

TWITTER SENTIMENT ANALYSIS

Gungone Satyasree , Jonnada Keerthana ,Kaja Sai Samanwitha,Asma Begum

Department of Computer Science and Engineering, Stanley College of Engineering and Technology for Women, Telangana, India

Abstract: With the advancement of web technology and its growth, there is a huge volume of data present in the web for internet users and a lot of data is generated too. Internet has become a platform for online learning, exchanging ideas and sharing opinions. Social networking sites like Twitter, Facebook, Google+ are rapidly gaining popularity as they allow people to share and express their views about topics, have discussion with different communities, or post messages across the world. There has been lot of work in the field of sentiment analysis of twitter data. This survey focuses mainly on sentiment analysis of twitter data which is helpful to analyze the information in the tweets where opinions are highly unstructured, heterogeneous and are either positive or negative, or neutral in some cases. In this paper, we provide a survey and a comparative analyses of existing techniques for opinion mining like machine learning and lexicon-based approaches, together with evaluation metrics. Using various machine learning algorithms like Naive Bayes, Max Entropy, and Support Vector Machine, we provide research on twitter data streams. We have also discussed general challenges and applications of Sentiment Analysis on Twitter.

Keywords : Twitter, Sentiment analysis (SA), Opinion mining, Machine learning, Naive Bayes (NB), Maximum Entropy, Support Vector Machine (SVM).

1 INTRODUCTION

1.1 ABOUT THE PROJECT

This project of analyzing sentiments of tweets comes under the domain of “Pattern Classification” and “Data Mining”. Both of these terms are very closely related and intertwined, and they can be formally defined as the process of discovering “useful” patterns in large set of data, either automatically (unsupervised) or semi automatically (supervised). The project would heavily rely on techniques of “Natural Language Processing” in extracting significant patterns and features from the large data set of tweets and on “Machine Learning” techniques for accurately classifying individual unlabeled data samples (tweets) according to whichever pattern model best describes them. This project will be helpful to the companies, political parties as well as to the common people. It will be helpful to political party for reviewing the program that they are going to do or the program that they have performed. Similarly companies also can get review about their new product on newly released hardware or software’s. Also the movie maker can take review on the currently running movie. By analyzing the tweets analyzer can get results on how positive or negative or neutral people are about it.

1.2 OBJECTIVE OF THE PROJECT

1.2.1 Primary objective

The primary objectives of the project are mentioned below: To fulfill the requirement for achieving the Bachelor’s degree of Computer Science and Engineering To know the fundamentals of the Initially designed to perform domain of pattern classification and data mining Little by little Python is becoming a language that allows us to do everything.

1.2.2 Secondary objective

The secondary objectives of this project are mentioned below:

To develop an application that deals with the day-to-day requirement of any production organization.

1.3 SCOPE OF THE PROJECT

This project will be helpful to the companies, political parties as well as to the common people. It will be helpful to political party for reviewing the program that they are going to do or the program that they have performed. Similarly companies also can get

review about their new product on newly released hardware or softwares. Also the movie maker can take review on the currently running movie. By analyzing the tweets analyzer can get results on how positive or negative or neutral people are about it.

ADVANTAGES

Understanding people's sentiments
make wiser decisions
providing better products/services,
ultimately to improve the citizen lifestyle
Flexibility

1.5 DISADVANTAGES

ambiguity of natural language
characteristics of the posted content.
creating difficulties in determining the expressed sentiment
require large datasets
require large datasets

1.6 APPLICATIONS

social media monitoring,
customer support management, and
analyzing customer feedback.

1.7 HARDWARE AND SOFTWARE REQUIREMENTS

1.7.1 Software Requirements:

Operating System:
Windows 7/8/8.1/10
Microsoft Visio (2016)
Microsoft Word (2016)
Anaconda
Spyder

1.7.2 Hardware Requirements

Processor : Intel i5 or more
Motherboard : Intel® Chipset Motherboard.
Ram : 8GB or more
Cache : 512 KB
Hard disk : 16 GB hard disk recommended
Disk Drive : 1.44MB
Floppy Disk Drive Monitor : 1024 x 720 Display
Speed : 2.7GHZ and more

2.LITERATURE SURVEY

2.1 PROPOSED SYSTEM

To overcome the drawbacks of the methods we have Reviewed above, we propose a new model for sentiment Analysis. In this model we combine many techniques to Reach our final goal of emotion extraction. The steps for the Process are documented below.

Retrieval of Data: Public Twitter data is mined using the Existing Twitter APIs for data extraction. Tweets would Be selected based on a few chosen keywords pertaining To the domain of our concern, i.e. product reviews. We Have elected to use the Twitter API due to ease of data Extraction.

Preprocessing: In this stage, the data is put through a Preprocessing stage in which we remove identifying Information such as Twitter handles, timestamps of the Message and embedded links and videos. Such Information is largely irrelevant and may cause false Results to be given by our system.

Tweet Correction: As tweets are written for human Perusal, they often contain

slang, misspellings and other Irrelevant data. Thus we correct the misspellings in the Sentences and look to replace the slang in the sentences With words from standard english that may roughly Relate to the slang in question. As slang itself can be Used to display a wide variety of sentiment, often with Greater emotional impact, this process is necessary so That slang words may be considered as part of the Emotion expressed.

Polarity detection: In this step we begin the second Phase of our proposed system, in which we try to Identify the polarity of the sentence in question. If Emoticons exist in the statements, they will be used as Well to compute the overall polarity of the statement. We aim to find sentences where the polarity detection is Not very clear or where the expressed sentiment may be Low. We also try to isolate the opinion words in the Sentence in relation to a given concept in the sentence.

