**A Project Report**

**On**

***SENTIMENT ANALYSIS ON TWITTER DATA***

Submitted by:

DHRUVAJYOTI BORDOLOI

Under the guidance of

**Mr. Prabhat Das**

(PGT-IP)



**ARMY PUBLIC SCHOOL JORHAT**

**MES GATE ROWRIAH JORHAT**

**Assam-785004**

**CERTIFICATE BY PRINCIPAL**

This is to certify that the project report entitled “**Sentiment Analysis On Twitter Data**”

submitted to Army Public School, Jorhat by Dhruvajyoti Bordoloi has been successfully completed and examined.

This project report is according to the CBSE guidelines

Date-

Place-

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Mrs. Ferdausi Sultana Hazarika

(Principal of APS, Jorhat)

**CERTIFICATE BY EXAMINERS**

This is to certify that this project report entitled “**Sentiment Analysis on Twitter Data**” is a bonafide work done by Dhruvajyoti Bordoloi who carried out the project work under my supervision.

It is certified further that, to the best of my knowledge the work reported here has not been submitted to any other institute for award of any other degree.

Date-

Place-

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ External Examiner Internal Examiner

(Mr. Prabhat Das)

**ACKNOWLEDGEMENT**

I am extremely thankful to Mr. Prabhat Das (PGT-IP) for his valuable guidance and support without his help our project might not be completed.

I would like to extend my gratitude to Army Public School, Jorhat for giving me the opportunity.

**DECLARATION BY CANDIDATE**

I hereby declare that the project report entitled “**Sentiment Analysis On Twitter Data**” submitted by me to Army Public School, Jorhat was not submitted to any other institution for award of any other degree.

I also declare that the sources used in this project report have been fully acknowledged.

Date-

Place**-**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

Signature of Candidate

**ABSTRACT**

Sentiment analysis also referred as opinion mining is the computational study of people’s opinions, sentiments, evaluations, attitudes and emotions expressed in textual form. It is one of the most active research areas in natural language processing (NLP) and text mining in recent years.

It is popular mainly due to two reasons. First, it has wide range of applications because opinions are central to almost all human activities and are key influencers of our behaviours. Before making a decision we want to hear others opinions.

Second, it presents many challenging research problems, which have never been attempted before the year 2000. Part of the reason for the lack of study before was because of the little opinionated texts in digital forms. In fact this research has spread outside of computer science to the management sciences and social sciences due to its importance to business and society as a whole.

**CONTENTS**

**Pg. No**

1. Tools and libraries used 1-2
2. Introduction 3
3. Project Overview 4-7
4. Source code 8-10
5. Commands used in MySQL 11
6. Conclusion and future work 12
7. References 13

**LIST OF FIGURES**

**Pg. No**

Fig 1. User information table 4

Fig 2. User authentication 4

Fig 3. Output options 5

Fig 4. Output for option 1 5

Fig 5. Output for option 2 5

Fig 6. Output for option 3 6

Fig 7. Output for option 4 6

Fig 8. Output for option 5 7

Fig 9. Output for option 6 7

**Tools and Libraries Used**

**MySQL**

MySQL is an open source relational database management system developed by Oracle. It is under the terms of the GNU General Public License and is also available under a variety of proprietary licenses. Its name is a combination of ‘My’, the name of co-founder Michael Widenius’s daughter and SQL is the abbreviation of structured query language (SQL). SQL is used by programmers to create, modify and extract data from the relational database as well as control user access to the database.[1]

**Pandas**

Pandas is a python library used for working with data sets. It allows us to analyze big data and make conclusions based on statistical theories. The name is derived from the term ‘panel data’ an econometrics term for data sets that include observations over multiple time periods for the same individuals.[2]

**PyCharm**

PyCharm is a Python Integrated Development Environment (IDE) providing a wide range of essential tools for python developers. It is developed by the Czech company JetBrains. It provides code analysis, a graphical debugger, an integrated unit tester, integration with version control systems (VCSes), and supports web development with Django as well as data science with Anaconda.

The community edition of PyCharm is released under the Apache License, and there is also Professional Edition with extra features- released under a proprietary license.[3]

**Matplotlib**

Matplotlib is a plotting library for the Python programming language and its numerical mathematics extension NumPy. It was introduced by John Hunter in the year 2002.

