iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** XI **Month of publication:** November 2022

DOI: https://doi.org/10.22214/ijraset.2022.47524

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue XI Nov 2022- Available at www.ijraset.com

Twitter Analysis On Real Time Data

Prof. Ravishankar Sir[1], Prathamesh P. Bhamanage[2], Pratik S. Patil[3], Suraj Y. Ahire[4], Durgesh P. Jadhav[5]
[1]Professor, [2, 3, 4, 5]Students, Department of Computer Engineering, SKN Sinhgad Institute of Technology and Science, Kusgaon (BK), Pune

Abstract: ***Sentiment analysis is mainly concerned with identifying and classifying opinions or emotions that are expressed within a text. These days, sharing opinions and expressing emotions through social networking websites has become very common. This paper presents an idea of extracting sentiments out of the tweet and an approach towards classifying a tweet into positive, negative or neutral. This approach can be in many ways useful to any organization, who gets mentioned or tagged in a tweet. Generally the tweets being unstructured in format, first of all the tweet needs to be converted into the structured format. In this paper, tweets are resolved using pre-processing phase and access of tweets has been accomplished via libraries using Twitter API.. We provide additional comparisons and extract alternatives. exams, apprenticeships, etc., are compared to find higher overall performance, and several scoring criteria have been developed for different techniques.***
Keywords: Twitter API, Data mining, Sentiment analysis, Supervised learning, Natural Language Processing, SVM.

I. INTRODUCTION

The age of the internet has changed the way people express their ideas. With the ever-increasing popularity of social networking, microblogging and blogging websites, a large amount of data is generated every day. These social networking websites depend largely on the user-generated content. Typically, when people intend to purchase a product, they browse through a lot of websites to gain some information about the products before they make their purchase. They take into consideration the available reviews and ratings of these products on these websites before making purchases. The amount of information is inordinate for a normal person to analyze it using naive techniques.

Thus, in order to make this process efficient and to automate it, several sentiment analysis techniques are used. Symbolic techniques or knowledge-based approach and machine learning techniques are usually used to develop such models. The objective of feeling investigation is recognizing content assumption extremity. Assumption examination could be taken as an order issue. Sentiment analysis is a process that isolates the content into positive, negative or neutral conclusion. Profound neural system and the Gaussian blend model is one of the hearty models for normal preparing language. Conviction analysis is useful for consumers who are trying to research a product or service, or marketers researching public opinion of their company/product. However, doing the analysis of tweets that express human emotions isn't an easy job. A lot of challenges are involved in terms of tonality, polarity, lexicon and grammar of the Twitter sentiment analysis using bag of words tweets to avoid this circumstance and to increase the revenue of the college, the proposed research is undertaken.

II. LITERATURE REVIEW

1) Most of the authors and expertise in this field state and discuss those opinions and attitude expressed in social media platform is growing in abundance nowadays and people decide everything based on these views fellow users make. This leads to the creation of enormous amount of data. Gleaning. information from huge storage of data is a big challenge for the companies nowadays. Here comes the area for analyzing data mostly on popular platforms such as twitter. There are several methods used in these cases recognizing various scenarios. Prototype of this tool is demonstrated by analyzing matter posted by student on Twitter is demonstrated. It used the Sentiment analysis tools: R and Rapid miner. Data is collected by crawling through Twitter API with R. Then, Rapid miner is used in data pre-processing, classification, for classifiers performance evaluation and Net Sentiment Score is calculated which associates with customer satisfaction using classification results. Tweets are collected and all unimportant words are deleted from the tweet collection to support the classification process, then tweets are filtered by means of a Bayes Naive classifier, which was earlier trained and whose intention is to select messages that represent news about fresh cyber-attacks and malware. Sentiment was measured which was expressed in the form of tweets posted from a specific region using the fast Text model. Then a simple clustering technique called latent Dirichlet allocation (LDA) was applied to recognize the topics that were gathered in every area using sentimental analysis.

©IJRASET: All Rights are Reserved

2) *Data Mining:* There are large amounts of databases available with the increase in Information Technology. Data is present in huge amounts which comprise various fields. For the purpose of future decision makings, the data needs to be stored and manipulated. For this, various databases have been developed and research has been carried out for their management. The process of extraction of useful information and patterns from large amounts of stored data is known as data mining. Data mining is a method of examining large existing databases to generate new information. It is a process of processing large volumes of data (usually data stored in a database). It used to locate relationships between styles and information. Various types of data are being analyzed with the help of certain data mining tools.

III. METHODOLOGY

Twitter goes about as stubborn data bank with enormous measure of information accessible, utilized for conclusion analysis. Twitter is very convenient for research in light of the fact that there are enormous quantities of messages, many of which are freely accessible, and acquiring them is actually basic contrasted with scarping sites from the web. Twitter information is gathered for investigation utilizing Twitter API. Two broadly utilized methodologies utilized for the equivalent are Machine Learning and Dictionary Based methodology. We are utilizing Dictionary Based methodology for dissecting the notions of information posted by various clients. At that point extremity arrangement of this information is done. For example Tweets gathered after examinations are grouped into three classes as Positive, Negative and Neutral.

