Sentiment Research on Twitter Data

Anchal Kathuria, *Research Scholar, MMU, Haryana, India.*

Dr. Avinash Sharma, *Professor, Department of Computer Science, MMU, Haryana, India*

Article Info
Volume 83
Page Number: 1046 - 1050
Publication Issue:
March - April 2020

Article History
***Article Received**: 24 July 2019*
***Revised**: 12 September 2019*
***Accepted**: 15 February 2020*
***Publication**: 14 March 2020*

Abstract:
The size of informal organization information that is being produced is expanding exponentially step by step. Open and private assessment of different subjects or issues are communicated in web-based social networking. Opinion investigation is a strategy for examining the feeling of an explanation that it typifies. Twitter is one of the social Medias that is picking up prevalence these days and the vast majority are utilizing this stage to communicate their sentiments. Notion examination on Twitter is an utilization of breaking down the estimation of twitter information (tweets) passed on by the client. The exploration on this issue articulation has developed reliably. The fundamental explanation for this is the difficult organization of tweets that are posted, and it makes the handling troublesome. The tweet configuration would be the quantity of characters, slangs, shortened forms, emoticons, http connects, etc. Right now mean to depict the philosophies embraced, the procedure and models applied, alongside a summed up approach utilizing python. Slant examination intends to decide or quantify the disposition of the essayist regarding some subject.

***Keywords**: P Sentiment Analysis, tweet, twitter, sentiment, social media.*

I. INTRODUCTION

The age of net has remodeled the approach individual’s specific their views. It’s currently done over web log posts, on-line discussion forums, product review websites etc. Socialmediaplatforms(alsocalledmicrobloggingsites)may be a media wherever individuals specific their opinions on onethingorsomebody.Twitteristhatthemosttypicallyused microblogging web site by the individuals to post their opinions [1]. Organizations need a tool that helps in analyzing the feedback given by the individuals on their product or service, and this analysis are often done by police investigationthefeelingoftheposts.Thefeelingispaintedas positive, negative, that square measure sub-categorized as powerfully positive, sapless positive, powerfully negative sapless negative, and neutral. It will analyze emotions towards entities like product, services, organizations,people, issues, events, topics and theirattitudes.

II. EXISTINGSYSTEM

The existing system „Sentiment Analysis‟ takes the static datawhichisalreadyextractedfromasocialmediaplatform. ThedataextractedisstoredinacsvfileorExcelfilewhichis the input to the program or application. For each statement the program analyses, the output would be afloating-pointnumber which is termed as polarity. The polarity values range from -1 to +1. Based on the polarity obtained the program determines the emotion of the statement.

- The emotion is classified as positive, negative,neutral.
- If polarity>0 then the emotion ispositive.
- If polarity= 0 then the emotion isneutral.

- If polarity<0 then the emotion isnegative.

Drawbacks:

- The user who is analyzing the statements has to go through the entire document (csv file) to get overall general report.
- The emotions are classified only into three categories i.e., positive, negative,neutral.
- The data is stored prior to theanalysis.

III.PROPOSEDSYSTEM

This framework manages performing capacities progressively through an online internet based life i.e., twitter. Twitter posts of electronic items makes a dataset. Tweets are short messages with slang words and incorrect spellings. Thus, the sentence level assumption examination is performed. This should be possible in seven stages. In the primary stage, input information is given. Here the information alludes to a username or a hashtag. At that point, the quantity of tweets to be dissected are indicated. Those tweets are recovered from the twitter database. At that point in the third stage, the recovered twitter information is put away in a database. In fourth stage, the tweet is prepared. This progression is performed before include extraction. Preparing steps incorporate expelling URLs, evacuating stop-words, evading mis-spellings and slang words. Misspellings square measure evaded by replacement ceaseless characters with 2 events. Slang words contribute rich to the sentiment of a tweet. Subsequently, a slang word dictionary is kept up to switch slang words happening in tweets with their related implications. Next part is highlight extraction. A component vector is shaped abuse applicable choices.

$$P(s|M) = \frac{P(s) \cdot P(M|s)}{P(M)}$$

Overcome The Drawbacks Of Existing System:

- It would be better if the result is represented in theform of bar graph or pie chart.
- To get a better understanding on the emotions, the emotionsshouldbeclassifiedintosevencategories.Theyare:
  - StronglyPositive
  - Positive
  - Weaklypositive
  - Neutral
  - StronglyNegative
  - Negative
  - WeaklyNegative
- Insteadofstoringthedatapriortotheanalysis,it canget realtimedatafromtwitterbygivingahashtagoruser nameto analyze the tweets of a person or a specifiedhashtag