We train the system to understand the relation Between words in various contexts. Pre-existing Dictionaries like SenticNet can be used in this Phase to segregate the emotion from the context It is in.

Once the opinion words are identified with Context, we can find the polarities of the words Using NLTK-SentiWordNet.

To help with detection of the concepts associated, We train our system on a large dataset that Expresses a wide variety of complex and Ambiguous emotions. The system is given this Data in an unsupervised fashion and will proceed By clustering.

2.2 EXISTING SYSTEM

The polarity values range from -1 to +1. polarity determines the emotion of the statement.

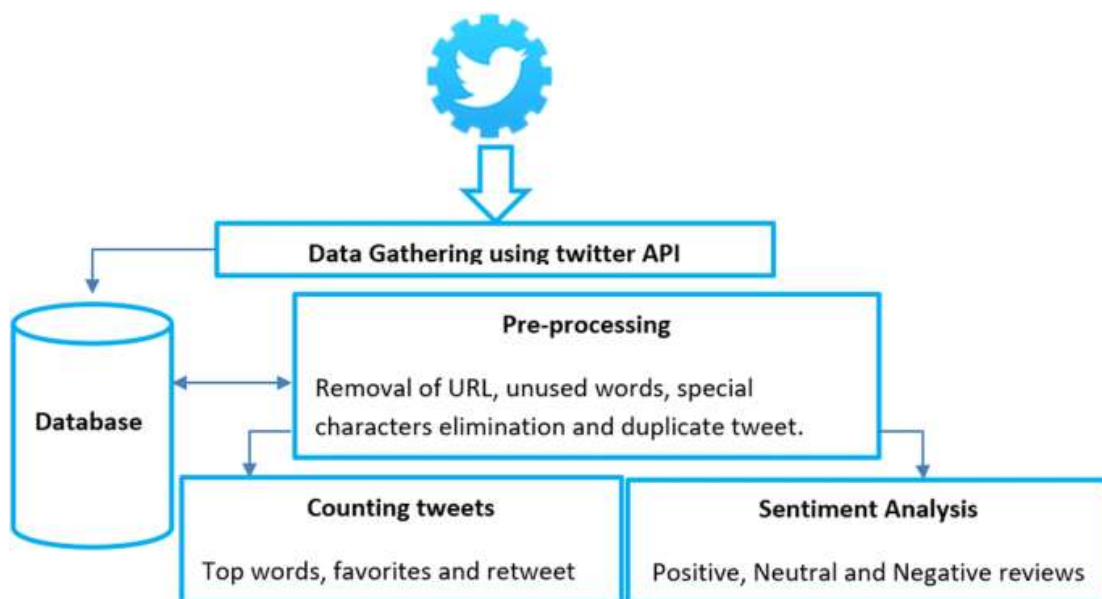
- The emotion is classified as positive, negative, neutral.
- If polarity > 0 then the emotion is Positive.
- If polarity = 0 then the emotion is Neutral.
- If polarity < 0 then the emotion is Negative.

Sentiment analysis is a hotspot in numerous research fields including natural language processing (NLP), data mining (DM), and information retrieval (IR).

4 IMPLEMENTATION

4.1 ALGORITHM

- The 12 sentiment analysis algorithms can be broken down into four categories:
- Use of sentiment lexicons
 - Off-the-shelf sentiment analysis systems including Amazon Comprehend, Google Cloud Services, and the Stanford CoreNLP system
 - Classical machine learning algorithms
 - Deep learning algorithms



4.2 CODE IMPLEMENTATION

```
import streamlit as st
import tweepy from textblob
import TextBlob from wordcloud
import WordCloud
import pandas as pd
import numpy as np
import re
import matplotlib.pyplot as plt
from PIL import Image
import seaborn as sns
consumerKey = '7Uo3HMQ3XtVb2vQn14DR7aIJz'
consumerSecret = 'xp9wjNSZNMDX9JZgw7MVDU8UtVISZawBBIFyN68SzhR5DtFIBx'
accessToken = '1276183506272391169-cuOzhfQ3rrJndadaW7IhM1Zm5hnhii'
accessTokenSecret = 'jrEvLfQgvD2Ah3LwGWQ2FzQfgYKzPi05T22yOUHFpQ3Rn'
#Create the authentication object
authenticate = tweepy.OAuthHandler(consumerKey, consumerSecret)
# Set the access token and access token secret
authenticate.set_access_token(accessToken, accessTokenSecret)
# Creating the API object while passing in auth information
api = tweepy.API(authenticate, wait_on_rate_limit = True)

#plt.style.use('fivethirtyeight')
def app():
    st.title("Tweet Analyzer 🗨️")
    activities=["Tweet Analyzer", "Generate Twitter Data"]
    choice = st.sidebar.selectbox("Select Your Activity",activities)
    if choice=="Tweet Analyzer":
```

st.subheader("Analyze the tweets of your favourite Personalities") st.subheader("This tool performs the following tasks :") st.write("1. Fetches the 5 most recent tweets from the given twitter handel") st.write("2. Generates a Word Cloud") st.write("3. Performs Sentiment Analysis a displays it in form of a Bar Graph") raw_text = st.text_area("Enter the exact twitter handle of the Personality (without @)") st.markdown("<----- Also Do checkout the another cool tool from the sidebar")