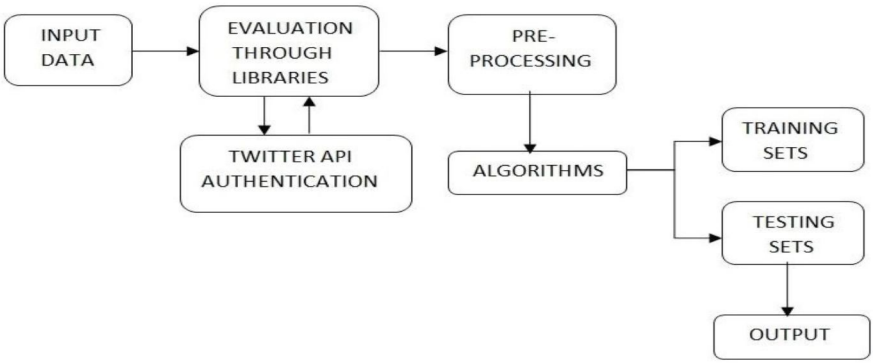


Fig1. Workflow

A. *Techniques of Sentiment Analysis*

The semantic concepts of entities extracted from tweets can be used to measure the overall correlation of a group of entities with a given sentiment polarity. Polarity refers to the most basic form, which is if a text or sentence is positive or negative. However, sentiment analysis has techniques in assigning polarity such as:

1) *Natural Language Processing (NLP):* NLP techniques are based on machine learning and especially statistical learning which uses a general learning algorithm combined with a large sample, a corpus, of data to learn the rules. Sentiment analysis has been handled as a Natural Language Processing denoted NLP, at many levels of granularity. Starting from being a document level classification task, it has been handled at the sentence level and more recently at the phrase level. NLP is a field in computer science which involves making computers derive meaning from human language and input as a way of interacting with the real world.
2) *Support Vector Machine (SVM):* Support Vector Machine is to detect the sentiments of tweets . together with stated SVM is able to extract and analyze to obtain upto70%-81.3% of accuracy on the test set. Collected training data from three different Twitter sentiment detection websites which mainly use some pre-built sentiment lexicons to label each tweet as positive or negative. Using SVM trained from these noisy labeled data, they obtained 81.3% in sentiment classification accuracy.

B. *Application Programming Interface(API)*

1) Alchemy API performs better than the others in terms of the quality and the quantity of the extracted entities. As time passed the Python Twitter Application Programming Interface (API) is created by collected tweets. Python can automatically calculated frequency of messages being retweeted every 100 seconds, sorted the top 200 messages based on there-tweeting frequency, and stored them in the designated database As the Python Twitter API only included Twitter messages for the most recent six days, collected the data needed to be stored in a different database .

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue XI Nov 2022- Available at www.ijraset.com

2) H. Python: Python was found by Guido Van Rossum in Natherland, 1989 which has been public in 1991. Python is a programming language that's available and solves a computer problem which is providing a simple way to write out a solution. mentioned that Python can be called as a scripting language. In addition, here mentioned that Python is a language that is great for writing a prototype because Python is less time consuming and working prototype provided, contrast with other programming languages. Many researchers have been saying that Python is efficient, especially for a complex project, as mentioned that Python is suitable to start up social networks or media steaming projects which most always are a web-based which is driving a big data. It gave the reason that because Python can handle and manage the memory used. Besides Python creates a generator that allows an iterative process of things, one item at a time and allow program to grab source data one item at a time to pass each through the full processing chain.

IV. RESULT AND DISCUSSION

A. Twitter Retrieved

To associate with Twitter API, developer need to agree in terms and conditions of development Twitter platform which has been provided to get an authorization to access a data. The output from this process will be saved in JSON file. The reason is, JSON (JavaScript Object Notation) is a lightweight data-interchange format which is easy for humans to write and read. Moreover, stated that, JSON is simple for machines to generate and parse. JSON is a text format that is totally language independent, but uses a convention that is known to programmers of the C-family of languages, including Python and many others. However, outputs size depends on the time for retrieving tweets from Twitter. Nevertheless, the output will be categorized into 2 forms, which are encoded and un-encoded. According to security issue for accessing a data, some of the output will be shown in an ID form such as string ID. Sentiment Analysis. The tweets will be assigned the value of each word, together with categorize into positive and negative word, according to lexicon dictionary. The result will be shown in .txt, .csv and html. Keyword extraction is difficult in Twitter due to misspellings and slang words. So to avoid this, a pre-processing step is performed before feature extraction. Pre-processing steps include removing URLs, avoiding misspellings and slang words. Misspellings are avoided by replacing repeated characters with 2 occurrences. Slang words contribute considerably to the emotion of a tweet. So they can't be simply removed. Therefore, a slang word dictionary is maintained to replace slang words occurring in tweets with their associated meanings. Domain information contributes much to the formation of slang word dictionary. Also, we use a technique in which if the overall sentiment of the tweet is obtained, we will be able to find out the sentiment score of the new term by just looking at its relative position in the sentence.

