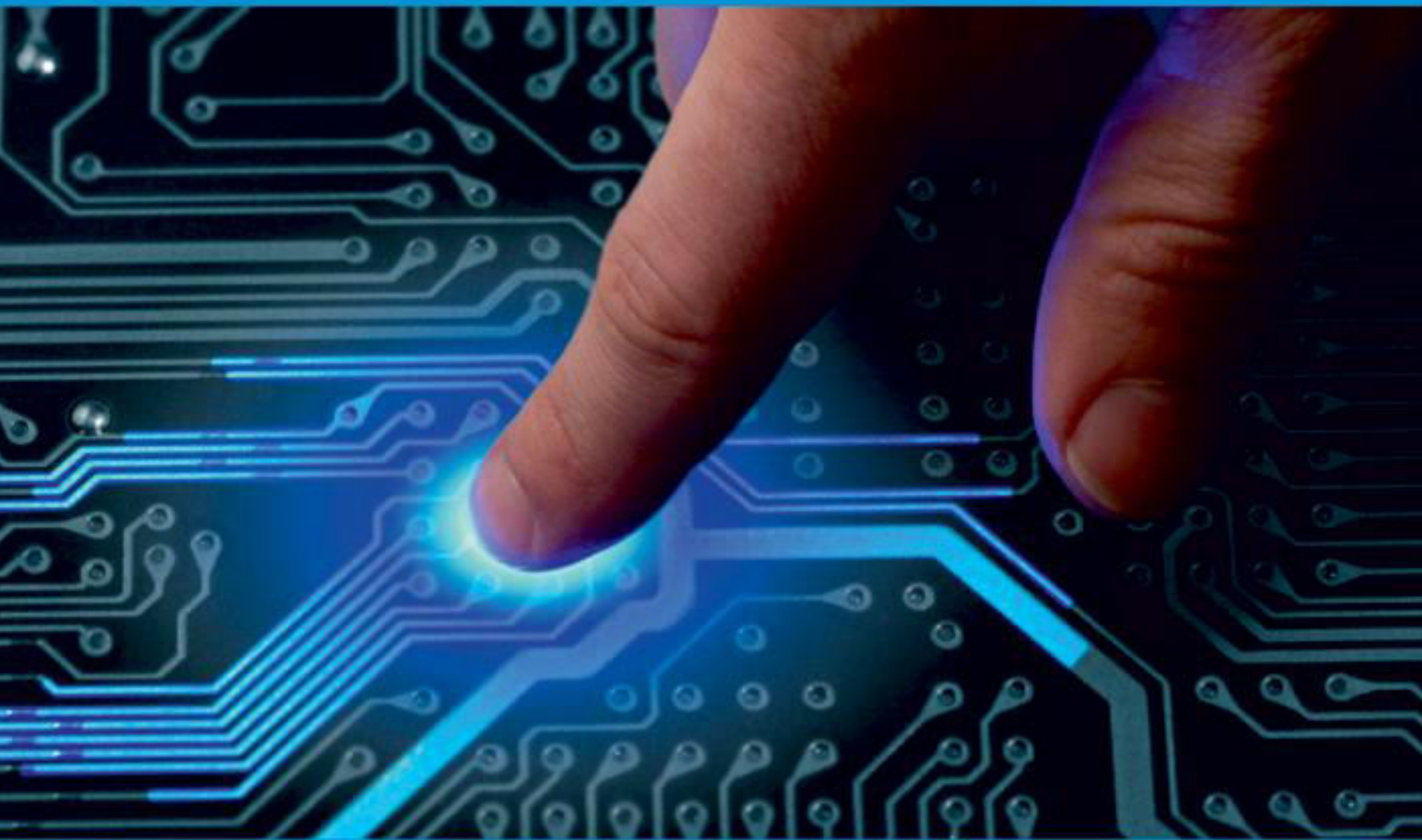
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 11, Issue 4, April 2023

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.379

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 4, April 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1104007 |