Analyzer_choice = st.selectbox("Select the Activities", ["Show Recent Tweets", "GenerateWordCloud", "Visualize the Sentiment Analysis"]) if st.button("Analyze"): if Analyzer_choice == "Show Recent Tweets":

st.success("Fetching last 5 Tweets") defShow_Recent_Tweets(raw_text): # Extract 100 tweets from the twitter user posts = api.user_timeline(screen_name=raw_text, count = 100, lang = "en",

tweet_mode="extended") defget_tweets(): l=[] i=1 for tweet in posts[:5]: l.append(tweet.full_text) i= i+1 return l recent_tweets=get_tweets() return recent_tweets recent_tweets= Show_Recent_Tweets(raw_text) st.write(recent_tweets) elifAnalyzer_choice=="Generate WordCloud":

st.success("Generating Word Cloud") defgen_wordcloud(): posts = api.user_timeline(screen_name=raw_text, count = 100, lang = "en", tweet_mode="extended") # Create a dataframe with a column called Tweets df = pd.DataFrame([tweet.full_text for tweet in posts], columns=["Tweets"]) # word cloud visualization

allWords = ' '.join([twts for twts in df['Tweets']]) wordCloud = WordCloud(width=500, height=300, random_state=21, max_font_size=110).generate(allWords) plt.imshow(wordCloud, interpolation="bilinear") plt.axis('off') plt.savefig('WC.jpg') img= Image.open("WC.jpg") return img img=gen_wordcloud() st.image(img) else:

defPlot_AnalysisAnalysis st.success("Generating Visualisation for Sentiment Analysis") posts = api.user_timeline(screen_name=raw_text, count = 100, lang = "en", tweet_mode="extended") df = pd.DataFrame([tweet.full_text for tweet in posts], columns=["Tweets"]) # Create a function to clean the tweets defcleanTxt(text): text = re.sub('@[A-Za-z0-9]+', "", text) #Removing @mentions text = re.sub('#', "", text) # Removing '#' hash tag text = re.sub('RT[\s]+', "", text) # Removing RT

text = re.sub('https?:\s\S+', "", text) # Removing hyperlink return text # Clean the tweets df['Tweets'] = df['Tweets'].apply(cleanTxt) defgetSubjectivity(text): return TextBlob(text).sentiment.subjectivity # Create a function to get the polarity defgetPolarity(text):

return TextBlob(text).sentiment.polagetPolarity # Create two new columns 'Subjectivity' & 'Polarity' df['Subjectivity'] = df['Tweets'].apply(getSubjectivity) df['Polarity'] = df['Tweets'].apply(getPolarity) defgetAnalysis(score): if score < 0: return 'Negative' elif score == 0: return 'Neutral' else: return 'Positive' df['Analysis']=df['Polarity'].apply(getAnalysis) return df

```
df= Plot_Analysis() st.write(sns.countplot(x=df["Analysis"],data=df))
st.pyplot(use_container_width=True) else:
    st.subheader("This tool fetches the last 100 tweets from the twitter handel& Performs the
following tasks") st.write("1. Converts it into a DataFrame") st.write("2. Cleans the text")
st.write("3. Analyzes Subjectivity of tweets and adds an additional column for it") st.write("4.
Analyzes Polarity of tweets and adds an additional column for it") st.write("5. Analyzes
Sentiments of tweets and adds an additional column for isidebar user_name =
st.text_area("*Enter the exact twitter handle of the Personality (without @)*")
#st.markdown("<----- Also Do checkout the another cool tool from the sidebar")
defget_data(user_name):
    posts = api.user_timeline(screen_name=user_name, count = 100, lang ="en",
tweet_mode="extended") df = pd.DataFrame([tweet.full_text for tweet in posts],
columns=["Tweets"]) defcleanTxt(text):
    text = re.sub('@[A-Za-z0-9]+', "", text) #Removing @mentions text = re.sub('#', "", text) #
Removing '#' hash tag text = re.sub('RT[\s]+', "", text) # Removing RT

    text = re.sub('https?:\V\S+', "", text) # Removing hyperlink return text # Clean the tweets
df['Tweets'] = df['Tweets'].apply(cleanTxt) defgetSubjectivity(text):
    return TextBlob(text).sentiment.subjectivity # Create a function to get the polarity
defgetPolarity(text):
    return TextBlob(text).sentiment.pol'Positive' # Create two new columns
'Subjectivity' & 'Polarity' df['Subjectivity'] = df['Tweets'].apply(getSubjectivity)
df['Polarity'] = df['Tweets'].apply(getPolarity) defgetAnalysis(score): if score < 0: return
'Negative' elif score == 0: return 'Neutral' else:
    return 'Positive' df['Analysis'] = df['Polarity'].apply(getAnalysis) return df

if st.button("Show Data"): st.success("Fetching Last 100 Tweets") df=get_data(user_name)
st.write(df) st.subheader(' -----BLACKHOLE PEACE-----
:sunglasses:') if __name__ == "__main__": app()
```

