

# Twitter Sentimental Analysis

SAHIL KADAM<sup>1</sup>, ISHANI SHIRSAT<sup>2</sup>, AKASH DWIVEDI<sup>3</sup>, DR. SAVITA SANGAM<sup>4</sup>

<sup>1, 2, 3, 4</sup> Department of Information Technology, Shivajirao S. Jondhale College of Engineering, Dombivli.

**Abstract-** This paper 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. Due to this large amount of usage we hope to achieve a reflection of public sentiment by analyzing the sentiments expressed in the tweets. Analyzing 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.

**Indexed Terms-** Twitter, Social Media, Analysis, Sentiment.

## I. INTRODUCTION

Data analysis is the process of applying organized and systematic statistical techniques to describe, recap, check and condense data. It is a multistep process that involves collecting, cleaning, organizing and analysing. Data mining is like applying techniques to mold data to suit our requirement. Data mining is needed because different sources like social media, transactions, public data, enterprises data etc. generates data of increasing volume, and it is important to handle and analyze such a big data. It won't be wrong to say that social media is something we live by. In the 21st century social media has been the game changer, be it advertising, politics or globalization, it has been estimated that data is increasing faster than before and by the year 2020; about 1.7 megabytes of additional data will be generated each instant for each person on the earth. More data has been generated in the past two years than ever before in the history of the mankind. It is clear from the fact that the number of internet users are

now grown from millions to billions. Database which is opted for the proposed study is from Twitter. It is now day's very popular service which provides facility of micro blogging. In this people write short messages generally less than 140 characters, about 11 words on average. It is appropriate for analysis as the number of messages is large. It is much easier task as compared to searching blogs from the net. The objective of the proposed analysis, 'Sentiment Analysis', is the analysis of the enormous amount of data easily available from social media. Algorithm generates an overall sentiment score from the inputted topic in terms of positive, negative or neutral, further it also works on finding the frequency of the words being used. Word cloud that is a pictorial representation of words based on frequency occurrence of words in the text is also generated. Calculation is actualized utilizing R attributable to its component rich, thorough and expressive abilities for measurable information.

## II. PROBLEM DEFINITION

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 about 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 result on how positive or negative or neutral are peoples about it.

## III. LITERATURE SURVEY

Sentiment analysis is a growing area of Natural Language Processing with research ranging from document level classification (Pang and Lee 2008) to learning the polarity of words and phrases (e.g., (Hatzivassiloglou and McKeown 1997; Esuli and Sebastiani 2006)). Given the character limitations on tweets, classifying the sentiment of Twitter messages is most similar to sentence level sentiment analysis

(e.g., (Yu and Hatzivassiloglou 2003; Kim and Hovy 2004)); however, the informal and specialized language used in tweets, as well as the very nature of the microblogging domain make Twitter sentiment analysis a very different task. It's an open question how well the features and techniques used on morewell-formed data will transfer to the microblogging domain. Just in the past year there have been a number of papers looking at Twitter sentiment and buzz (Jansen et al. 2009; Pak and Paroubek 2010; O'Connor et al. 2010; Tumasjan et al. 2010; Bifet and

Frank 2010; Barbosa and Feng 2010; Davidov, Tsur, and Rappoport 2010). Other researchers have begun to explore the use of part-of-speech features but results remain mixed. Features common to microblogging (e.g., emoticons) are also common, but there has been little investigation into the usefulness of existing sentiment resources developed on non-microblogging data. Researchers have also begun to investigate various ways of automatically collecting training data. Several researchers rely on emoticons for defining their training data (Pak and Paroubek 2010; Bifet and Frank 2010). (Barbosa and Feng 2010) exploit existing Twitter sentiment sites for collecting training data. (Davidov, Tsur, and Rappoport 2010) also use hashtags for creating training data, but they limit their experiments to sentiment/non-sentiment classification, rather than 3-way polarity classification, as we do. We use WEKA and apply the following Machine Learning algorithms for this second classification to arrive at the best result:

- K-Means Clustering
- Support Vector Machine
- Logistic Regression
- K Nearest Neighbours
- Naïve Bayes
- Rule Based Classifiers

#### IV. CONCEPTUAL OVERVIEW OF THE PROJECT

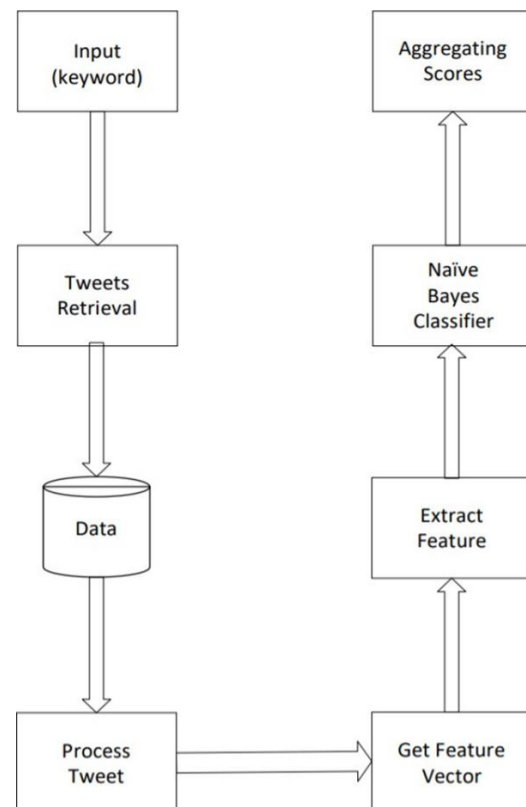
This proposal is an application which is used to analyze the tweets. We will be performing sentiment analysis in tweets and determine where it is positive, negative or neutral. This web application can be used by any organization office to review their works or by political leaders or by any others company to review about their products or brands. The main feature of our

web application is that it helps to determine the opinion about the peoples on products, government work, politics or any other by analyzing the tweets. Our system is capable of training the new tweets taking reference to previously trained tweets. The computed or analyzed data will be represented in various diagram such as Pie- chart, Bar graph and Word cloud.

- Project summary and purpose

This project of analysing 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 unlabelled data samples (tweets) according to whichever pattern model best describes them.

