



e-ISSN: 2319-8753 | p-ISSN: 2320-6710

IJIRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 10, Issue 5, May 2021

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 7.512

9940 572 462

6381 907 438

ijirset@gmail.com

www.ijirset.com



Tweet Sentiment Analysis Using Data Visualization

Sarthak Ahir¹, Chinmay Belekar², Dharmik Bhanushali³, Mihir Bhanushali⁴, Dr. Sandhya Kadam⁵

Student, Department of Electronics and Telecommunication Engineering, K.J Somaiya Institute of Engineering and Information Technology, Mumbai University, Mumbai, India¹

Student, Department of Electronics and Telecommunication Engineering, K.J Somaiya Institute of Engineering and Information Technology, Mumbai University, Mumbai, India²

Student, Department of Electronics and Telecommunication Engineering, K.J Somaiya Institute of Engineering and Information Technology, Mumbai University, Mumbai, India³

Student, Department of Electronics and Telecommunication Engineering, K.J Somaiya Institute of Engineering and Information Technology, Mumbai University, Mumbai, India⁴

Associate Professor, Department of Electronics and Telecommunication Engineering, K.J Somaiya Institute of Engineering and Information Technology, Mumbai University, Mumbai, India⁵

ABSTRACT: This project addresses the problem of sentiment analysis in twitter; that is classifying tweets according to the sentiment expressed in them: positive, negative or neutral. Twitter is an online micro-blogging and social-networking platform which allows users to write short status updates of maximum length 140 characters. It is a rapidly expanding service with over 200 million registered users - out of which 100 million are active users and half of them log on twitter on a daily basis - generating nearly 250 million tweets per day. Due to this large amount of usage we hope to achieve a reflection of public sentiment by analysing the sentiments expressed in the tweets. Analysing the public sentiment is important for many applications such as firms trying to find out the response of their products in the market, predicting political elections and predicting socioeconomic phenomena like stock exchange. The aim of this project is to develop a functional classifier for accurate and automatic sentiment classification of an unknown tweet stream

KEYWORDS: Sentiment analysis, Twitter.

I. INTRODUCTION

We have chosen to work with twitter since we feel it is a better approximation of public sentiment as opposed to conventional internet articles and web blogs. The reason is that the amount of relevant data is much larger for twitter, as compared to traditional blogging sites. Moreover, the response on twitter is more prompt and also more general (since the number of users who tweet is substantially more than those who write web blogs on a daily basis). Sentiment analysis of public is highly critical in macro-scale socioeconomic phenomena like predicting the stock market rate of a particular firm. This could be done by analysing overall public sentiment towards that firm with respect to time and using economics tools for finding the correlation between public sentiment and the firm's stock market value. Firms can also estimate how well their product is responding in the market, which areas of the market is it having a favourable response and in which a negative response (since twitter allows us to download stream of geo-tagged tweets for particular locations. If firms can get this information, they can analyze the reasons behind geographically differentiated response, and so they can market their product in a more optimized manner by looking for appropriate solutions like creating suitable market segments. Predicting the results of popular political elections and polls is also an emerging application to sentiment analysis.



II. LITERATURE SURVEY

[1] Pak and Alexander et al. proposed, by using the corpus, Author builds a sentiment classifier, which is capable of determining positive, neutral and negative sentiments for the whole document. Experimental results show that the proposed techniques are more efficient and perform better as compared to previously proposed techniques

[2] Agarwal and Apoorv explained one such popular micro blog named as Twitter and build models to classifying the “tweets” into positive and negative sentiment or they can be neutral. Author build novel models for two classification: first one is a binary task of classifying sentiment of users into positive and negative classes and second is a 3-way task of classifying sentiment of users into positive, negative

Author experiment with two types of models: (1) unigram model which is a feature-based model (2) a tree 30° kernel-based model.

[3] Ortigosa and Alvaro et. al proposed a novel method for sentiment analysis in social site giant Facebook that, starting from the messages written by its users, supports: (i) to extract useful information about the Facebook users’ sentiment polarity (whether it is positive, neutral or negative), which reflected from the messages written by users; and (ii) to model the users’ normal sentiment polarity and to analyze significant emotional changes in user.