RESULTS

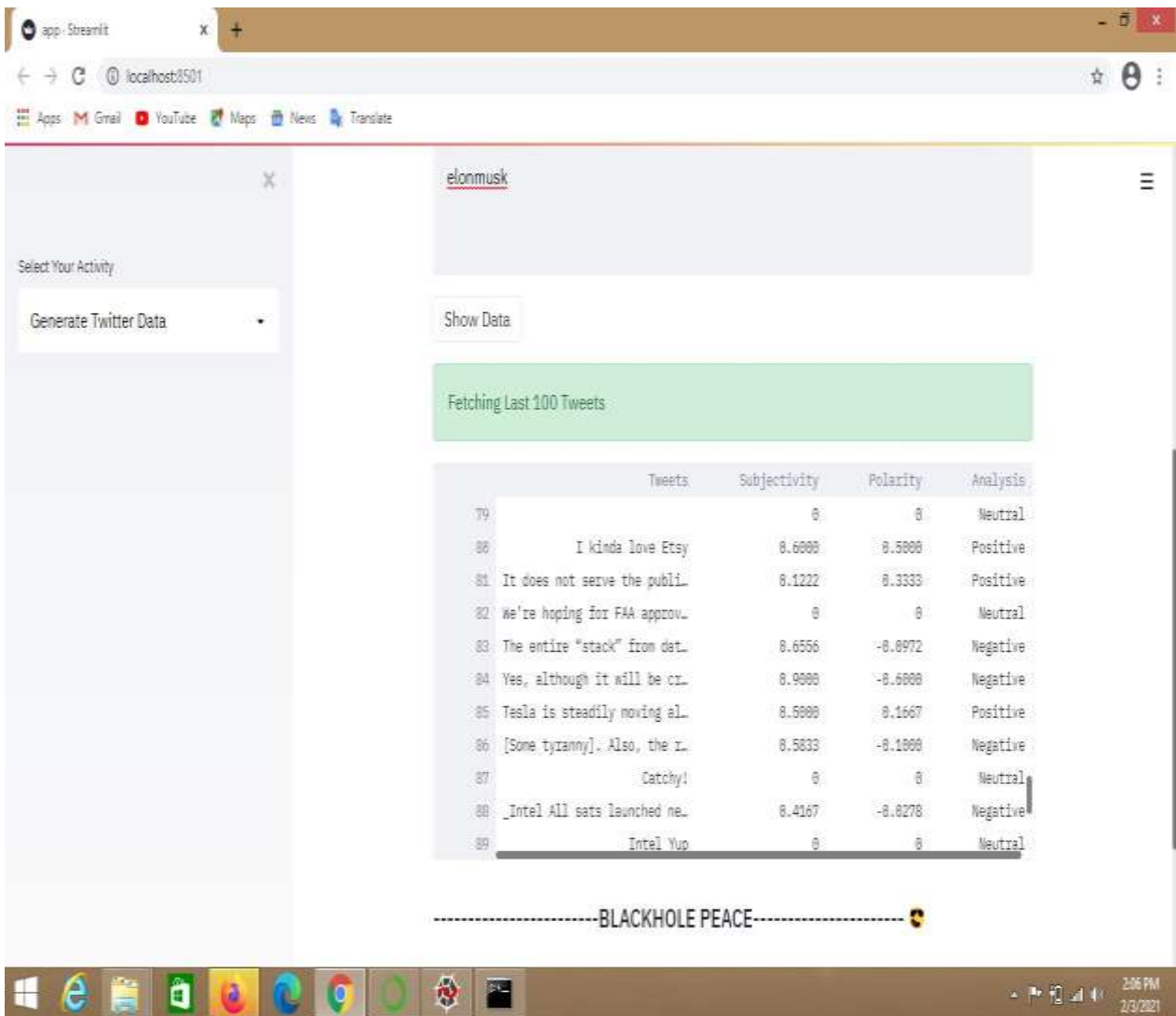
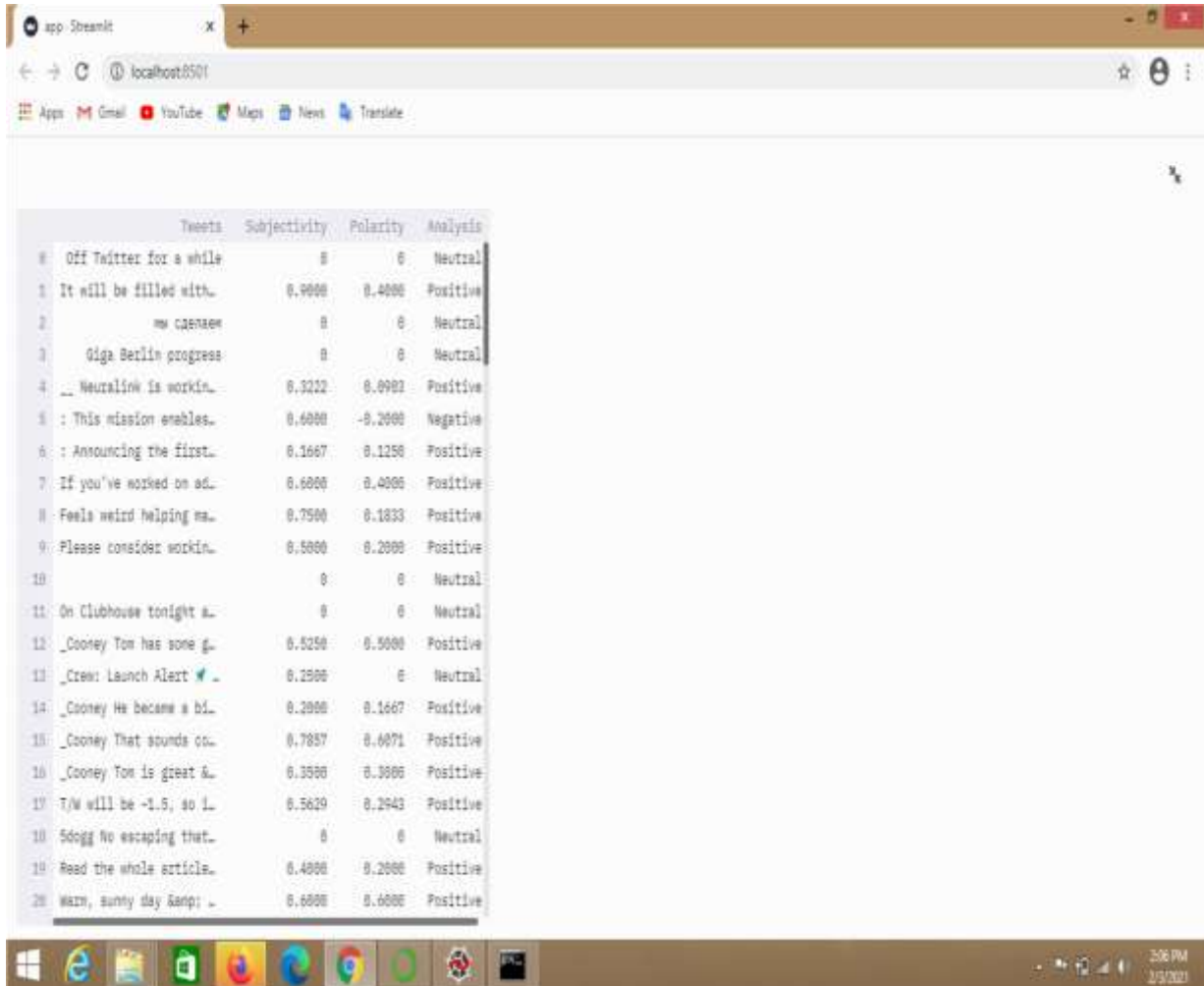