- Design



```

main - Notepad
File Edit Format View Help
from textblob import TextBlob
import tweepy
import matplotlib.pyplot as plt

def percentage(part, whole):
    return 100 * float(part) / float(whole)

consumerKey = "C1c7p05F0hTahuqcajz1hi61"
consumerSecret = "F17gJaw5Tg5ZWhqzZhiY6dJAVgWEG9MB8m3Pzpk8BcFH0rX9F"
accessToken = "4209947295-1ZfZygrIht0IKQu1hIdgsnyG1zKxa2Z87pLE882"
accessTokensSecret = "RrtVvypnPGAS1j53L9J8Sr10jiB56zudmPkSq1u1z9r"

auth = tweepy.OAuthHandler(consumerKey, consumerSecret)
auth.set_access_token(accessToken, accessTokensSecret)
api = tweepy.API(auth)

searchTerm = input("Enter keyword/hashtag to search about:")
noofSearchTerms = int(input("Enter how many tweets to analyze:"))
tweets = tweepy.Cursor(api.search, q=searchTerm).items(noofSearchTerms)

positive = 0
negative = 0
neutral = 0
polarity = 0

for tweet in tweets:
    # print(tweet.text)
    analysis = TextBlob(tweet.text)
    polarity += analysis.sentiment.polarity

    if(analysis.sentiment.polarity == 0):
        neutral += 1
    elif(analysis.sentiment.polarity < 0):
        negative += 1
    elif(analysis.sentiment.polarity > 0):
        positive += 1

positive = percentage(positive, noofSearchTerms)
negative = percentage(negative, noofSearchTerms)
neutral = percentage(neutral, noofSearchTerms)

positive = format(positive, '.2f')
neutral = format(neutral, '.2f')
negative = format(negative, '.2f')

print("how people are reacting on " + searchTerm + "by analyzing" + str(noofSearchTerms) + "Tweets.")

if(polarity == 0):
    print("Neutral")
    
```

```

main - Notepad
File Edit Format View Help
auth.set_access_token(accessToken, accessTokensSecret)
api = tweepy.API(auth)

searchTerm = input("Enter keyword/hashtag to search about:")
noofSearchTerms = int(input("Enter how many tweets to analyze:"))
tweets = tweepy.Cursor(api.search, q=searchTerm).items(noofSearchTerms)

positive = 0
negative = 0
neutral = 0
polarity = 0

for tweet in tweets:
    # print(tweet.text)
    analysis = TextBlob(tweet.text)
    polarity += analysis.sentiment.polarity

    if(analysis.sentiment.polarity == 0):
        neutral += 1
    elif(analysis.sentiment.polarity < 0):
        negative += 1
    elif(analysis.sentiment.polarity > 0):
        positive += 1

positive = percentage(positive, noofSearchTerms)
negative = percentage(negative, noofSearchTerms)
neutral = percentage(neutral, noofSearchTerms)

positive = format(positive, '.2f')
neutral = format(neutral, '.2f')
negative = format(negative, '.2f')

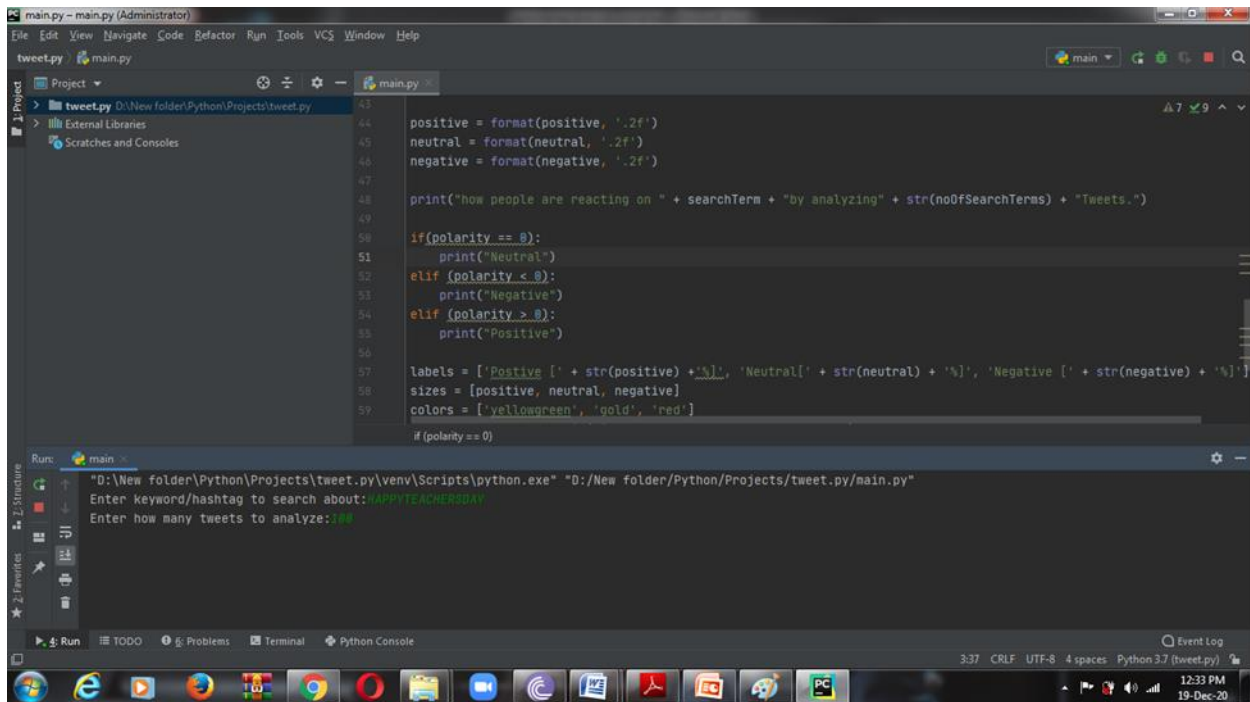
print("how people are reacting on " + searchTerm + "by analyzing" + str(noofSearchTerms) + "Tweets.")

if(polarity == 0):
    print("Neutral")
elif (polarity < 0):
    print("Negative")
elif (polarity > 0):
    print("Positive")

labels = ['Positive [' + str(positive) + '%]', 'Neutral[' + str(neutral) + '%]', 'Negative [' + str(negative) + '%]']
sizes = [positive, neutral, negative]
colors = ['yellowgreen', 'gold', 'red']
patches, texts = plt.pie(sizes, colors=colors, startangle=90)
plt.legend(patches, labels, loc="best")
plt.title('How people are reacting on ' + searchTerm + ' by analyzing ' + str(noofSearchTerms) + ' Tweets. ')
plt.axis('equal')
plt.tight_layout()
plt.show()
    
```