Matplotlib comes with a variety of plots which helps to understand trends, patterns and to make correlations. It presents a complex data in form of simple graphs.[4]

**TextBlob**

TextBlob is a Python library for processing textual data. It actively uses Natural Language ToolKit (NLTK) to achieve its tasks. TextBlob provides a consistent API for diving into common Natural Language Processing (NLP) tasks such as part-of-speech tagging, noun phrase extraction, sentiment analysis and more.[5]

**Smtplib**

Python comes with the built in module Smtplib module for sending emails using the Simple Mail Transfer Protocol (SMTP).Smtplib uses the RFC 821 protocol for SMTP.

**INTRODUCTION**

Sentiment analysis refers to the use of natural language processing, text analysis and computational linguistics to identify, extract, quantify and study subjective information. Sentiment analysis determine and categorize the opinions and attitude of people expressed in a part of the text. It is the process of detecting a piece of writing for positive negative or neutral feelings bound to it.

The program is developed to analyse the sentiments expressed in a given dataset and assign its polarity (Positive, negative, neutral) as its not realistic to have people individually read tens of thousands of reviews. This program includes simple registration system where the user can register him/her into the program with the help of an one time password (OTP) sent to the email address used for the registration process. The program then gives us the number of positive, negative or neutral statements present in the given dataset. The program uses the python library TextBlob for this purpose. A graph is also generated using the Matplotlib library that gives us a visual representation of the polarity of the given data.

**Project Overview**



Figure 1: All the information regarding the users is stored in this table

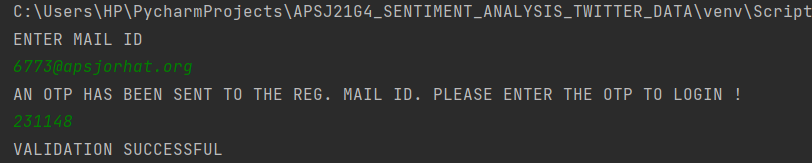


Figure 2 : When the home page is opened program asks for user’s email address, on entering an email id, program sends an OTP configuration to user’s email.

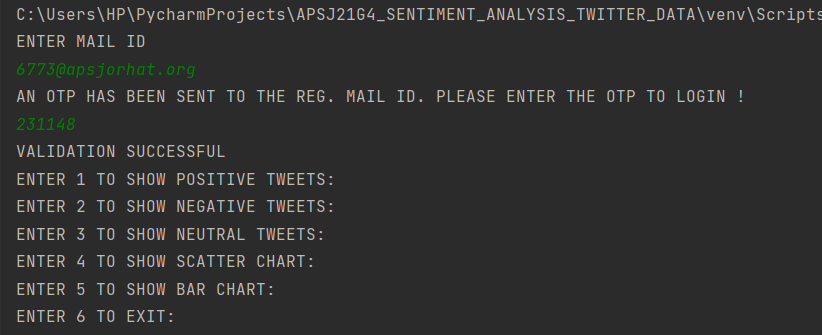


Figure 3: After the login procedure is completed the user is presented with the following

list of options

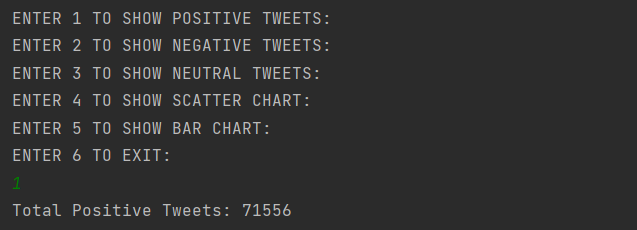


Figure 4: Output for option 1 (Total positive Tweets)

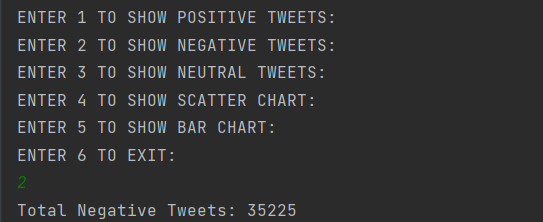


Figure 5: Output for option 2 (Total negative Tweets)

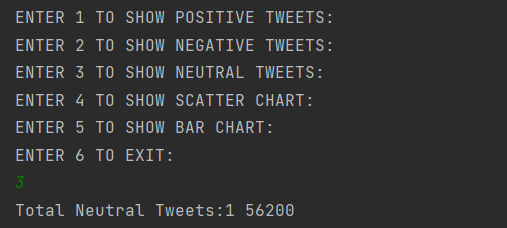


Figure 6: Output for option 3 (Total neutral Tweets)