Sentimental Analysis on Twitter

Dr.Ch. Anusha[1], K. Sree Swetha[2], D. Anusha[3], Ch. Satish[4], K. Venkat[5]

Associate Professor, Department of Information Technology, Kallam Haranadhareddy Institute of Technology, Chowdavaram, Guntur, Andhra Pradesh, India[1]

IV B.Tech, Department of Information Technology, Kallam Haranadhareddy Institute of Technology, Chowdavaram, Guntur, Andhra Pradesh, India[2,3,4,5]

ABSTRACT The computational handling of views, feelings and subjectivity in the text is known as Sentimental Analysis. The operation of computationally recognizing and streamlining opinions expounded in a piece of text, particularly to choose whether the author's attitude towards a peculiar content, product, etc. is positive, negative, or neutral. With the advancement of the social networking epoch and its growth, the Internet has set an auspicious platform for online literacy, swapping ideas and sharing opinions. A huge quantity of sentiment data exists in social media in the form/aspect of tweets, blogs, and updates on the status, posts, etc. In this paper, the most liked microblogging platform Twitter is used. Twitter sentiment analysis is an appeal of sentiment analysis on data from Twitter, to extract users' opinions and sentiments. The main aim is to survey how text analysis techniques can be used to dig into some of the data in a series of posts exciting on distinctive trends of tweets languages and tweets volumes on Twitter. Experimental evaluations display that the proposed machine learning classifiers are well systematized and perform/execute better in terms of accuracy and classify the tweets into positive, strongly positive, weakly positive, negative, strongly negative, weakly negative and neutral..

KEYWORDS: Machine Learning, Python, Sentimental Analysis, Natural Language Processing (NLP).

I. INTRODUCTION

People make judgments/sentences around them like positive and negative attitudes about people, products, places and events. Attitudes of these types are considered/counted sentiments. The study of automated techniques for extracting sentiments from written languages is called Sentiment analysis. The growth of social media has redounded in a blowup of intimately available, user-generated text on the WWW.

Blogs and comment sections on Facebook and Twitter are considered social media which can capture millions of peoples' views or words. For an enterprise, it is now becoming an important source of information. With the entire world through social media, people are happy to share the facts about their lives, experiences/gests, knowledge, and thoughts/studies.

By expressing their opinions and stating their views in the form of comments that take place in society they actively/laboriously participate in events. This way of sharing/participating drives businesses/firms to collect more information/data about their companies, and products and thereby take decisions to go on with their businesses effectively. This sentiment analysis is a crucial element of leading/directing innovative/creative Customer/Client Experience Management and Customer/Client Relationship Marketing concentrated enterprises/businesses. For numerous enterprises, information/data has become the main trading thing. This paper reveals an approach/way which is implemented/enforced as a tool that can analyze/dissect sentiments on Twitter social media and then develop an application to generate/induce knowledge that can be used for business/custom surroundings/environments using people's attitudes about their products and services.

II. PROPOSED METHOD

The project would depend on and count on techniques of "Natural Language Processing" in extracting significant patterns and features from the large dataset of tweets and on "Machine Learning" techniques for directly/accurately assaying/analyzing individual unlabeled tweets according to whichever pattern model best describes them.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 4, April 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1104007 |