## V. RESULT

- Input:

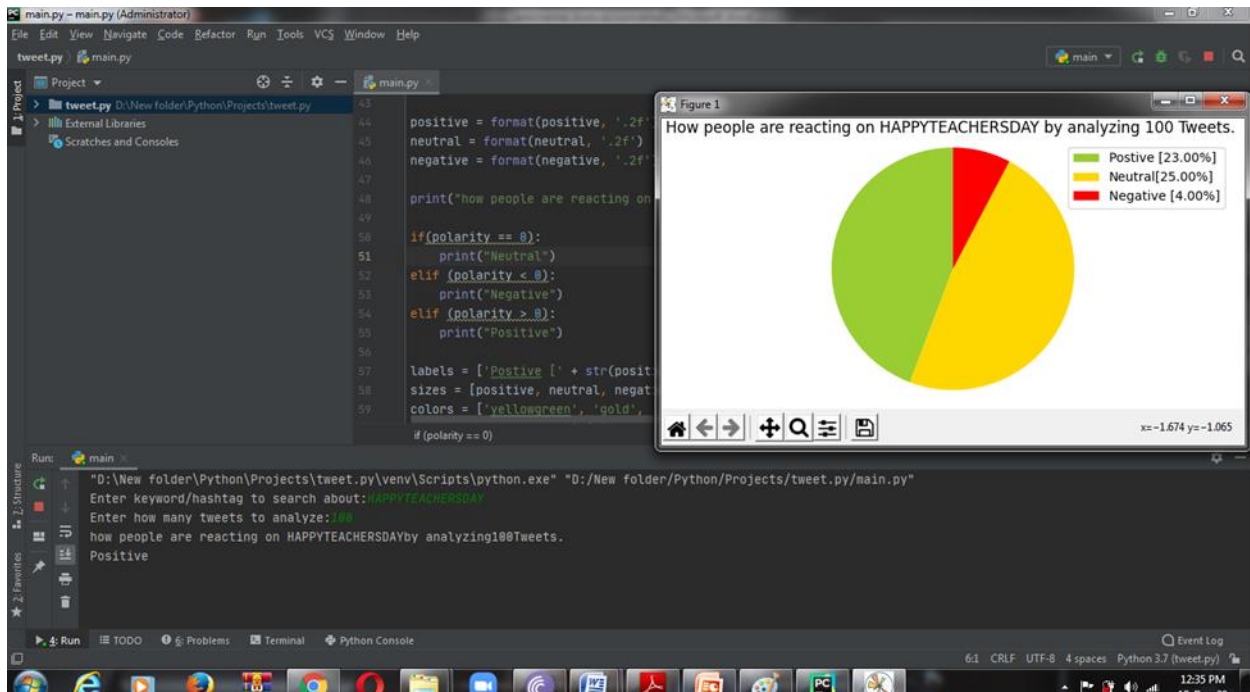


```
43
44 positive = format(positive, '.2f')
45 neutral = format(neutral, '.2f')
46 negative = format(negative, '.2f')
47
48 print("how people are reacting on " + searchTerm + "by analyzing " + str(noOfSearchTerms) + "Tweets.")
49
50 if(polarity == 0):
51     print("Neutral")
52 elif (polarity < 0):
53     print("Negative")
54 elif (polarity > 0):
55     print("Positive")
56
57 labels = ['Positive [' + str(positive) + '%]', 'Neutral[' + str(neutral) + '%]', 'Negative [' + str(negative) + '%]']
58 sizes = [positive, neutral, negative]
59 colors = ['yellowgreen', 'gold', 'red']
60
61 if (polarity == 0)
```