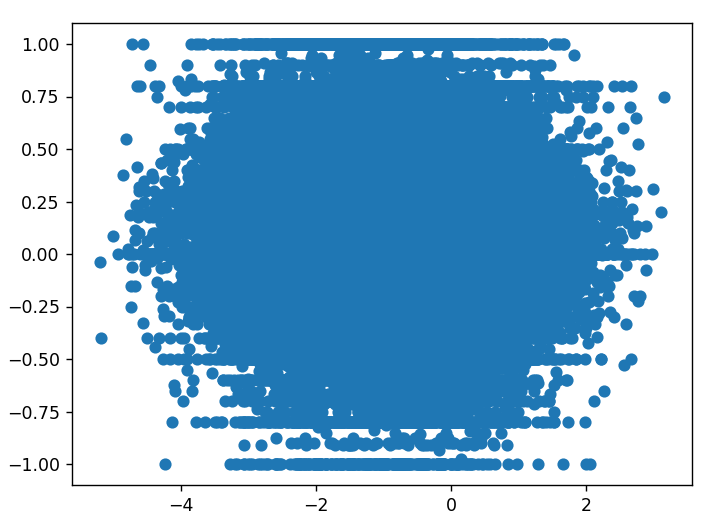


Figure 7: Output for option 4 (Scatter Chart)

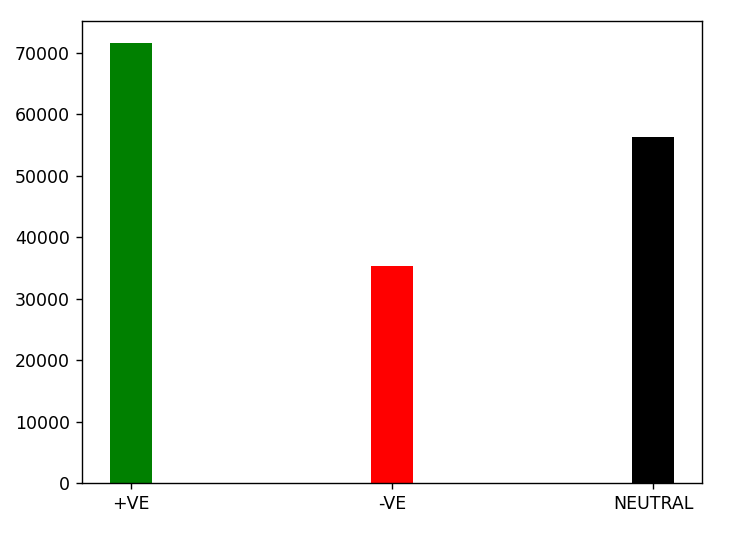


Figure 8: Output for option 5 (Bar Graph)

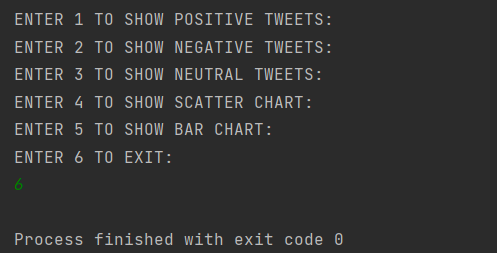


Figure 9: Output for option 6 (Exit)

**Source Code**

**main**

import auth  
import otp\_sender  
import dashboard  
print("ENTER MAIL ID")  
email=input()  
if auth.auth\_user(email)==1:  
 rcv\_otp=otp\_sender.otp\_sender(email)  
 print("AN OTP HAS BEEN SENT TO THE REG. MAIL ID. PLEASE ENTER THE OTP TO LOGIN !")  
 inp\_otp=input()  
 if rcv\_otp == inp\_otp:  
 print("VALIDATION SUCCESSFUL")  
 dashboard.main\_sa()  
 else:  
 print("INVALID OTP")  
 exit()

**Authentication**

import connector as con  
def auth\_user(email):  
 count=0  
 query = "select \* from reg\_users;"  
 con.cursor.execute(query)  
 for i in con.cursor:  
 if i[0]==email:  
 count=1  
 return count

**Connector**

import mysql.connector as mc  
dbc = mc.connect(host=”localhost”,user=”root”,passwd=”root”,database=”twitter\_data”)  
cursor=dbc.cursor()

**OTP Sender**

import smtplib  
import random  
def otp\_sender(email):  
 otp=str(random.randint(100000,999999))  
 SUBJECT = 'OTP FOR LOGIN'  
 TEXT = 'YOUR OTP TO LOGIN IS:' + otp  
 s = smtplib.SMTP('smtp.gmail.com', 587)  
 s.starttls()  
 s.login('group4@apsjorhat.org', 'apsj#12345678')  
 message = 'Subject:{} \n\n{}'.format(SUBJECT, TEXT)  
 s.sendmail('group4@apsjorhat.org', email, message)  
 s.quit()  
 return otp