The proposed system overcomes the drawbacks of the prevailing system

Now, it has fell upon our coaching set then, it's required to extract helpful options from it which might be utilized in the method of classification. however, 1st let‟s discuss some text format techniques which is able to aid America in featureextraction:

- **Tokenization:** it's the method of breaking a stream of text into words, symbols and alternative substantive parts referred to as tokens. they will be separated by whitespace characters and/or punctuation characters. Tokenization is performed so it will examine tokens as individual parts that structure atweet.
- **URLs**anduserreferences(identifiedbytokenshttpand @) are removed if the user is interested in only analyzingthe text of thetweet.
- Punctuationmarksanddigits/numeralsmayberemoved ifforinstancetheuserwishestocomparethetweetto alistof English words.
- **Lowercase Conversion:** Tweet may be normalizedby converting it to lowercase which makes its comparison with an English dictionary easier.
- **Stop-words removal:** Stop words area unit a category of some very common words that

embrace nosupplementary data once employed in a text and area unit so claimed to be useless. Examples comprise “a”, “an”, “the”, “he”, “she”, “by”,“on”,“there”,“here”etc.it'stypicallyconvenienttoget ridofthesewordsasaresultoftheyholdnofurtherdatasince they're used virtually equally altogether categories of text, as an example, once computing prior-sentiment-polarity of words in a very tweet per their frequency of incidence in several categories and victimization this polarity to calculate the typical sentiment of the tweet over the set of words employed in thattweet.

Finallyusingdifferentclassifiers,tweetsareclassifiedinto strongly positive, weakly positive, neutral, strongly negative and weakly negative classes. Based on the number of tweets in each class, the final sentiment isderived.

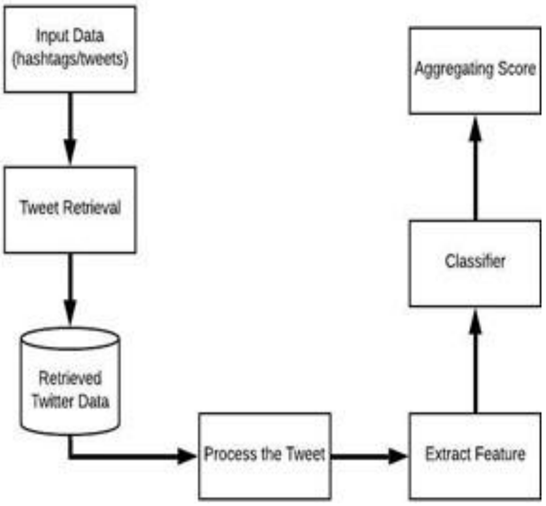

Figure 1: System Architecture

IV.IMPLEMENTATION

Input: Give the input either the username or the hashtags and the no of tweets that you want to analyse the tweets.

Step 1: Tweet cursor from the tweetpy package will retrieve the tweets related to the given search word.

Step 2: The Retrieved tweets will undergo tokenization and then the process of cleaning where the punctuation marks, emoticons, URLs, stop- words will be removed.

Step 3: The output of the above process are the features where they have sent for the analysis to the Naïve Bayes classifier.

Step 4: Now the Classifier will be processing the features andthenassigningthepolaritiestoeachofthemrangingfrom -1 to 1.

Step5:Nowaddingeachpolarityofthegivencertaintweet net polarity will be used to classify the given tweet into its respectivecategory.

Output: The output will be the classification of % of tweets in the particular category and the pie chart which will depict all the categories that have been classified.

4.1. *Building the Classifier:*

The classifier used is Naïve Bayes Classifier. The Bayes Theorem tells that

Where s is Sentiment and M is the Messages.

Here the dataset with videogames reviews is used. Filtering all the reviews by scores and dividing the examples equally between positive class (score = 1) and negative class (score = -1).

4.2. *Preparing Data:*

Then will be preparing the data by tokenization, cleaning of the data and then move to Bag-of - Words. Generate the featurevectorforeachofthedocument.BagofWordswillbe countingthefrequencyofnumberoftimesthetokenhasbeen appeared in eachdocument.

No of columns is tokens that are unique in collection of documents.

No of rows is total documents in whole collection

Now converting the X_train data into a vector called tf_train and X_test data into a vector called tf_test

4.3. *Building:*

Naive Bayes approach is based on Bayes‟ theorem which uses probabilistic learning function. Here, Multinomial approach is used.

