

Sentiment Analysis of Trending Hashtags on Twitter

Pritpal Singh, Chandigarh University, Mohali
Diwanshi, Chandigarh University
Vanshika Rathore, Chandigarh University
Mohammed Saif Uddin Siddiqui, Chandigarh University

Article Info**Volume 82****Page Number: 16080 - 16087****Publication Issue:****January-February 2020****Article History****Article Received:** 18 May 2019**Revised:** 14 July 2019**Accepted:** 22 December 2019**Publication:** 28 February 2020**Abstract:**

When the modern day world was hit upon by bigger data sets , we all found a strong need to handle data effectively and storage problems aroused .Main motive was to build a framework and support in order to store and handle data . Data science served as a secret helping agent here. Data science was foreseen as the future of artificial intelligence . It is a blend of various working tools, algorithmic modules , and machine learning principles with the major goal to discover the hidden patterns from the raw available data. It is one of the best methodologies to ensure that the data is searched and sorted upon in a valid and verified way so as to maintain the accuracy and integrity of data

Keywords: WBAN, Privacy, Security, Location

I. INTRODUCTION

The project “ Sentimental Analysis of Trending Hashtags on Twitter” aims at analysing the data sets in a generic and efficient way . This project will focus on analysis of the viewpoints of the normal people on specific topics that may be presented as trending hashtags on an independent platform that we have chosen here as Twitter which includes a stream of tweets [1] .

We aim at providing sentimental analysis of the entered hashtags that we want to search for about whether people are positive or negative about a certain topic. We even provided the option for entering the number of tweets to be looked upon then the result is calculated based on Twitter data . It uses packages like Tweepy to fetch data from Twitter and display results accordingly with the help of pie chart so as to easily depict results.

Due to the fact that micro blogging websites have been emerging out to become a major source of various kinds of information [1]. This is because of the nature of the micro blogs on which much population post and present their real time messages

and sentiments about their personal opinions on a variety of topics , discussion of currently occurring issues, complaining , and expressing positive sentiment for products they admire in daily life. In fact, companies producing such products have all begun to poll these micro-blogs to gain a sense of generalised sentiment for their intended product. Many a times these companies tend to study user reactions and replies to other users on microblogs. One challenge is to create a technology to detect ,analyse and summarize an overall sentiment of the same[2].Although there are numerous ways to extract out data about the sentiments of a person regarding a particular product or any other service , still many organisations or field workers find it difficult to make it to the results of a person's sentiment.

- Sentiment Analysis of Applications that are Web Based Focus mainly on a Single Tweet Only. With the tremendous extension of the World Wide Web, individuals are now utilising social media platforms like Twitter to cultivate huge volumes of opinion based texts in the way of tweets that are accessible for the process of sentiment analysis .

This interprets to a big volume of data and information from an individual point of view which makes it extremely difficult to pull out sentences, then read them, followed by the process to analyze them tweet by tweet, and summarize these analysis to finally organize them into a suitable format which is easily understandable in a time conscious manner[2].

• Issue of Sentiment Analysis with ill suited English Language

Ill suited language is defined as the utilisation of conversational and argot in communication, hiring the traditional protocols of spoken language just like 'is not' and 'isn't'. Not every system is capable of detecting the sentiments by using informal and ill suited language and it may result in disturbing the analysis and decision forming or making process. For instance, if we talk about emoticons or emojis as we call them, they are a picture based representation of what human facial expressions look like, that in the non availability of body language and metrics serve the purpose of drawing a recipient's attention towards the temper or intent of a sender's official verbal communication, enhancing and converting the above's interpretation. For instance, ☺ implies a happy or joyous state of mind. Systems that are currently in working do not have enough data to be able to analyse data to use the analysis to extract feelings out of the emojis. As we the human beings, commonly turn towards the emoji use to easily and to the point expression of what we are unable to put down into words [3]. Being unable to analyze, it lays the company or any other work field at a huge loss. Short-forms sometimes referred to as slangs are widely being used even in the facilities like WhatsApp, Facebook, Instagram, SMS or any other text exchanging app. The application of short forms if used more often on platforms like Twitter will greatly help in minimising the number of characters that are used. This fact is because of the reason