**Dashboard**

import textblob as tb  
import matplotlib.pyplot as plt  
import numpy as np  
import csv  
try:  
 def main\_sa():  
 delimiters = ["[", "'", "]", "(", ")"]  
 pos = 0  
 neg = 0  
 neu = 0  
 y = []  
 print("ENTER 1 TO SHOW POSITIVE TWEETS:\r\nENTER 2 TO SHOW NEGATIVE TWEETS:\r\nENTER 3 TO SHOW NEUTRAL TWEETS:\r\nENTER 4 TO SHOW SCATTER CHART:\r\nENTER 5 TO SHOW BAR CHART:\r\nENTER 6 TO EXIT:")  
 a = int(input())  
  
 with open('Twitter\_Data.csv', 'r',errors='ignore') as file:  
 reader = csv.reader(file)  
 for row in reader:  
 data = row  
 string\_data = str(data)  
  
 for i in delimiters:  
 string\_data = string\_data.replace(i, '')  
 input\_to\_textblob = tb.TextBlob(string\_data)  
 sentence\_polarity = input\_to\_textblob.sentiment.polarity  
  
 if (sentence\_polarity > 0):  
 y.append(sentence\_polarity)  
 pos += 1  
 elif (sentence\_polarity == 0):  
 y.append(sentence\_polarity)  
 neu += 1  
 elif (sentence\_polarity < 0):  
 y.append(sentence\_polarity)  
 neg += 1  
  
 if a == 1:  
 print("Total Positive Tweets:", pos)  
 main\_sa()  
 elif a == 2:  
 print("Total Negative Tweets:", neg)  
 main\_sa()  
 elif a == 3:  
 print("Total Neutral Tweets:1", neu)  
 main\_sa()  
 elif a == 4:  
 x = np.random.normal(min(y), max(y), len(y))  
 plt.scatter(x, y)  
 plt.savefig("scatter\_sentiment\_analysis.pdf")  
 plt.show()  
 main\_sa()  
 elif a == 5:  
 x = [5, 10, 15]  
 y = [pos, neg, neu]  
 plt.bar(x, y, color=['g', 'r', 'k'])  
 plt.xticks(x, ['+VE', '-VE', 'NEUTRAL'])  
 plt.savefig("bar\_sentiment\_analysis.pdf")  
 plt.show()  
 main\_sa()  
 elif a == 6:  
 exit()  
except Exception as e:  
 print(e)

**COMMANDS USED IN MySQL**

**Creating database**

Create database twitter\_data;

**Using database**

Use twitter\_data;

**Creating table and inserting values**

Create table reg\_users(email varchar(30));

**To fetch all values**

Select \* from reg\_users;

**Conclusion and Future Work**

In this project we have created a program that can be used to analyse the sentiments of people in twitter. The analysed statements are then assigned a polarity that describes whether the statement is positive, negative or neutral. Most of the world’s data is unstructured and unsorted and a program like ours can help to sort the data in an efficient manner and generate useful information in the form of the sentiments expressed by people. Sentiment analysis is extremely important because it helps businesses quickly understand the overall opinions of their customers and it helps to make faster and more accurate decisions on a particular topic by automatically sorting the data behind reviews. The field of sentiment analysis is an exciting new research direction. What we have developed here is an elementary sentiment analysis program.

There is a lot of scope for upgrading our program, we can use advanced sentiment analysis models that are more accurate with better semantic knowledge. We can also add the functionality to link our program with more social media sites for sentiment analysis of social media posts. There is also the scope to develop a web-based application for sentiment analysis.

**REFERENCES**

[1]WIKIPEDIA, <https://www.en.wikipedia.org/wiki/MySQL>, Accessed on 5-3-2022

[2] W3Schools, <https://www.w3schools.com/Pandas>, Accessed on 5-3-2022

[3]WIKIPEDIA, <https://www.en.m.wikipedia.org/PyCharm>, Accessed on 5-3-2022

[4]GEEKSFORGEEKS, <https://www.geeksforgeeks.org/Matplotlib>, Accessed on 6-3-2022

[5]ANALYTICSVIDHYA, <https://www.analyticsvidhya.com/TextBlob>, Accessed on 6-3-2022

[6]ZETCODE, <https://zetcode.com/Smtplib>, Accessed on 5-3-2022