International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-8 Issue-6, March 2020

Sentiment Classification System of Twitter Data using Python

G. Adhithya Balaji, Rashmita Khilar

Abstract*: Social media has developed drastically over the years. These days, individuals from all around the globe utilize online networking destinations to share data and information. Twitter is a well known communication site where users update information or messages known as tweets. Users share their day by day lives, post their opinions on everything, for example, brands and places. Various purchasers and advertisers utilize these tweets to accumulate bits of knowledge of their items and opinions on them. The aim of this paper is to exhibit a model that can perform sentiment analysis of real-time data collected from twitter and classify the tweets into positive, negative or neutral based on the sentiment expressed in them.

Keywords: Sentimental analysis, opinion, twitter, tweets.

I. INTRODUCTION

Sentimental analysis can be defined as the process of segregating the emotions in a phrase or a sentence to determine the way of thinking or the feelings expressed by a person. With the increase in the web technologies, Huge volumes of data are available on various websites where users are sharing & exchanging their ideas and opinion. These data are valuable for businessmen, marketing groups and even to the governments to know about the views and opinions of the people. An average of 6,000 tweets are tweeted on twitter per second all over the world, which makes it a significant source of information. Sentimental analysis can be carried out through different algorithms. Sentimental analysis is categorized into document, phrase and sentence level. In document level, a complete document is taken and processed to determine whether the sentiment is positive, negative or neutral. In phrase level, a particular phrase is chosen and analysed to see the polarity in the phrase. In sentence level, a sentence is considered for analysis and grouped into different categories of emotions expressed. In this paper, we are doing the sentence level analysis, with input to our model being the raw data extracted from the real-time tweets. All the sentences are taken into a .CSV file, then preprocessing is applied on these sentences and later using SVM machine learning algorithm the data is classified into positive, negative and neutral. The objective of this paper is to tackle the data to gather crucial information on popular conclusion that would help settle on smart business choices, political crusades, and better item utilization.

Manuscript received on February 10, 2020.
Revised Manuscript received on February 20, 2020.
Manuscript published on March 30, 2020.
* Correspondence Author
G. Adhithya Balaji*, Department of Information Technology, Saveetha School of Engineering, Chennai, Tamil Nadu, India.
Email: im.adhitya98@gmail.com
Rashmita Khilar, Department of Information Technology, Saveetha School of Engineering, Chennai, Tamil Nadu, India.
Email: rashmitakhilar.sse@saveetha.com

II. LITERATURE SURVEY

V. Kharde and S. Sonawane [1], has given an overview and a similar examinations of existing methods for assessment mining like AI and dictionary based methodologies, together with assessment measurements. The paper likewise gave an examination on twitter information stream utilizing different AI calculations like Naive Bayes, Max Entropy, and Support Vector Machine.

Pitiphat Santidhanyaroj, Talha Ahmad Khan [2], proposed a model framework to empower robotized sentiment based investigation of informal community information. They have developed a prototype system that utilizes the APIs provided by popular social networks such as Facebook and Twitter whereas the system retrieves and analyses the incoming data in realtime. They have also designed it to interact with the APIs and execute an on going basis to act as real-time service.

Subramaniam.G, Ranjitha.M [3], proposed a model to predict the user emotion using twitter data. and implemented the proposed system using HTML, CSS and python framework to analyse the data. The model works based on the twitter media and interprets the sentimental emotions to update the information on their survey website.

M.Trupthi, Suresh Pabboju [4], has given an intuitive programmed framework which predicts the estimation of the survey/tweets of the individuals posted in online networking utilizing hadoop. The system proposed extracted the data from SNS services which is done using Streaming API of twitter. First the extracted tweets are loaded into hadoop and later has been preprocessed using map reduce. A naïve bayes classification has been used in this system.

Po-Wei Liang, Bi-Ru Dai [5], proposed a framework that can consequently investigate the notions of messages. The issue of feeling mining calculations not having the option to function admirably in a miniaturized scale blog has been recognized and corrected. Now, the system could figure out how to consequently separate the arrangement of messages which contain suppositions, non-conclusion messages and decide their assumption.

Kunpeng Zhang, Yu Cheng, Yusheng Xie, Daniel Honbo, Ankit Agrawal, Diana Palsetia , Kathy Lee , Wei-keng Liao , Alok Choudhary [6], proposed to build up a recognizable proof framework called SES which executes three distinctive assumption distinguishing proof calculations.

The essential semantic standards were laid in the calculations utilized. They arranged the tweets as well as built up a supposition degree. Further, an AI model was utilized on highlights got from the yields of the three unique calculations utilized.

Apoorv Agarwal, Boyi Xie, Ilia Vovsha, Owen Rambow, and Rebecca Passonneau [7], examined the sentiment analysis on twitter data by introducing a POS- specific prior polarity feature. They have also explored the use of a tree kernel and also showed that the tree kernel outperformed the state of art baseline.