that the Twitter platform has been put down a specific limit on the characters ranging with in t o 1 4 0 . For example, 'Ttyl' refers to "talk to you later". Main objective here is to mainly investigate the sentiment analysis in microblogging websites like Twitter that in the view to gain an effective feedback from the user of the product followed by the second one, is basically to develop a service to facilitate user's review or view on that thing that simply allows the company, any organisation, any field group or individual human being to sentiment first and analyze next a good number of tweets into a utilisable data set format.

The project revolves around basically usage of natural language processing and data science. These two areas have been of immense importance in the ever growing technological advancements.

Data science refers to the investigation of data or what we call as information. It revolves around the development of methodologies for recording data, or its storage, and analysis of data for efficient extraction of important and utilisable information. The main aim of this field is to gather insights and data sets or knowledge bases from any source of data — may be structured or unstructured or both.

Data sciences revolves around having various applications :

1. InternetSearching
2. Recommendationsystems
3. SpeechRecognition
4. Advanced ImageRecognition
5. Airline routeplanning

Natural language processing often acronymed as NLP can be defined as the classification of artificial intelligence or AI that trade in analysis, interpretation and generation of the languages which human beings utilise naturally so as to interact with computers or machines in both written and readable or spoken forms with the help of natural human languages and not with the computerlanguages.

Sentiment analysis which is used here what is also known as opinion mining is referred to as the broad area of natural language processing or NLP, ciphericlinguistics, text mining all of which revolves

around the calculative study of emotions or opinions and sentiments expressed in the most preferred form i.e. texts[3].

Twitter sentiment analysis is basically that the sentiment or emotion is likely to be found in the section of comments or tweets so as to facilitate helpful indicators to fulfill many varied purposes [4]. It is also the case that as stated that an emotion or more particularly an emotion can be classified into three broad categories, that are negative, neutral and positive ones. It is a very basic and NLP technique to elaborate and analyse an already expressed emotion, feelings or sentiment within a fixed and selected range of tweets.

Techniques of Sentiment Analysis

Polarity can be defined as the most used form, that is if a sentence or text is positive, neutral or negative. However, sentiment analysis has ways for the assignment of polarity like as follows:

1. Natural Language Processing(NLP) These kind of techniques are primarily based on ML and they use a general learning algorithm which is combined with a large collection of data or bigger data sets to memorise the set of rules. It has been evaluated as a NLP i.e. natural language processing, at most levels of modularity or distinction. Starting from being a documentation based activity, it has been moved up to the word group level and more precisely nowadays at the phrase based level also. It is an area in cse field which revolves around making the computers able to derive meaning and valid results from human usable language [4].
2. Artificial Neural Networks also acronymed as ANN stated that ANN or widely called as artificial neural network is basically an arithmetic technique that inter relates and connects the group of many artificial neurons. It thus processes data sets or information with the help of the approach of connections to computation[4].
3. Support Vector Machine also referred as SVM are primarily used for detection of the sentiments or emotions hidden in the tweets. SVM is also capable to pull and interpret to gather upto 70%-

81.3% of efficiency on the given or chosen test set. Using SVM which is given training from the labeled data set, they are capable to obtain 81.3% in classification of emotion or sentiment accuracy[5].

II. RELATEDWORK

1. Sentimental Analysis of TwitterData

In that they scrutinised sentiment analysis on Twitter based data sets.

2. Twitter as a corpus for sentiment analysis and opinion mining

They aimed at gathering a collection of approx 300000 texts in the form of various posts from Twitter that were evenly split among three classifications of texts like ones having positive emotions, such as happiness, joy or amusement, negative emotions, such as sadness, tears or anger or the neutral ones neither expressing smiles nor sorrows.