[4] Aisopos and Fotis et al. presented with Microblog content, some serious challenges are associated. Some of these are the applicability of sentiment analysis used in past and different classification methods caused by their inherent characteristics of content. To resolve them, author introduces a method that relies on two orthogonal and complementary sources of evidence: context-based method captured by polarity ratio and content-based features acquired by n-gram graphs. Both the methods are language-neutral and tolerant to noise; guarantee high robustness and effectiveness in the manner author are considering.

[5] Horakova and Marketa et al. Presents model which collects tweets from social networking sites and thus provide a view of business intelligence.

In their framework, there are two layers in the sentiment analysis tool, the data processing layer and sentiment analysis layer. Data processing layer deals with data collection and data mining.

[6] Arijit Chatterjee and Dr. William Perrizo discuss the effect investors’ bias has on the volatility of stocks in the market, sentiment analysis was done on tweets of the potential investors and also why they used Microsoft Azure over other sentiment analyzer tools. Twitter is one of the largest social media platforms with over 280 million active users with almost 500 million tweets created every day. Some investors use Twitter to share their opinion on some ticker symbols every day, this paper discusses how these opinions of the investors affect the stock market. Investors are assumed to be sentiment driven.

III. ARCHITECTURE

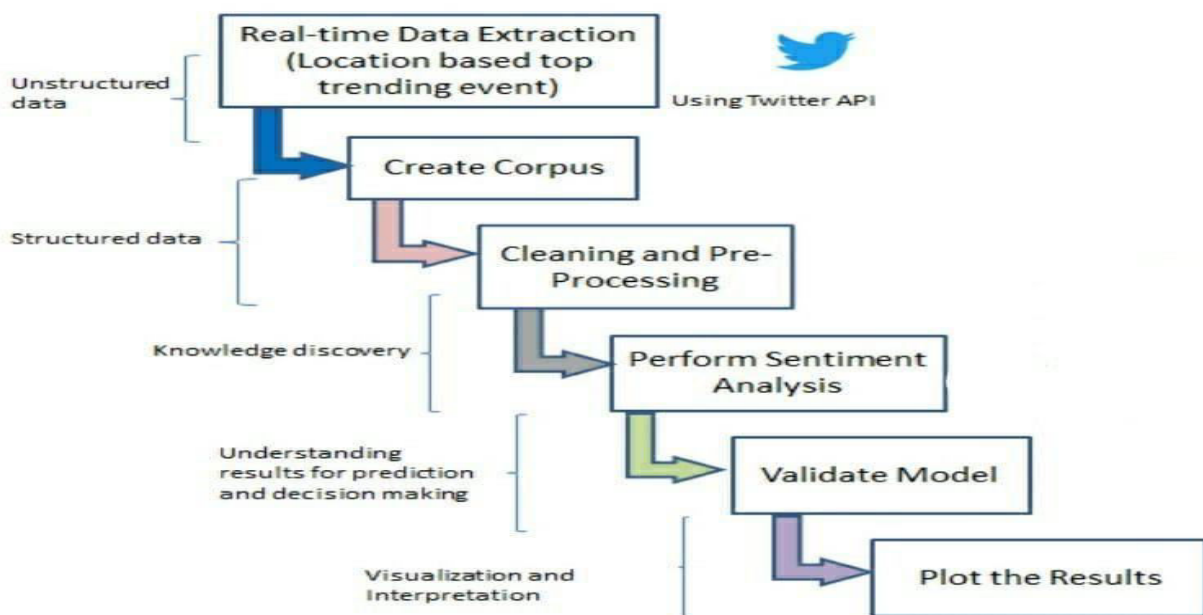


Fig 1: Architecture of the proposed system.



The process of designing a functional classifier for sentiment analysis can be broken down into three basic categories. They are as follows:

- I. Data Acquisition
- II. Feature Extraction
- III. Assigning the polarity

Data Acquisition:

Data in the form of raw tweets is acquired by using the python library “tweepstream” which provides a package for simple twitter streaming API. This API allows two modes of accessing tweets: SampleStream and FilterStream. SampleStream simply delivers a small, random sample of all the tweets streaming at a real time. FilterStream delivers tweet which match a certain criteria.