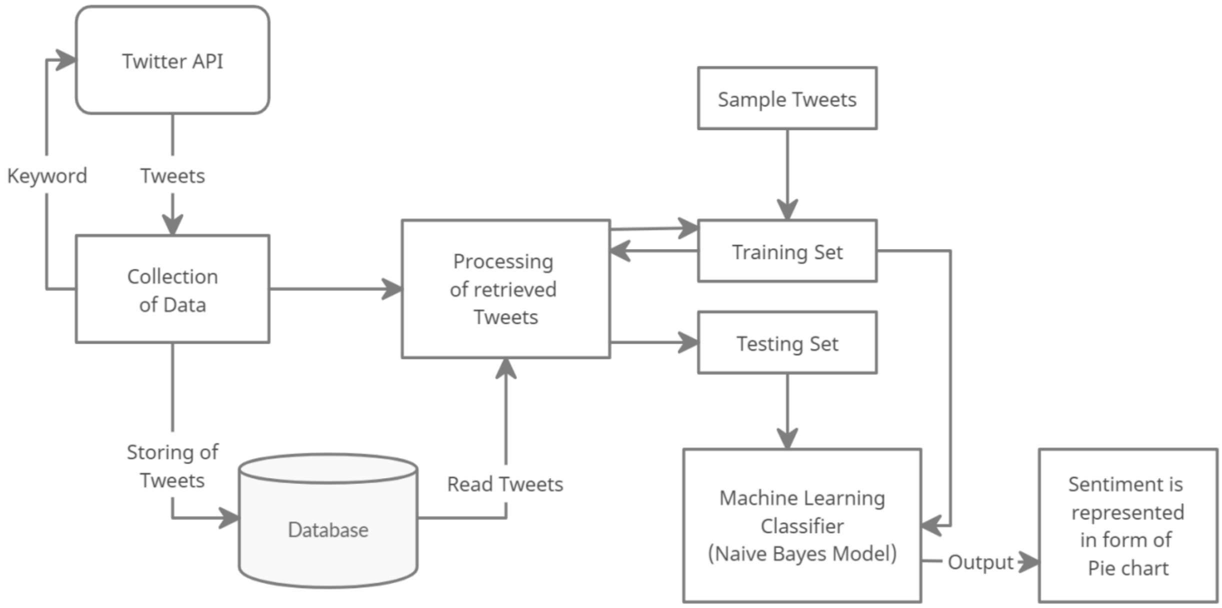


Figure 1 Proposed System Architecture

A. Extraction of Tweets

To extract the tweets we need to access the Twitter API because this Twitter API gives developers access to most of Twitter's functionality. You can use the API for access to read and write information related to Twitter entities such as tweets, users, and trends. An app is created in the developer's Twitter account and we generate the keys and tokens and these keys and tokens are used to access the Twitter API. Now we can draw out the tweets from Twitter.

B. Preprocessing of Tweets

Twitter Due to misspellings and slang words keyword extraction is tough. So to avoid this, a preprocessing step is performed ahead of feature extraction. Preprocessing steps include removing URLs and keeping away from misspellings and slang words. Misspellings are avoided by returning repeated characters with 2 occurrences. The emotion of a tweet is given much by slang words. So they can't be simply removed. Therefore a slang word dictionary is maintained to replace slang words occurring in tweets with their related meanings. Domain information accords much to the formation of slang word dictionaries.

C. Creation of Feature Vector

Feature extraction is done in two steps. In step one, Twitter-specific features are extracted. Hashtags and emoticons are relevant Twitter-specific features. Emoticons can be negative or positive. So they are given different weights. Positive emoticons are given a weight of '1' and negative emoticons are given a weight of '-1'. There may be positive and negative hashtags. Therefore the count of negative and positive hashtags are combined as two separate features in the feature vector.

Twitter-specific features may not be available in all tweets. So further feature extraction is to be done to get other features. After extracting Twitter-specific features, they are excluded from the tweets. Tweets can be then considered/taken as simple text. Using the unigram approach, then tweets are represented/shown as a collection of words. In unigrams, a tweet is represented/shown by its keywords. We maintain a negative keyword to represent negation. Counts of negative and positive keywords in tweets are applied as two distinct features in the feature vector. The presence of negation contributes much to the sentiment, so their presence is also included as a relevant feature.

All keywords cannot be treated equally in the presence of multiple negative and positive keywords. Therefore from all the tweets, a special keyword is selected. A search is done to identify a keyword having a relevant part of speech if the tweets have only negative keywords or only positive keywords. A relevant part of speech is an adjective,

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 4, April 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1104007 |

verb or adverb. Such a relevant part of speech is defined/[explained] based on its relevance in determining sentiment.