Photo6.1

Photo6.2



	Tweets	Subjectivity	Polarity	Analysis
0	Off Twitter for a while	0	0	Neutral
1	It will be filled with...	0.9000	0.4000	Positive
2	... my capsule	0	0	Neutral
3	Giga Berlin progress	0	0	Neutral
4	... Neuralink is workin...	0.3222	0.0902	Positive
5	: This mission enables...	0.6000	-0.2000	Negative
6	: Announcing the first...	0.1667	0.1250	Positive
7	If you've worked on ad...	0.6000	0.4000	Positive
8	Feels weird helping ma...	0.7500	0.1833	Positive
9	Please consider workin...	0.5000	0.2000	Positive
10		0	0	Neutral
11	On Clubhouse tonight a...	0	0	Neutral
12	_Cooney Tom has some g...	0.5250	0.5000	Positive
13	_Crew: Launch Alert 🚀	0.2500	0	Neutral
14	_Cooney He became a bi...	0.2000	0.1667	Positive
15	_Cooney That sounds co...	0.7087	0.6071	Positive
16	_Cooney Tom is great &...	0.3000	0.3000	Positive
17	T/W will be -1.5, so I...	0.5629	0.2943	Positive
18	Sdogg No escaping that...	0	0	Neutral
19	Read the whole article...	0.4000	0.2000	Positive
20	Warm, sunny day &rsq...	0.6000	0.6000	Positive

