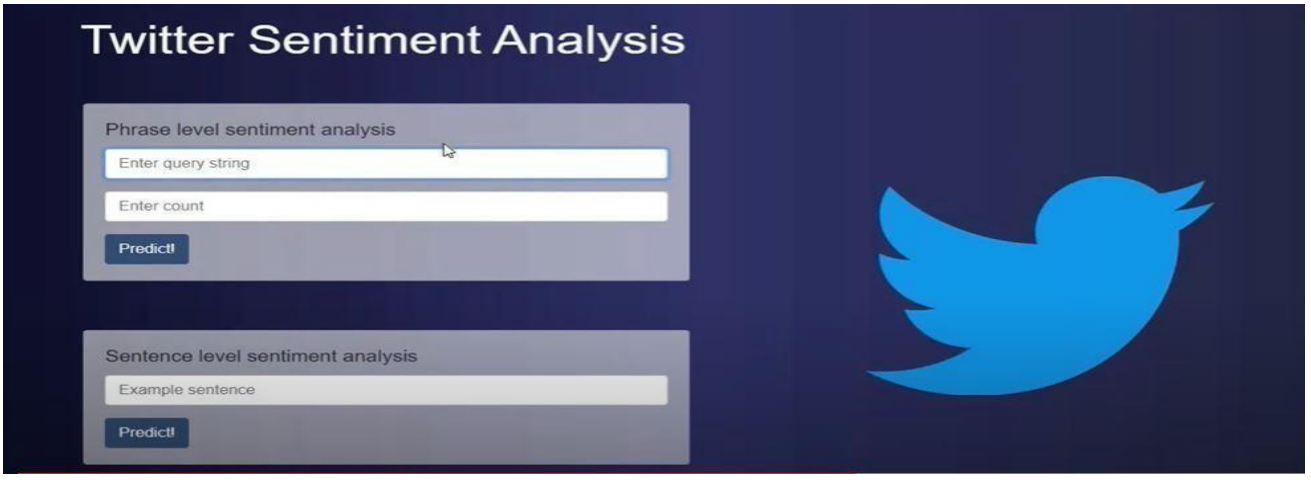


Figure 8.4: Output 1

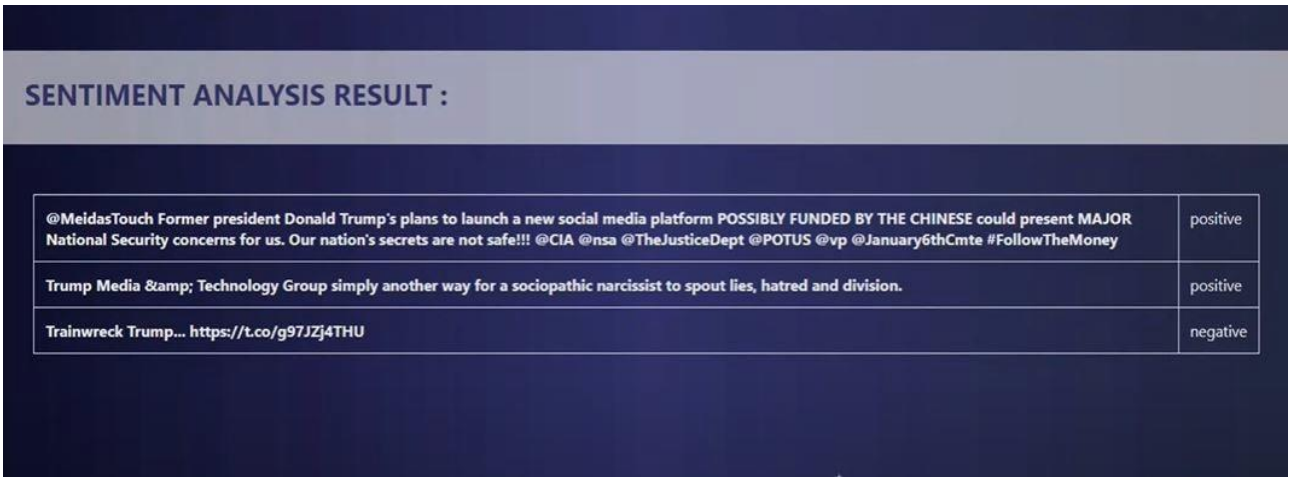


Figure 8.5: Output 2

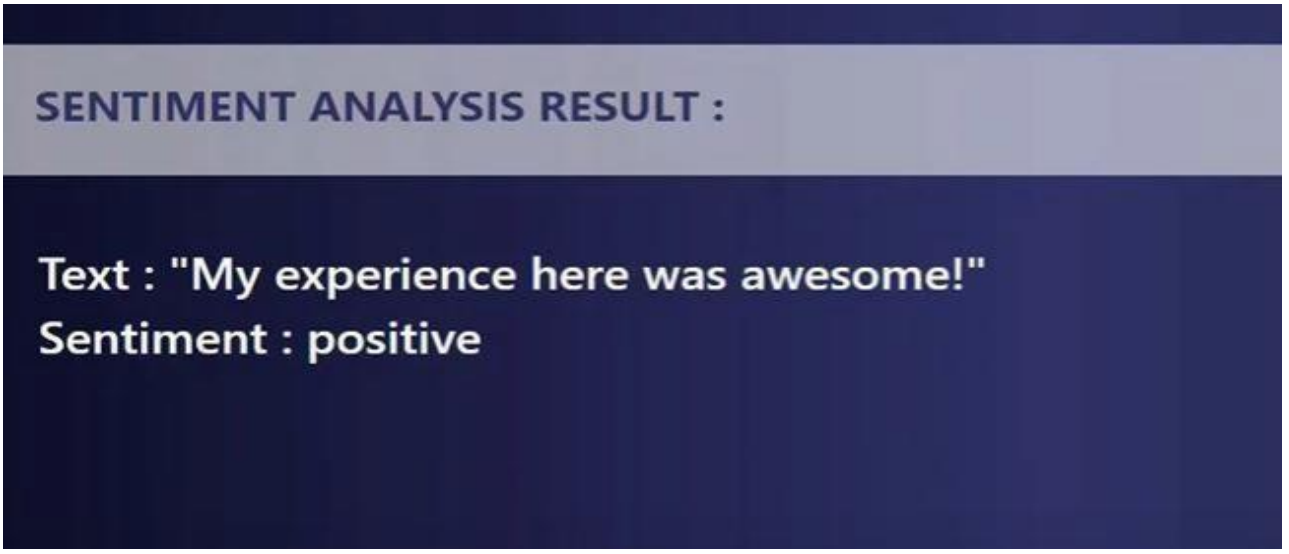


Figure 8.6: Output 3

Chapter 9

Future Scope

The future of sentiment analysis is going to continue to dig deeper, far past the surface of the number of likes, comments and shares, and aim to reach, and truly understand, the significance of social media interactions and what they tell us about the consumers behind the screens.

Sentiment analysis can help companies understand how customers feel about a brand: positive, negative, or neutral. Brand monitoring, including sentiment analysis, is one of the most important ways to keep customers engaged and interested.

Sentiment analysis has been an important tool for brands looking to learn more about how their customers are thinking and feeling. It is a relatively simplistic form of analytics that helps brands find key areas of weakness (negative sentiments) and strengths (positive sentiments).

Sentiment analysis is a uniquely powerful tool for businesses that are looking to measure attitudes, feelings and emotions regarding their brand. To date, the majority of sentiment analysis projects have been conducted almost exclusively by companies and brands through the use of social media data, survey responses and other hubs of user-generated content. By investigating and analyzing customer sentiments, these brands are able to get an inside look at consumer behaviors and, ultimately, better serve their audiences with the products, services and experiences they offer.