3. Twitter sentiment analysis: The good the bad and the omg!

In this, they interrogated the usability of some linguistic features for detection of the sentimental analysis of Twitter based posts or text messages.

4. Semantic sentiment analysis of twitter

They used the application of the strategy to predict sentiment analysis of three distinguish kinds of Twitter data sets. They even compared with the help of an approach that was based on the sentiment centric topic analysis.

5. Topic based sentiment analysis using twitter: which is a graph-based hashtag utilisable sentiment classification approach

Their task focussed on automatically generating the overall sentiment polarity for a desired hashtag in a fixed or limited period of time.

6. Twitter brand sentiment analysis

An efficient and mixed system based on n-gram analysis plus dynamic artificial neural network so as to demonstrate the overall efficiency and accuracy of the so created Twitter-specific lexicon in comparison to a conventional sentiment lexicon, thus developed a resultant comparison based

sentimental distinguishing model using SVM technique.

7. Emoticon smoothed language models for twitter sentimentanalysis

They efficiently integrated the two distinguished data sets and laid them into the same learning framework and also presented a novel based model, called an emoticon smoothed language model (ESLAM).

8. Contextual semantics for sentiment analysis of Twitter

They provided a semantic sentiment representation of words that were known as

SentiCircle. SentiCircle was capable of capturing the contextual based semantic of words from their corresponding occurrences.

All these categories carry mentioned below :-the particular meanings as

III. PROPOSEDWORK

We presented a method to collect and analyse a data set consisting of a mix of positive plus negative sentimental phases , and a collection of objective based texts in the form we call as Twitter tweets . Our method facilitates us to collect both the negative and positive sentiment expressions so that no man power or effort is required to analyse and distinguish the documents. Texts are also collected and then analysed on their own. The size of the collected data set can be effectively large. Also , we carried out a statistical linguistic analysis of the gathered data set. We used the gathered data sets to build out a sentiment based distinction system for websites or web platforms based on microblogging. We thus carried out an experimental or practical based evaluation on the data sets of realistic world micro blogging posts in order to prove that our own method is way more accurate and performs much better than the methodologies that are already proposed .

We focussed on providing options to the user of our project to enter any hashtag that may be trending actually or is of interest to the user or is of any relevance or irrelevance but has to looked upon to get the statistical analysis of sentiments on the

related tweets to the hashtags . After the entering of the related hashtag we ask the user the number of tweets he would like to search upon like 100 , 1000 or 100000 any number he can search for , then the twitter data is looked upon to analyse the corresponding results to gain the stats in the form of piecharts.

The opinions are classified as follows : -

1. Stronglypositive
2. Weaklypositive
3. Stronglynegative
4. Weaklynegative
5. Positive
6. Negative
7. Neutral
- 8.

These are the classification of opinions.

Table 1 : Parameters presented in table includes Emoticon and their corresponding Polarity that can be like Positive , negative or neutral with respect to the Emoticons as mentioned in the table :

Emoticon	Polarity
:-) :o) :] :c) :3 :)	Positive
C: :D	Extremely Positive
:c :[:(:-(	Negative
DX D8 D= v.v D;	Extremely Negative
: :!	Neutral

Strongly positive indicates that the views regarding the particular hashtag are completely in favour i.e. totally positive response from the audience regarding that particular hashtag .Generally , the percentage in this category is not that high .

Strongly negative indicates that the views regarding the particular hashtag are completely against the particular hashtag and are totally in contrary to the particular hashtag .Generally , it has a lower percentage as compared to the other categories.

Weakly positive indicates that the views are not negative but are a little positive just to imply that they are not negative

.This is a favourable condition in an unfavourable way such that the audience in majority is against the topic .

Weakly negative indicates that the views are not positive but are a little negative just to imply that they are not positive

.This is a un-favourable condition in a favourable way such that the audience in majority is in favour of the topic .