Feature Extraction:

Now that we have arrived at our training set we need to extract useful features from it which can be used in the process of classification. But first we will discuss some text formatting techniques which will aid us in feature extraction:

1. Tokenization: It is the process of breaking a stream of text up into words, symbols and other meaningful elements called “tokens”. Tokens can be separated by whitespace characters and/or punctuation characters. It is done so that we can look at tokens as individual components that make up a tweet.
 2. URL's and user references (identified by tokens “http” and “@”) are removed if we are interested in only analyzing the text of the tweet.
 3. Punctuation marks and digits/numerals may be removed if for example we wish to compare the tweet to a list of English words.
 4. Stop-words removal: Stop words are class of some extremely common words which hold no additional information when used in a text and are thus claimed to be useless. Examples include “a”, “an”, “the”, “he”, “she”, “by”, “on”, etc. It is sometimes convenient to remove these words because they hold no additional information since they are used almost equally in all classes of text, for example when computing prior-sentiment-polarity of words in a tweet according to their frequency of occurrence in different classes and using this polarity to calculate the average sentiment of the tweet over the set of words used in that tweet.
- Parts-of-Speech Tagging: POS-Tagging is the process of assigning a tag to each word in the sentence as to which grammatical part of speech that word belongs to, i.e. noun, verb, adjective, adverb, coordinating conjunction etc.

Assigning the polarity:

Design of dictionary: After the data acquisition and feature extraction, the broken stream of data passes through a bag of dictionary. This bag of dictionary consists of bag of words separated by two columns of negative words and positive words. The stream of data passes through the dictionary and polarity is assigned to each word of the stream by the dictionary. Thus depending on the polarity of the stream the sentiment is decided. If the polarity is greater than zero, the tweet is said to be positive, if the polarity is less than zero the tweet is said to be negative and if the polarity is zero the tweet is assigned as a neutral tweet. We labelled the tweets in three classes according to sentiments expressed/observed in the tweet: positive, negative, neutral/objective. We gave the following guidelines to our labellers to help them in the labelling process:

- **Positive:** If the entire tweet has a positive/happy/excited/joyful attitude or if something is mentioned with positive connotations. Also, if more than one sentiment is expressed in the tweet but the positive sentiment is more dominant. Example: “4 more years of being in dangerous Australia then I move to the USA! :D”.
- **Negative:** If the entire tweet has a negative/sad/displeased attitude or if something is mentioned with negative connotations. Also, if more than one sentiment is expressed in the tweet but the negative sentiment is more dominant. Example: “I want an android”.



IV. METHODOLOGY

Downloading Twitter Datasets using Tweepy: We will be extracting data from Twitter from the APIs obtained by the Twitter by using either a Twitter handle or by searching for a trending hashtag. To download the following Twitter datasets, we ought to use the Tweepy libraries.

Processing the downloaded data: The data received will be then be used for processing. Before processing, we will convert the unstructured data into structured data to process it further using tokenization.

Performing Sentiment Analysis: Further, this clean data will be fit to undergo the Sentiment analysis using various sentiment analysis technologies and prediction models to obtain the best results with accurate efficiency.

Creating an analytical Dashboard: For easy to access of data, a visual Dashboard will be made with unique visuals so that the individuals who want to access data can see their results and access it at their convenience.

Deploying the Web App: With a Visual Dashboard, there will be an addition of a Web Application, so that the individuals who want to access the data can do so at their own pace without any issues.

Now, for the display of the analysis, we have used a web application called Heroku. All of the extracted data in the form of tweets which is analysed is displayed on the web application.