References

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

2) S. Batra and D. Rao, ”Entity Based Sentiment Analysis on Twitter”, Stan- ford University,2010

3) Saif M.Mohammad and Xiaodan zhu ,Sentiment Analysis on of social media texts:,2014

4) Ekaterina kochmar,University of Cambridge,at the Cambridge coding Academy Data Science.2016

5) Manju Venugopalan and Deepa Gupta ,Exploring Sentiment Analysis on Twitter Data, IEEE 2015

6) Brett Duncan and Yanqing Zhang, Neural Networks for Sentiment Analysison Twitter.2017

7) Afroze Ibrahim Baqapuri, Twitter Sentiment Analysis: The Good the Bad and the OMG!,

8) Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media.2011

Acknowledgement

We would like take up this opportunity to express gratitude to everyone who supported us throughout the course of this project. We are thankful for their aspiring guidance, invaluably constructive criticism and friendly advice during the project work. We are highly indebted to our project guide Prof. Merrin Solomon for her guidance and constant supervision as well as for providing necessary information regarding the project also for her support in working towards the project. We also express our sincere gratitude to our friends and family who provided us with the facilities being required and conductive conditions for our project and being a part of the Mini Project.

Date:

Shruti Rajesh Rahate (2015057)
Komal Bharat Rote (2015060)