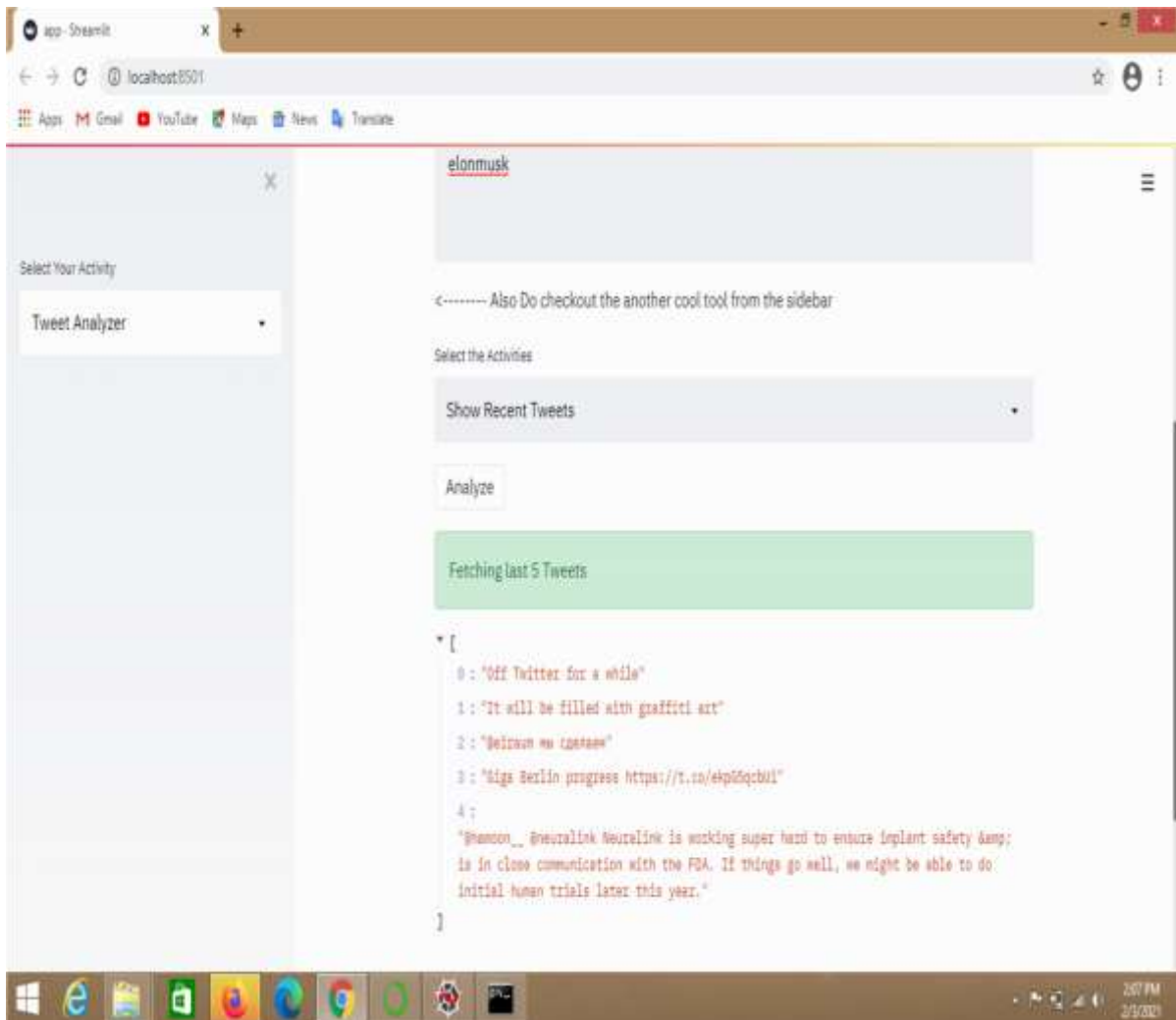


Photo 6.3

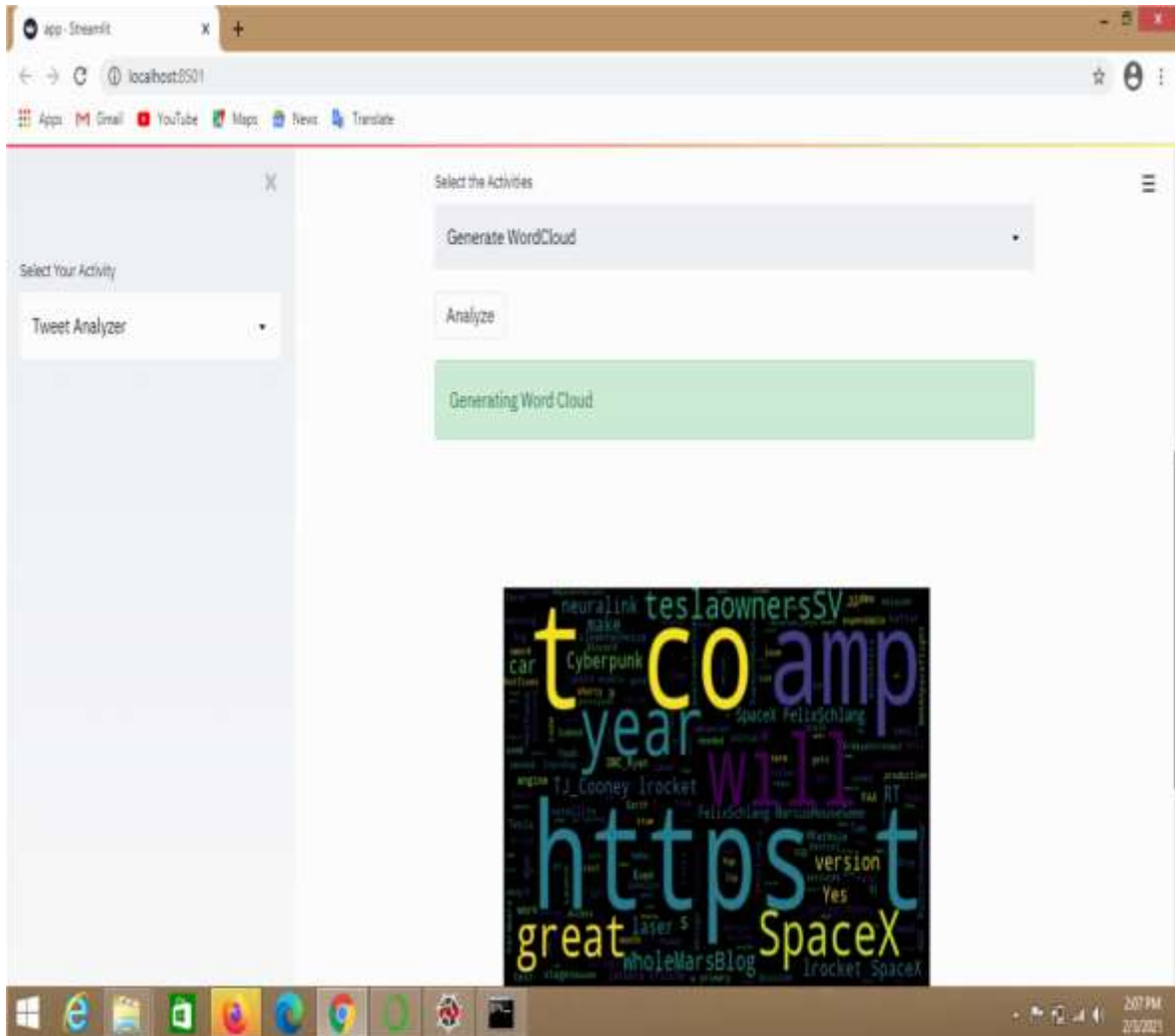


Photo 6.4

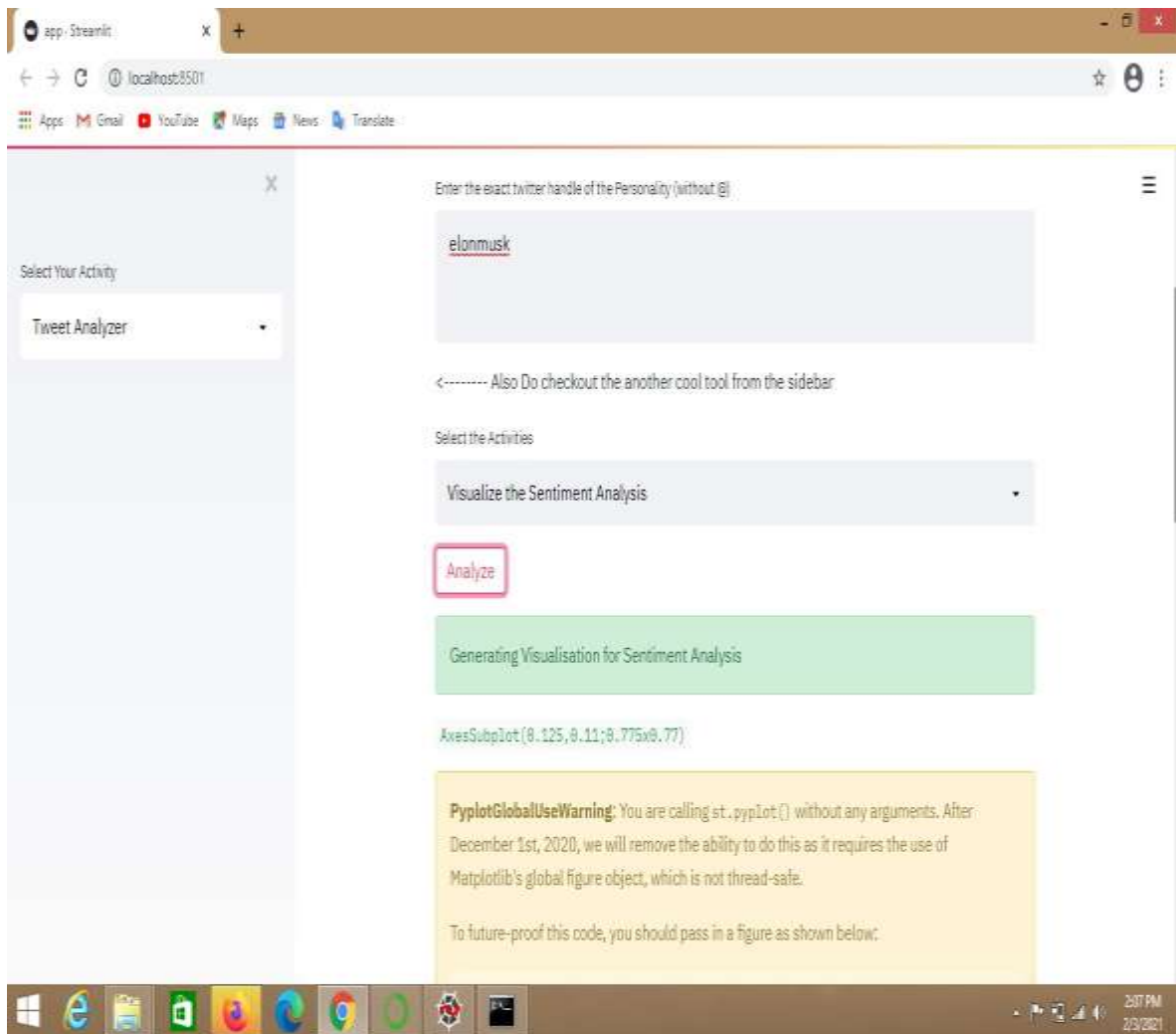


Photo 6.5

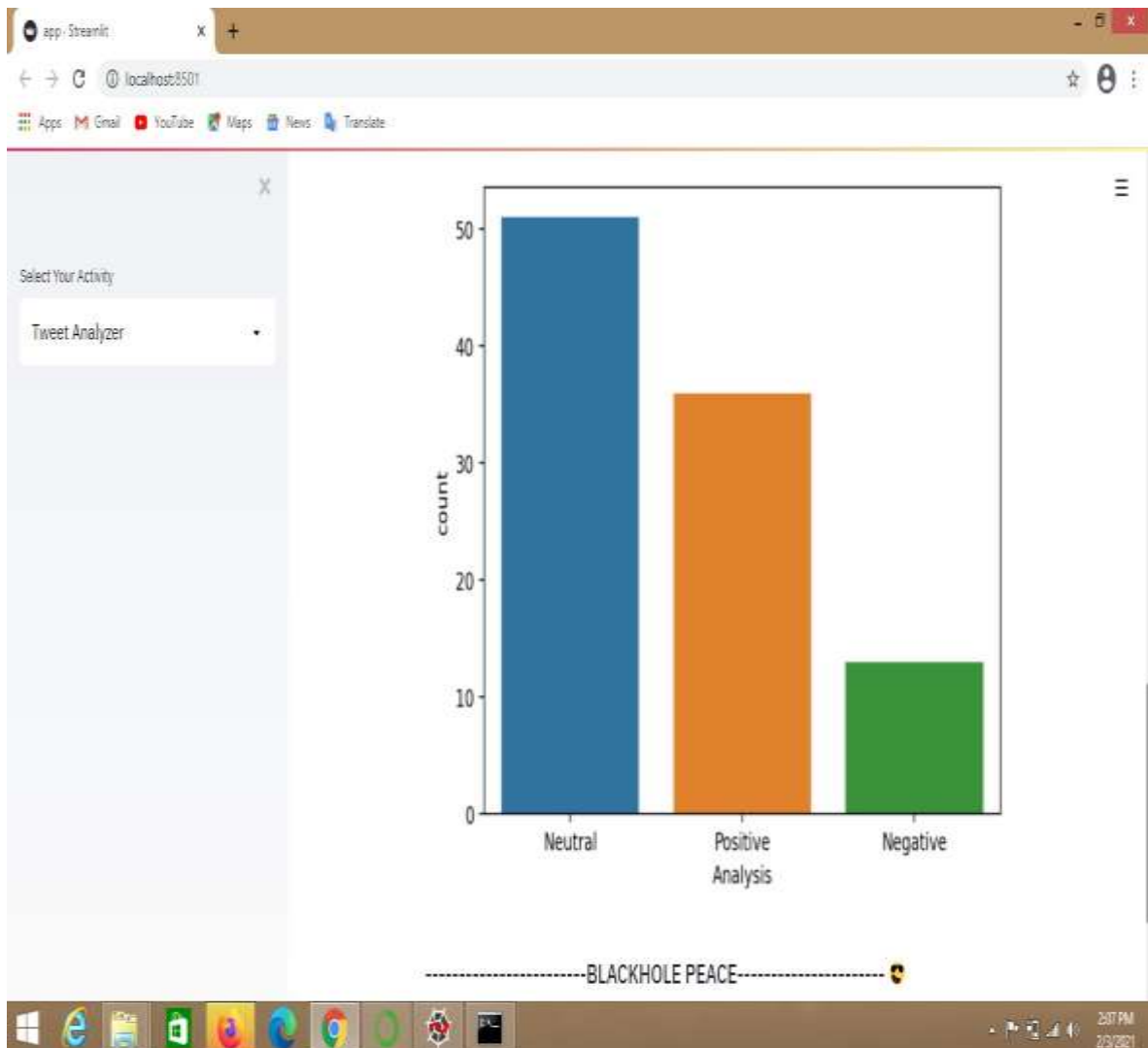


Photo 6.6

7 CONCLUSION

Twitter sentiment analysis comes under the category of text and opinion mining. It focuses on analyzing the sentiments of the tweets and feeding the data to a machine learning model to train it and then check its accuracy, so that we can use this model for future use according to the results. It comprises of steps like data collection, text preprocessing, sentiment detection, sentiment classification, training and testing the model. This research topic has evolved during the last decade with models reaching the efficiency of almost 85%-90%. But it still lacks the dimension of diversity in the data. Along with this it has a lot of application issues with the slang used and the short forms of words. Many analyzers don't perform well when the number of classes are increased. Also, it's still not tested that how accurate the model will be for topics other than the one in consideration. Hence sentiment analysis has a very bright scope of development in future.