V. RESULTS

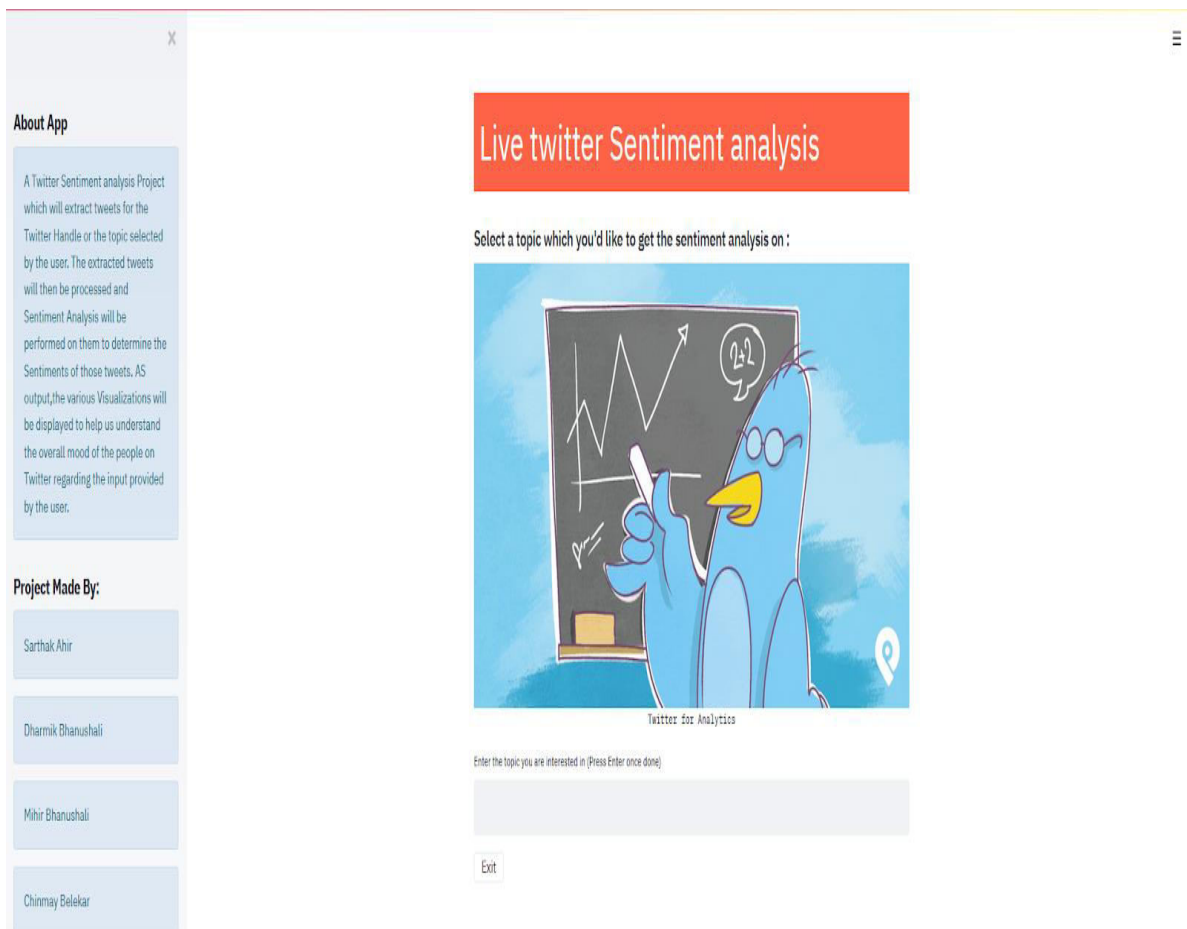


Fig 2: Homepage for the Webapp.



About App

A Twitter Sentiment analysis Project which will extract tweets for the Twitter Handle or the topic selected by the user. The extracted tweets will then be processed and Sentiment Analysis will be performed on them to determine the Sentiments of those tweets. AS output, the various Visualizations will be displayed to help us understand the overall mood of the people on Twitter regarding the input provided by the user.

Project Made By:

- Sarthak Ahir
- Dharmik Bhanushali
- Mihir Bhanushali
- Chinmay Belekar

Enter the topic you are interested in (Press Enter once done)

india

Tweets have been Extracted !!!!