Keywords that are adjectives, adverbs or verb show more emotion than others. If a relevant part of speech can be determined for a keyword, then that is taken as a special keyword. Otherwise, a keyword is selected randomly from the available keywords as a special keyword. We select any keyword having a relevant part of speech if both negative and positive keywords are present in a tweet. If a relevant part of speech is present for both negative and positive keywords, none of them is chosen. The special keyword feature is given a weight of '-1' if it is negative and '1' if it is positive and '0' in its absence. Part of the speech feature is given a value/weight of '1' if it is relevant and '0' otherwise.

Thus the feature vector is composed/consists of 8 relevant features; they are the presence of negation, emoticon, part of speech (pos) tag, special keyword, number of positive keywords, number of positive hashtags, number of negative keywords and number of negative hashtags.

D. Sentiment Classification

Later generating a feature vector, classification is done using the Naive Bayes classifier, normally used for text classification which can be also used for Twitter sentiment classification.

Naive Bayes Classifier assembles the use of all the features in the feature vector and analyzes them individually as they are uniformly independent of each other. The conditional probability for Naive Bayes can be explained as

$$P(X|y_j) = \Pi^{m}_{i=1}P(x_i|y_j)$$

'X' is the feature vector explained as X=*{x*1,*x*2,....*xm}* and *yj* is the class label. There are different dependent features like emotional keywords, emotions count/number of negative and positive keywords, and count/number of negative and positive hashtags which are successfully utilized by the Naive Bayes classifier for classification. Naive Bayes does not contemplate the relationships between features. So it cannot use the relationships between a part of emotional keyword, speech tag and negation.

In this classification procedure, the TextBlob package is also used. We classify the tweets into, strongly positive, positive, weakly positive, negative, strongly negative, weakly negative and neutral based on the disparity range.

If the disparity of the tweet is equal to 0 then the tweets are treated as neutral, If the disparity of the tweet is greater than 0 and less than or equal to 0.3 then the tweets are treated as weakly positive, If the disparity of the tweet is greater than 0.3 and less than or equal to 0.6 then the tweets are treated as positive If the disparity of the tweet is greater than 0.6 and less than or equal to 1 then the tweets are treated as strongly positive If the disparity of the tweet is greater than -0.3 and less than or equal to 0 then the tweets are treated as weakly negative, If the disparity of the tweet is greater than -0.6 and less than or equal to -0.3 then the tweets are treated as negative, If the disparity of the tweet is greater than -1 and less than or equal to -0.6 then the tweets are treated as strongly negative.

III. RESULT

We will first present our results for the subjective/objective and negative/positive classifications. These results act as the primary step of our classification approach. We only utilize the short-listed features for both of these outcomes. This process that for the subjective/objective classification we have five features and for negative/positive classification we have three features. For both of these outcomes, we utilize the Naïve Bayes classification algorithm, because that is the algorithm we are engaging in our actual classification approach at the first step. Likewise, all the figures reported/described are the conclusion/result of 10-fold cross-validation. We take an average of each of the ten values we get from the cross-validation.

We build a condition while reporting the outcomes of disparity classification (which differentiates between negative and positive classes) that only subjective labelled tweets are used to calculate these outcomes. However, in the case of the last classification approach, any such condition is detached and both objectivity and disparity classifications are used for all tweets regardless of whether they are labelled subjective or objective.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 4, April 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1104007 |