8 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 Political parties, Companies, Movies, Likes share and comments.

9 REFERENCES

1. Jansen,B.J.; Zhang,M.; Sobel,K.; and Chowdury,A. (2009), "Twitterpower: Tweets as electronic word of mouth", Journal of the American Society for Information Science and Technology 60(11):2169–2188.

2. Pak, A., and Paroubek, P (2010), "Twitter as a corpus for sentiment analysis and opinion mining". In Proc. of LREC.
3. Pang, B., and Lee, L. (2008), "Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval" 2(1-2):1– 135.
4. Wilson, T. Wiebe, J.; and Hoffmann, (P. 2009),"Recognizing contextual polarity: An exploration of features for phrase-level sentiment analysis. Computational Linguistics", 35(3):399–433.
5. M Hu and B Liu. (2004),"Mining and summarizing customer reviews. KDD".
6. L. Barbosa, J. Feng. "Robust Sentiment Detection on Twitter from Biased and Noisy Data". COLING 2010: Poster Volume, pp. 36-44.
7. J. Kamps, M. Marx, R. J. Mokken, and M. De Rijke, "Using wordnet to measure semantic orientations of adjectives," 2004.
8. Alec Go, Richa Bhayani, and Lei Huang. 2009. Twitter sentiment classification using distant supervision. Technical report, Stanford.