Prerna Mishra [8], discussed the sentiment analysis of twitter data and also performed a case study on Digital India. An opinion mining framework with modules such as pre-processing, feature extraction and polarity classification has been deployed to perform the sentimental analysis. The main focus of the paper is to capture the polarity of the sentiments captured from twitter data.

M.Lovelin Ponn Felciah,R.Anbuselvi [9], a survey which focuses on the different strategies and techniques for arranging the assessment from internet media datasets.

Anant Arora, Chinmay Patil, Stevina Correia [10], This paper exhibits an audit that covers the various strategies and approaches that are utilized in sentiment mining frameworks. Likewise, this paper features the different application regions and moves identified with the Opinion Mining.

III. ALGORITHM

Support Vector Machine is one of the most mainstream supervised learning calculations, which is utilized for classification and regression issues. In any case, essentially, it is utilized for Classification issues in Machine Learning. The objective of the SVM calculation is to make the best line or choice limit that can isolate n-dimensional space into classes so we can undoubtedly put the new information point in the right classification later on. This best choice limit is known as a hyperplane. SVM picks the extraordinary focuses/vectors that help in making the hyperplane. These extreme cases are called as support vectors, and thus calculation is named as Support Vector Machine

IV. PROPOSED SYSTEM

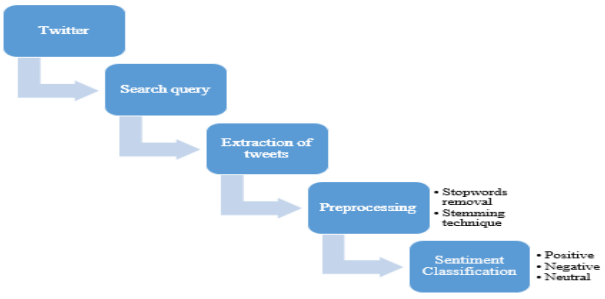

Figure 1: Proposed Block Diagram

A. Search Query:

In this module, user can search for the required query in the application. The search query then carries the process and displays the appropriate results of the query.

B. Extraction of Tweets:

Once the query is processed the live tweets from the real time are extracted using the Twitter API. The parameters are set to extract a minimum of 100 tweets related to the search keyword.

C. Preprocessing:

As soon as all the tweets are extracted from twitter, they are processed and the noise data are removed.

(1) Stopwords Removal:

A word reference based methodology has been used to expel stopwords from tweets. A nonexclusive stopword list containing 75 stopwords made utilizing mixture approach is utilized. The SVM calculation is executed as follows. The objective content is tokenized and individual words are put away in exhibit. One prevent word is perused from the stopword list. The stop word is then contrasted with target message in type of cluster utilizing successive hunt method. In the event that it coordinates, the word in exhibit is evacuated and furthermore the correlation is continued until length of the cluster. After expulsion of stopword totally, another stopword is perused from stopword list and again calculation runs persistently until all the stopwords are looked at. Resultant content without stopwords are shown, likewise required measurements like stopword evacuated, number of stopwords expelled from target content, absolute include of words in target content, include of words in resultant content, singular stop word include found in target content are shown.

(2) Stemming Technique:

After the removal of stopwords, the stemming process is carried out. Stemming is that the method of reducing inflected (or generally derived) words to their word stem, base or root kind typically a word kind. In simple, the postfix from each words like “ing”,“tion” etc. are removed.

D. Sentiment Classification:

After the stemming process, all the tweet terms containing the keyword are classified into positive, negative and neutral tweets. Collaborative filtering algorithm is used for classification. Here there is a positive and negative word dataset is present. By comparing with this, we can classify the tweets into positive, negative and neutral tweets.

V. RESULTS

In this paper, the data was extracted directly from real time twitter and were used for analysis. The proposed methodology used the SVM algorithm to classify the tweets into positive, negative and neutral.

As a result, the sentimental analysis carried out has shown significant and accurate results.

Example: Results for the filter "Samsung"

All the tweets that contain the word samsung are extracted and used for sentiment analysis.

Based on the analysis, the tweet is classified. The results and graph are displayed below.

```
# We print the most recent 5 tweets:
print("5 recent tweets:\n")
for tweet in tweets[:5]:
    print(tweet.text)
    print()
```

```
Number of tweets extracted: 200.

5 recent tweets:

Whether you're celebrating the new year or just want to personalize your
s://t.co/27aEYEDlqJ

The #YouthOlympics #TeamGalaxy athletes and influencers descended upon Sw
t.co/GSRwXwPLUC

The acquisition will help network operators accelerate and simplify rollc
s://t.co/JYnGCPsPJl

The new Galaxy XCover Pro, featuring premium design, business functionali
s://t.co/JjVkBeLrFG

Introducing Galaxy Book Flex alpha, a new variant of #GalaxyBookFlex that
co/NQS1nF4OFV
```