Positive indicates favourability to the topic as per the stats and implies that the audience is in favour of the topic .

Negative indicates unfavourable response to the topic as per the stats and implies that the audience is against the topic . Neutral indicates neither favourability nor unfavourable response to the topic as per the stats and implies that the audience is neutral to the topic [6].

There are different types of hashtags :-

First type is Campaign specific hashtags , second one is Trending hashtags and the last one is Always on hashtags .

IV. RESULTS

We thus created an analyser for sentiments on Twitter platform called as “Sentiment Analysis of Trending HashtagsonTwitter” so as to find the weakly positive , weakly negative , strongly positive , strongly negative , negative , positive and neutral percentage of views on a particular topic . Here , we used natural language processing which is entailing and applying out algorithms in order to identify and also extract out the natural language rules so as to convert the unstructured language data into a form that computers are able to understand.

Thus we enabled our user to get data interpretations regarding the hashtags he desired to check positive and negative user opinion ratio and understand the hashtags’s impact accordingly. PyCharm platform helped us in

laying down our code to get the intended output. thus we were able to perform according to the requirement and to analyse our tweets accordingly

.

This project will help the easy and efficient analysis of Twitter data . If anyone wants to know about how people feel about a certain topic then this is surely the best way to find that out as it will be easiest and fastest way to find it out. Also , pictorial representation of data enables us to easily understand the results and present the final outcome with ease. Many a times we require certain viewpoints of people regarding certain implementations . The traditional data collection methods are way too complicated and time consuming thus is not much implemented and preferred by people . The better way is to present the analysis pictorially or graphically so as to keep our representations simple and easily manageable . Thus, elaborating the analysis via sentiment prediction is by far the most efficient way to depict the collective response regarding the topic we desired from the desired data set.

Hashtags are an easy way to create awareness about a particular topic and with the help of the tweets related to those hashtags help us to generate result in our project [6]. For example - Searching Hashtag like “acid attacks” gave result in the form of a general report i.e. a report presenting overall percentage like how many tweets are positive or negative or weakly or strongly positive/negative or neutral and so on. Also , it is represented that the overall major opinion is positive or negative or anything else .

Procedure to be followed :-

1. Enter trending hashtag for the analysis
 2. Enter the number of tweets to be searched
- Packages used for the working of our proposed work are:-

1. Tweepy

It is an open sourced package in python that provides us a very easy way to access usage of the Twitter API with the help of Python. Tweepy consists of a set of classes and

methods that represents the models of Twitter and API endpoints, and it efficiently handles various implementation details, such as: Data encoding methods and decoding methods.

2. Textblob

TextBlob is a Python (2 and 3) based library for processing of text based data. It provides an easy and simplified API for entering into natural language processing (NLP) functions such as parts of speech, extraction of noun phrase, sentiment based analysis, classification of things, translation of sentences, and many more.

3. Matplotlib

It is a plotting facilitation library for the Python programming language and its numeric based mathematical extension called NumPy. It ensures to avail us with an object oriented API for putting the various plots down into applications with the help of general purpose GUI tools such as Tkinter, wxPython, or GTK+etc.

4. Jinja2

It is a modernised template based language for helping Python developers to ease and lighten their work. It was launched after Django's template. It is used to make HTML (Hypertext Markup Language), XML and many other markup formats which are turned back to the user through a Hypertext Transfer Protocol request.

5. Mpld3

This package is very easy to use i.e. we can simply choose any script to generate a matplotlib plot, then run that plot via one of the mpld3's convenience routines, and put back the results in a web page.

These packages facilitated us to pull out information from Twitter platform and thus helped them in shaping it to a desired conclusion so as to create an accurate and efficient outcome which is also desired as per the entered hashtags.

Platform used - PyCharm

PyCharm basically provides help to the programmers to use Python more efficiently in big data analysis and data science projects.