Total Tweets Extracted for Topic 'india' are :195

Total Positive Tweets are : 63

Total Negative Tweets are : 9

Total Neutral Tweets are :123

See the Extracted Data

Below is the Extracted Data :

	Date	User	IsVerified	Tweet	Likes
0	2021-03-21 06:36:19	Bee.com	False	@CryptoIndia https://t.co...	0
1	2021-03-21 06:34:51	Bee.com	False	@rsad15710191 @T261 @Engl...	0
2	2021-03-21 03:29:39	Bee.com	False	@Yragi_Munidev @PMOIndia @...	0
3	2021-03-20 15:32:10	Bee.com	False	@umiransaxi206 hi	1
4	2021-03-20 15:31:06	Bee.com	False	@rijinderPathan @MortyPan...	0
5	2021-03-20 15:30:08	Bee.com	False	@hosaleshree @vitenoesen...	0
6	2021-03-20 02:36:18	Bee.com	False	@rod_____ open soucce...	0
7	2021-03-20 02:23:09	Bee.com	False	@rod_____ try https://...	0
8	2021-03-18 06:15:04	Bee.com	False	@msudhirbocza @lonnusk m...	0
9	2021-03-17 22:26:12	Bee.com	False	@SaaGwir @Nigeria why?	1

Get Count Plot for Different Sentiments

Get Pie Chart for Different Sentiments

Get Count Plot Based on Verified and unverified Users

Get WordCloud for all things said about india

Get WordCloud for all Positive Tweets about india

Get WordCloud for all Negative Tweets about india

Fig 3: Extracted Data for the output.

Number of recent tweets to use for analysis:

5 100 1000

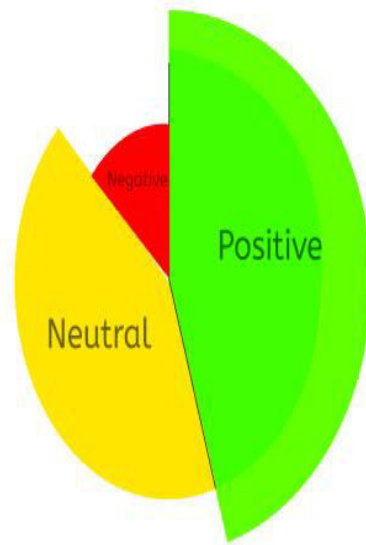


Fig 4: Output in Pie-chart.



About App

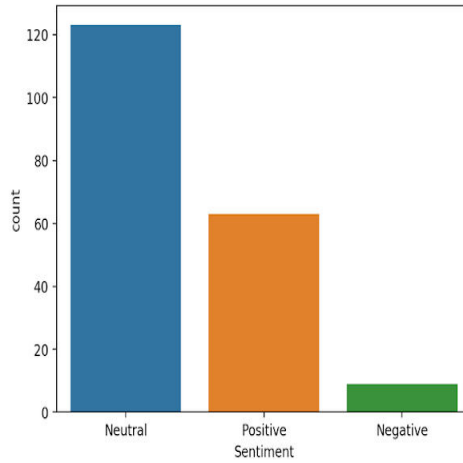
A Twitter Sentiment analysis Project which will extract tweets for the Twitter Handle or the topic selected by the user. The extracted tweets will then be processed and Sentiment Analysis will be performed on them to determine the Sentiments of those tweets. AS output, the various Visualizations will be displayed to help us understand the overall mood of the people on Twitter regarding the input provided by the user.

Project Made By:

- Sarthak Ahir
- Dharmik Dhanushali
- Mihir Bhanushali
- Chinmay Belekar

Count Plot for Different Sentiments

AxisSubPlot (0.125, 0.11; 0.775x0.77)



- Get Pie Chart for Different Sentiments
- Get Count Plot Based on Verified and unverified Users
- Get WordCloud for all things said about india
- Get WordCloud for all Positive Tweets about india
- Get WordCloud for all Negative Tweets about india
- Exit

Fig 5: Output in Count plot

Number of recent tweets to use for analysis:

Slider: 500 (range 1 to 1000)