9. David Zimbra, M. Ghiassi and Sean Lee, "Brand-Related Twitter Sentiment Analysis using Feature Engineering and the Dynamic Architecture for Artificial Neural Networks", IEEE 1530-1605, 2016.
10. Varsha Sahayak, Vijaya Shete and Apashabi Pathan, "Sentiment Analysis on Twitter Data", (IJIRAE) ISSN: 2349-2163, January 2015.
11. Peiman Barnaghi, John G. Breslin and Parsa Ghaffari, "Opinion Mining and Sentiment Polarity on Twitter and Correlation between Events and Sentiment", 2016 IEEE Second International Conference on Big Data Computing Service and Applications.
12. Mondher Bouazizi and Tomoaki Ohtsuki, "Sentiment Analysis: from Binary to Multi-Class Classification", IEEE ICC 2016 SAC Social Networking, ISBN 978-1-4799-6664-6.
13. Nehal Mamgain, Ekta Mehta, Ankush Mittal and Gaurav Bhatt, "Sentiment Analysis of Top Colleges in India Using Twitter Data", (IEEE) ISBN -978-1-5090-0082-1, 2016.
14. [https://www.geeksforgeeks.org/twitter-sentiment-analysis-usingpython.](https://www.geeksforgeeks.org/twitter-sentiment-analysis-usingpython/) Twitter Sentimental analysis for Realdonaldtrump" Devaki P, Ilakiya, J, Indumathi, R and Arul Priya, M,
15. "Geospatially and literally analysing tweets", Journal of Advanced Research in Dynamical and Con