It supports some of the widely used scientific and most important libraries for Python like — NumPy, Anaconda

,Tweepy, mpld3, Jinja2, Textblob and Matplotlib etc. Statistical Representation using :- Pie chart

A pie chart (also called as circle chart) is a circular figure presenting statistical graphics, which is further divided into pieces to elaborate the numerical proportions. In these circle charts, the length of the arc of each slice is proportional directly to the quantity that it is made to represent. Hashtags begun as an easy way to search and find out things on Twitter platform and are now available for cross-platform purposes and also other multi-purpose, utilised by organisations and people alike for almost everything; starting from major globalised events like the World Cup Final to solidarity moments like #MeToo.

A hashtag which is god is capable of capturing the imagination of the public residing online and helps out the various companies or organizations or individuals to drastically increase and emphasize their social media presence.

Yet tracking out these hashtags efficiently can be a major challenge provided the number of times they are used and mentioned in different countries or continents, in different languages, and in different contexts.

Analyzing the sentiment about our Twitter hashtags simply allow us to analyse whether our proposed idea will have either a positive or a negative impact.

A big amount of mentions on Twitter with negative sentiments indicates that our idea is bringing out certain bad impacts and we need to analyze this in much detail.

How people are reacting on various entered hashtags by analyzing different number of

```

Enter Keyword/Tag to search about: Acid attacks
Enter how many tweets to search: 100
How people are reacting on Acid attacks by analyzing 100 tweets
()
General Report:
Weakly Negative
()
Detailed Report:
16.00% people thought it was positive
21.00% people thought it was weakly positive
1.00% people thought it was strongly positive
2.00% people thought it was negative
11.00% people thought it was weakly negative
10.00% people thought it was strongly negative
37.00% people thought it was neutral
    
```

Figure 1 : Input for gathering sentiment analysis for hashtag “Acid attacks”

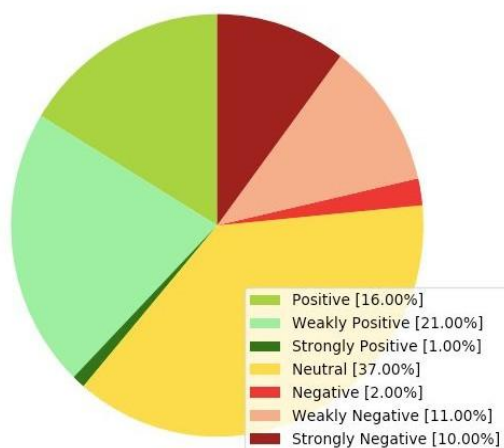


Figure 2 : Output for gathering sentiment analysis for hashtag “Acid attacks”

```

Enter Keyword/Tag to search about: Narendra Modi
Enter how many tweets to search: 1000
How people are reacting on Narendra Modi by analyzing 1000 tweets.
()
General Report:
Weakly Positive
()
Detailed Report:
8.70% people thought it was positive
8.00% people thought it was weakly positive
3.40% people thought it was strongly positive
1.30% people thought it was negative
2.50% people thought it was weakly negative
4.70% people thought it was strongly negative
71.20% people thought it was neutral
    
```

Figure 3 :Input for gathering sentiment analysis for hashtag “Narendra Modi”

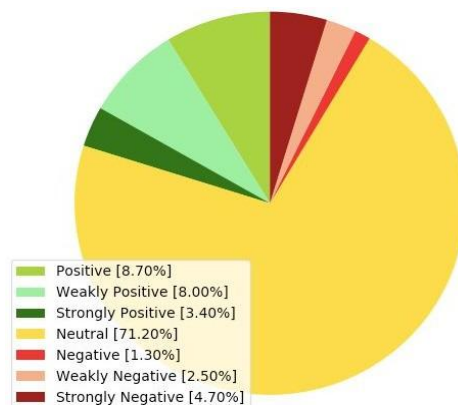


Figure 4 : Output for gathering sentiment analysis for hashtag “Narendra Modi”

```

Enter Keyword/Tag to search about: Racism
Enter how many tweets to search: 200
How people are reacting on Racism by analyzing 200 tweets.
()
General Report:
Weakly Positive
()
Detailed Report:
18.00% people thought it was positive
17.50% people thought it was weakly positive
0.50% people thought it was strongly positive
4.00% people thought it was negative
25.50% people thought it was weakly negative
1.00% people thought it was strongly negative
33.00% people thought it was neutral
    
```