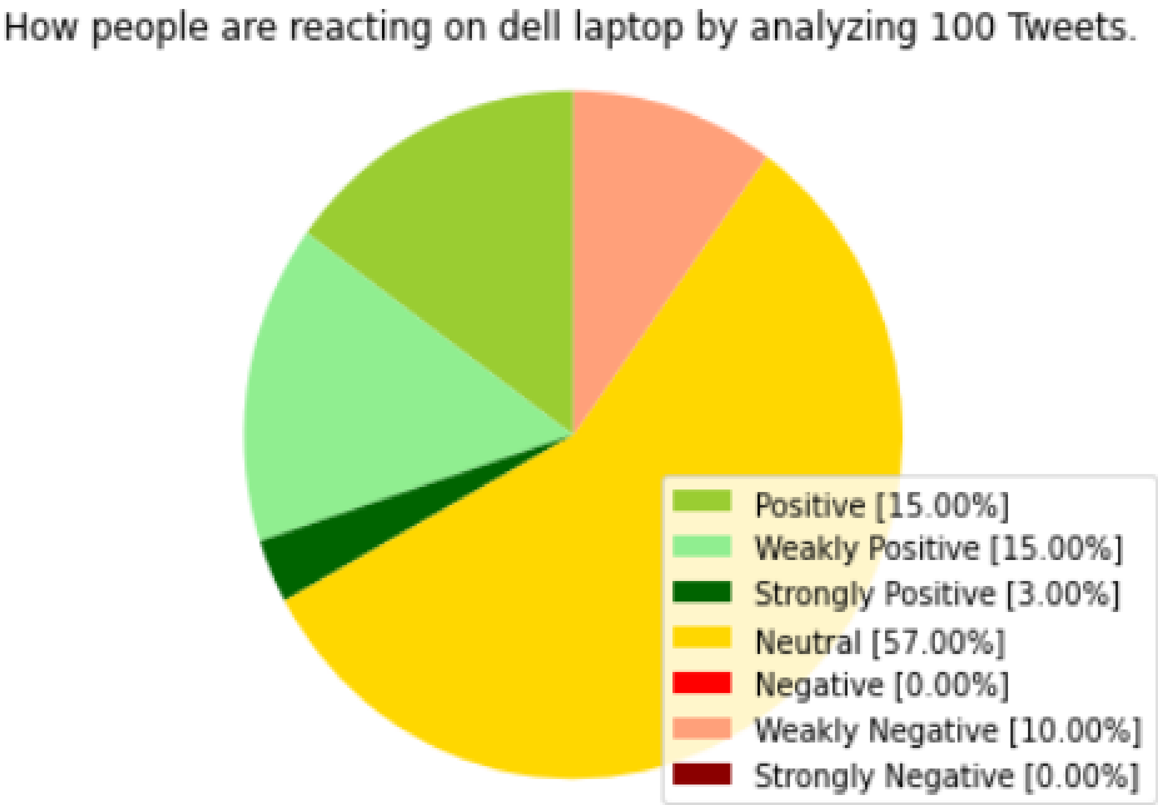


Figure 2 Sentiment in the Pie chart

IV. CONCLUSION

Twitter sentiment analysis is developed to analyze people's reactions towards a particular place, people, product, events etc. In this Naive Bayes machine learning classifier is handed down which is more accurate for analyzing a sentiment; at the same instant natural language processing techniques will be used. As an outcome, the model will categorize sentiment into strongly positive, positive, weakly positive, strongly negative, negative, weakly negative and neutral which is represented in a pie chart.

REFERENCES

[1] EfthymiosKouloumpis and Johanna Moore, IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 4, No 3, July 2012
[2] S. Batra and D. Rao, "Entity Based Sentiment Analysis on Twitter", Stanford University,2010
[3] Saif M.Mohammad and Xiaodan zhu, Sentiment Analysis on of social media texts: 2014
[4] Ekaterina kochmar, University of Cambridge, at the Cambridge coding Academy Data Science.2016
[5] Manju Venugopalan and Deepa Gupta, Exploring Sentiment Analysis on Twitter Data, IEEE 2015
[6] Brett Duncan and Yanqing Zhang, Neural Networks for Sentiment Analysis on Twitter.2017
[7] Afroze Ibrahim Baqapuri, Twitter Sentiment Analysis: The Good the Bad and the OMG!, Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media.2011

INNO SPACE

SJIF Scientific Journal Impact Factor

Impact Factor: 8.379

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

Scan to save the contact details