P is sum of all feature vectors with score-1P=∑ $tf\_[ytarin= 1]$ +1

Q issumofallfeaturevectorswithscore-1Q=∑ $tf\_[ytarin= -1]$ +1

Here 1 is added to both P, Q to ensure that each token has been taken at least once. ***log-count ratio r:***

r = log ((P/ ∑ P)/(Q/∑Q)) And b:

b=(log$lengt$(P))$lengt$ (Q)

Now coefficients calculated then will produce predictions on testing set. A linear classifier is fit, the linear equation is:

y = mx + b

pre_preds= tf_test. T + b

This is a "naïve" method as it assumes that features are independent, that is they will not interact. Also, the assumption made by BOW that order of the tokens does not matter. This method achieves good results.

V.RESULTS

The result is going to be shown within the variety of apie- chart in Figure a pair of, that constitutes seven major emotions: powerfully positive, sapless positive, positive, neutral, negative, sapless negative and powerfully negative. For neutral, it represents that the tweet / hashtag‟s aggregate score is that of zero. However, this project will list desired variety of recent tweets as such by the tipuser.

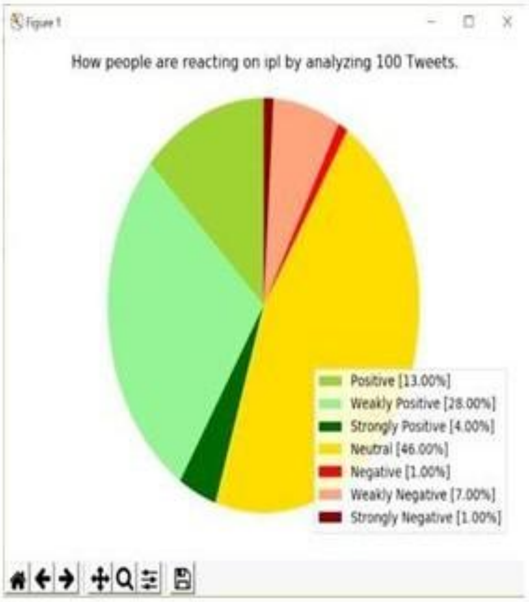

Figure 2: Pie graph of the analyzed tweets

VI.CONCLUSION

Twitter sentiment analysis is developed to investigate public‟s views towards a tweet / hashtag. Input is given i.e., either the username or a hashtag. Then the tweet is retrieved from twitter information that undergoes feature extraction. Associate in Nursing economical feature vector is formed by doing feature extraction in 2 steps when correct pre-processing. within the start, twitter specific options area unit extracted and additional to the feature vector. After that, these options area unit aloof from tweets and once more featureextractionisfinishedasifit'sdoneontraditionaltext. These options are additional to the feature vector. Classification accuracy of the feature vector is tested victimisation Naïve Thomas Bayes classifier. Associate in Nursing accuracy of seventy-eight.38 it had beenreached

REFERENCES

1. Hamid Bagheri, MdJohirulIslam,"Sentiment analysis of twitter data", Annual International Conference "Dialogue"(2017) (pp.14-28)
2. David Zimbra, M. Ghiassi and Sean Lee, "Brand- connected Twitter Sentiment Analysis

mistreatment FeatureEngineeringandthereforetheDynamicdesignfor Artificial Neural Networks", IEEE 1530-1605,2016.

3. BhumikaGupta,MonikaNegi,KanikaVishwa,"Study of Twitter Sentiment Analysis mistreatment Machine Learning Algorithms on Python" International Journal of laptop Applications 0975-8887,2017
4. Aliza Sarlan, ChayanitNadam and ShuibBasri, "Twitter Sentiment Analysis", 2014 International Conference on data Technology and Multimedia (ICIMU), Putrajaya, Malaya Gregorian calendar month eighteen – twenty, 2014.
5. Alexander Pak, Apostle Paroubek, "Twitter as a Corpus for Sentiment Analysis and Opinion Mining", Proceedings of the International Conference on Language Resources and analysis,
6. LREC 2010, 17-23 could 2010, Valletta,Malta,2010
7. Mining Twitter information with Python (Parthalf-dozen – Sentiment Analysis Basics), marcobonzanini 2015, https://marcobonzanini.com/2015/05/17/mining-twitter-data- link "https://marcobonzanini.com/2015/05/17/mining-twitter-data-with-python-part-6-sentiment-analysis-basics/" with-python-part-6-sentiment-analysis-basics/
8. Naive Bayes for Sentiment Analysis, Medium Corporation,2018, https://medium.com/@martinpella/naive-link"https://medium.com/@martinpella/naive-bayes-for-sentiment-analysis-49b37db18bf8" bayes-for-sentiment-analysis-49b37db18bf8