Figure 5 :Input for gathering sentiment analysis for hashtag “Racism”

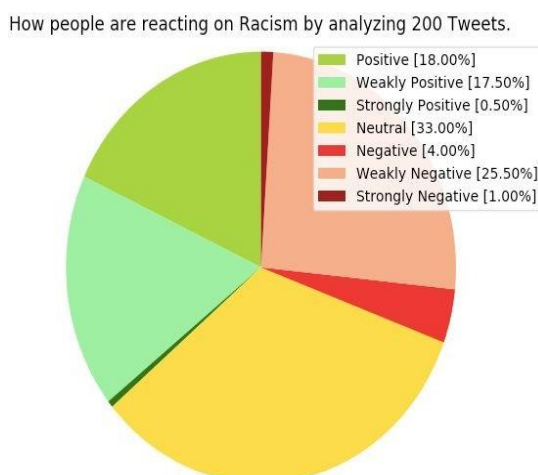


Figure 6 :Output for gathering sentiment analysis for hashtag “Racism”

V. CONCLUSION

We elaborated the results for analysing the sentiments of trending hashtags on Twitter. We used natural language processing for analysis of the tweets for searching entered hashtags and presented our results with the help of a pie chart . We showed out an effective and appropriate collection of experiments for running out the following tasks on twitter data that is nothing but a random chosen sample consisting of cluster of tweets. We found out various entered hashtags and looked upon the corresponding responses from people . For facilitating this feature dependent approach, we did feature based analysis that revealed that the most radiant features are the ones that provides a combination of the polarity of sentences / words and their corresponding parts of tags of speech. We strongly provides a conclusion that sentiment based analysis of Twitter based data sets is not much distinguishable from sentiment based analysis for other social or data gathering platforms . As a future plan , we will try and explore even better methods for analysis and processing natural languages to gain effective outputs out of the all available methods to us.

REFERENCES

- [1] https://www.researchgate.net/publication/301408174_Twitter_sentiment_analysis
<https://pdfs.semanticscholar.org/c114/7f3d9b46ff0a0c7c43b668123cb15a26120d.pdf>https://thesai.org/Downloads/Volume10No2/Paper_48-A
- [2] _Study_on_Sentiment_Analysis_Techniques.pdf
f Twitter as a corpus for sentiment analysis and opinion mining.
- [3] Robust sentiment detection on twitter from biased and noisy data
- [4] L Barbosa, J Feng - Proceedings of the 23rd international conference on ..., 2010 - dl.acm.org
- [5] Sentiment analysis on twitter
- [6] A Kumar, TM Sebastian - ... Journal of Computer Science Issues (IJCSI), 2012 – Citeseer
- [7] Singh, H.. (2017). Overcoming challenges in 3D NAND volume manufacturing. Solid State Technology. 60. 18-21.
- [8] Chen, Bill & Bottoms, Bill & Armstrong, Dave & Isobayashi, Atsunobu. (2015). ITRS 2.0: Heterogeneous Integration. Solid State Technology. 58. 13-17.
- [9] Choe, Jeongdong. (2015). Comparison of 1Ynm NAND architectures and beyond. Solid State Technology. 58. 24-25.
- [10] Vogler, Debra. (2013). A Preview of Semicon West 2013. Solid State Technology. 56. 30-32.
- [11] Monahan, Kevin. (2006). Addressing 32nm half-pitch challenges with double-patterning lithography. Solid State Technology. 49. 28-30.