Figure 2: Tweets Extracted

```
# We print percentages:

print("Percentage of positive tweets: {}%
print("Percentage of neutral tweets: {}%"
print("Percentage of negative tweets: {}%
```

```
Percentage of positive tweets: 66.5%
Percentage of neutral tweets: 27.5%
Percentage of negative tweets: 6.0%
```

Figure 3: Sentiment Percentage

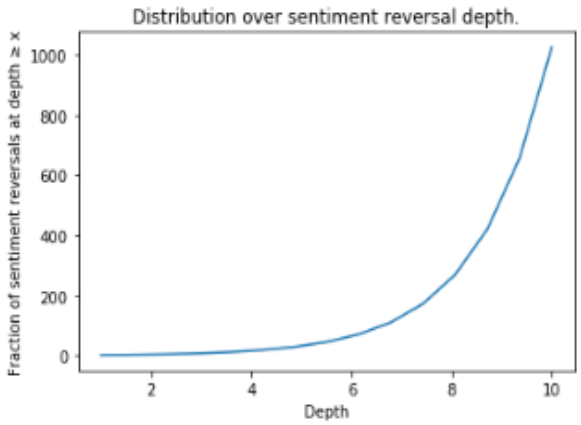

Figure 4: Sentiment Reversal Depth Graph

```
Accuracy Score : 0.8
Report :
              precision    recall  f1-score   support

          -1       0.67      0.67      0.67         3
           0       1.00      1.00      1.00         3
           1       0.75      0.75      0.75         4

    accuracy                           0.80        10
   macro avg       0.81      0.81      0.81        10
weighted avg       0.80      0.80      0.80        10
```

Figure 5: The accuracy score for our proposed model

VI. CONCLUSION

We have seen various steps used to perform sentiment analysis in this paper. Our main focus was to tackle the data from real time twitter to obtain important information regarding public opinion and sentiments. The proposed system classified the tweets into positive, negative, and neutral to ease the process of analysing sentiments with improved accuracy. This work can be handy and useful to the businessmen, marketing groups and even to the government towards crucial decision making.

REFERENCES

1. V. Kharde and S. Sonawane, "Sentiment Analysis of Twitter Data: A Survey of Techniques", International Journal of Computer Applications, vol. 139, p. 11, 2016.
2. Pitiphat Santidhanyaroj, Talha Ahmad Khan, "A Sentiment Analysis Prototype System For Social Network
3. Data," Faculty of Engineering and Applied Science University of Regina Regina, Saskatchewan, Canada.
4. Subramaniam.G, Ranjitha.M," User Emotion Analysis using Twitter Data", IFET College of Enginerring Cuddalore, India.
5. M.Trupthi, Suresh Pabboju," Sentiment Analysis On Twitter Using Streaming API" 2017 IEEE 7th InternationalAdvance Computing Conference
6. Po-Wei Liang, Bi-Ru Dai, "Opinion Mining on Social MediaData", IEEE 14th International Conference on Mobile Data
7. Kunpeng Zhang, Yu Cheng, Yusheng Xie, Daniel Honbo, Ankit Agrawal, Diana Palsetia , Kathy Lee , Wei-keng Liao , Alok Choudhary, SES: Sentiment Elicitation System for Social Media Data, Proceedings of the 2011 IEEE 11th International Conference on Data Mining Workshops, p.129-136, December 11-11, 2011.
8. Apoorv Agarwal, Boyi Xie, Ilia Vovsha, Owen Rambow, and Rebecca Passonneau. 2011. Sentiment analysis of twitter data. In Proceedings of the Workshop on Languages in Social Media, pages 30–38. Association for Computational Linguistics.
9. Prerna Mishra, Sentiment Analysis of Twitter Data:Case Study on Digital India, proceedings of the International Conference on Information Technology (InCITe), 2016.
10. M.Lovelin Ponn Felciah,R.Anbuselvi, "A Study onSentiment Analysis of Social Media Reviews," IEEESecond Conference on Innovations in Information Embedded & Communication Systems, 2015
11. Anant Arora, Chinmay Patil, Stevina Correia, "Opinion Mining: An Overview," International Journal of Advance Research in Computer and Communication Engineering, 2015.
12. Prerna Awasthi, Ranjana Rajnish, "Systematic Study on approach & tools used for opinion mining,". https://www.scribd.com/doc/315237279/Opinion-Mining
13. B.Pang and L.Lee, "Opinion Mining and Sentiment Analysis," The Essence of Knowledge, 2008.
14. Erik Cambaria, Bjourn Schuller, Yunqing Xia, Catherine Havasi, "New Avenues in Opinion Mining and Sentiment Analysis," IEEE, 2013
15. Megha Rathi, Aditya Malik, Daksh Varshney, Rachita Sharma, Sarthak Mendiratta, "Sentiment Analysis of Tweets using
16. Machine Learning Approach", Jaypee Institute of Information Technology JIIT Sec-62 Noida, India.
17. https://en.wikipedia.org/wiki/Twitter