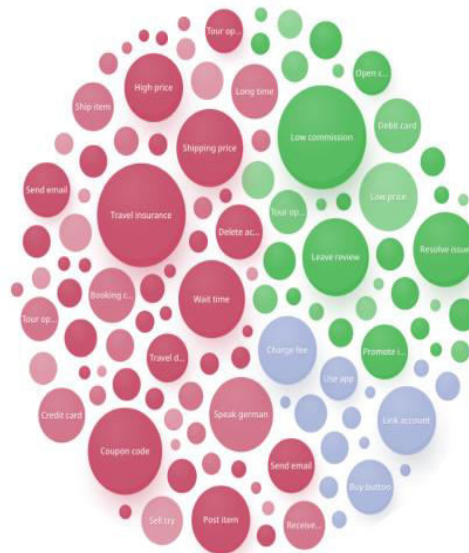


Fig 6: Output in Word Cloud



VI. CONCLUSION

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.

As of now we are exploring Parts of Speech separate from the unigram models, we can try to incorporate POS information within our unigram models in future. So, say instead of calculating a single probability for each word like $P(\text{word} | \text{obj})$ we could instead have multiple probabilities for each according to the Part of Speech the word belongs to. For example we may have $P(\text{word} | \text{obj}, \text{verb})$, $P(\text{word} | \text{obj}, \text{noun})$ and $P(\text{word} | \text{obj}, \text{adjective})$ used a somewhat similar approach and claims that appending POS information for every unigram results in no significant change in performance (with Naive Bayes performing slightly better and SVM having a slight decrease in performance), while there is a significant decrease in accuracy if only adjective unigrams are used as features. However, these results are for classification of reviews and may be verified for sentiment analysis on micro blogging websites like Twitter.

REFERENCES

- [1] Li Yang; Ying Li; Jin Wang; R. Simon Sherratt; "Sentiment Analysis for E-Commerce Product Reviews in Chinese Based on Sentiment Lexicon and Deep Learning" **Published in:** IEEE Access (Volume: 8, Jan. 2020)
- [2] Abdul Razzaq; Muhammad Asim; Zulqarnain Ali; Salman Qadri; Imran Mumtaz; Dost Muhammad Khan; Qasim Niaz, "Text sentiment analysis using frequency-based vigorous features" **Published in:** China Communications (Volume: 16, Issue: 12, Dec. 2019)
- [3] Shihan Elbagir Saad; Jing Yang "Twitter Sentiment Analysis Based on Ordinal Regression" **Published in:** IEEE Access (Volume: 7, Nov. 2019)
- [4] Pak Alexander, and Patrick Paroubek. "Twitter as a Corpus for Sentiment Analysis and Opinion Mining." LREC. Vol. 10, 2010.
- [5] Agarwal, Apoorv, et al. "Sentiment analysis of twitter data." Proceedings of the Workshop on Languages in Social Media. Association for Computational Linguistics, 2011.
- [6] Ortigosa, Alvaro, José M. Martín, and Rosa M. Carro. "Sentiment analysis in Facebook and its application to e-learning." Computers in Human Behavior 31 (2014)
- [7] Aisopos, Fotis, et al. "Content vs. context for sentiment analysis: a comparative analysis over microblogs." Proceedings of the 23rd ACM conference on Hypertext and social media. ACM, 2012.
- [8] Jebaseeli, A. N isha, and E. Kirubakaran. "A Survey on Sentiment Analysis of (Product) Reviews. " International Journal of Computer Applications 47.11 (2012).
- [9] Arjit Chatterjee and Dr. William Perrizo "Investor classification and sentiment analysis"
- [10] Efthymios Kouloumpis, Theresa Wilson and Johanna Moore. Twitter Sentiment Analysis: The Good the Bad and the OMG! *In Proceedings of AAAI Conference on Weblogs and Social Media (ICWSM)*, 2011.
- [11] Johann Bollen, Alberto Pepe and Huina Mao. Modelling Public Mood and Emotion: Twitter Sentiment and socio-economic phenomena. *In Proceedings of AAAI Conference on Weblogs and Social Media (ICWSM)*, 2011.



INNO  **SPACE**
SJIF Scientific Journal Impact Factor

**Impact Factor:
7.512**

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 **9940 572 462**  **6381 907 438**  **ijirset@gmail.com**



www.ijirset.com

Scan to save the contact details