Run: main

```
"D:\New folder\Python\Projects\tweet.py\venv\Scripts\python.exe" "D:\New folder\Python\Projects\tweet.py/main.py"
Enter keyword/hashtag to search about:HAPPYTEACHERSDAY
Enter how many tweets to analyze:100
```

- Output:



## CONCLUSION

The experimental studies performed through the successfully show that hybridizing the existing machine learning analysis and lexical analysis techniques for sentiment classification yield comparatively outperforming accurate results. For all the datasets used, we recorded consistent accuracy of almost 90%.

The first method that we approached for our problem is Naïve Bayes. It is mainly based on the independence assumption. Training is very easy and fast In this approach each attribute in each class is considered separately. Testing is straightforward, calculating the conditional probabilities from the data available. One of the major task is to find the sentiment polarities which is very important in this approach to obtain desired output. In this Naïve Bayes approach we only considered the words that are available in our dataset and calculated their conditional probabilities. We have obtained successful results after applying this approach to our problem.

Clearly from the success of Hybrid Naive Bayes, it can positively be applied over other related sentiment analysis applications like financial sentiment analysis (stock market, opinion mining), customer feedback services, and etc.

## ACKNOWLEDGEMENT

We take this opportunity to place our sincere gratitude towards our Project Guide Dr. Savita Sangam for her motivation and constant support throughout the project work.

We also express our deepest thanks to our head of the Department Dr. Savita Sangam whose benevolent helps us making available the computer facilities to us for our project in our laboratory and making it true success. Without her kind and keen coordination our project would have been stifled to standstill.

We would also like to sincerely thanks to our Project Coordinator Prof. Prajka Khaire for her constant support and motivation throughout the project work.

We would also like to thanks our college Principal Dr.

J. W. Bakal for providing lab facilities and permitting us to go on with our project. We would also like to thank our colleagues to help us directly or indirectly during our project.

## REFERENCES

- [1] Albert Biffet and Eibe Frank. Sentiment Knowledge Discovery in Twitter Streaming Data. *Discovery Science, Lecture Notes in Computer Science*,2010, Volume 6332/2010, 1-15, DOI: 10.1007/978-3-642-16184-1\_1
- [2] Alec Go, Richa Bhayani and Lei Huang. Twitter Sentiment Classification using Distant Supervision. *Project Technical Report, Stanford University*,2009.
- [3] Alexander Pak and Patrick Paroubek. Twitter as a Corpus for Sentiment Analysis and Opinion Mining. *In Proceedings of international conference on Language Resources and Evaluation (LREC)*,2010.
- [4] Andranik Tumasjan, Timm O. Sprenger, Philipp G. Sandner and Isabell M. Welp. Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment. *In Proceedings of AAAI Conference on Weblogs and Social Media (ICWSM)*,2010.
- [5] Bo Pang, Lillian Lee and Shivakumar Vaithyanathan. Thumbs up? Sentiment Classification using Machine Learning Techniques. *In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*,2002.
- [6] Chenhao Tan, Lilian Lee, Jie Tang, Long Jiang, Ming Zhou and Ping Li. User Level Sentiment Analysis Incorporating Social Networks. *In Proceedings of ACM Special Interest Group on Knowledge Discovery and Data Mining (SIGKDD)*, 2011.
- [7] 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.