B. Sentiment Analysis

Tweets from JSON file will be assigned the value of each word by matching with the lexicon dictionary. As a limitation of words in the lexicon dictionary which is not able to assign a value to every single word from tweets. However, as a scientific language of python, which is able to analyze a sense of each tweet into positive or negative for getting a result.

C. Information Presented

The result will be shown in a pie chart which is representing a percentage of positive, negative and null sentiment hash tags. For null hash tag is representing the hash tags that were assigned zero value. However, this program is able to list a top ten positive and negative hash tags.

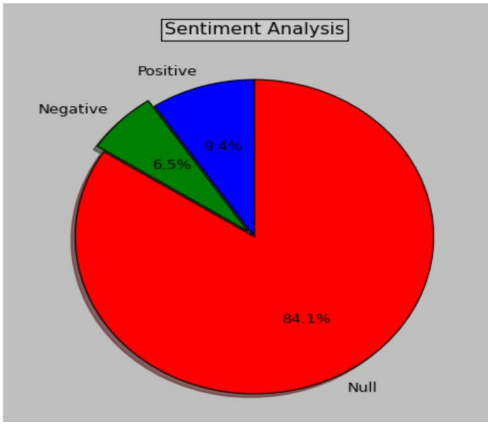


Fig2. Pie Chart

©IJRASET: All Rights are Reserved

D. Data Pre-Processing

Pre-processing is the progression of concentrated effort to the data from redundant elements. it increases the accuracy of the consequences by means of reducing mistakes inside the statistics. Pre-processing of data is one of the most important tasks that must be done before the dataset be able to be used for machine learning. The real-world statistics are incomplete and incompatible. So, it is necessary to be cleaned. Not by means of pre-processing, such as enchantment corrections, may lead the system to disregard important words. Preprocessing and concentrating the effort of data is one of the most important tasks that must be done before the dataset be able to be used for machine learning. The real-world statistics are strident, incomplete, and incompatible. So, it is necessary to be cleaned. It must be done before the dataset is able to be used for machine learning. The real-world statistics are improper and incompatible. So, it is necessary to be cleaned.

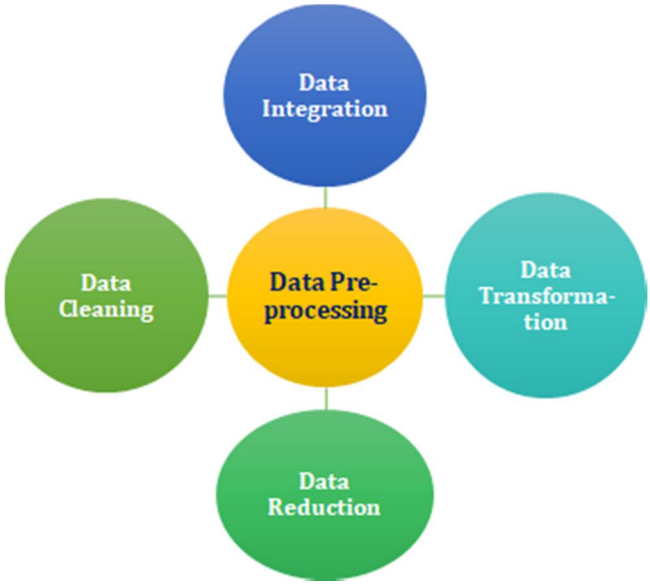


Fig3. Data Pre-Processing

V. CONCLUSION

There are different Symbolic and machine learning techniques to identify sentiments from text. Machine Learning techniques are simpler and efficient than Symbolic techniques. A combination of these two techniques can be used to achieve an accuracy of 100%. In this paper we took the Sanders analytic dataset in order to analyze the tweets. After pre-processing the data we created the feature vector that is used for evaluating Twitter sentiments using Machine Learning techniques. As a result, program will be categorized sentiment into positive and negative, which is represented in a pie chart and html page Although, the program has been planned to be developed as a web application, due to limitation of Django which can only work on Linux server or LAMP. Thus, it cannot be realized. Therefore, further enhancement of this element is recommended in future study. Amongst the various algorithms available, KNN algorithm is used to increase the efficiency of sentiment analysis whereas Naïve Bayes for simple and efficient sentiment analysis by classifying the tweets as either positive, negative or zero. Whenever a tweet is fed for sentiment analysis, it goes through various phases of sentiment analysis. For analyzing a tweet it is very necessary to know the morph and elements of the tweet. Each of these components and phases of sentiment analysis are briefly described in this review paper.

VI. ACKNOWLEDGEMENT

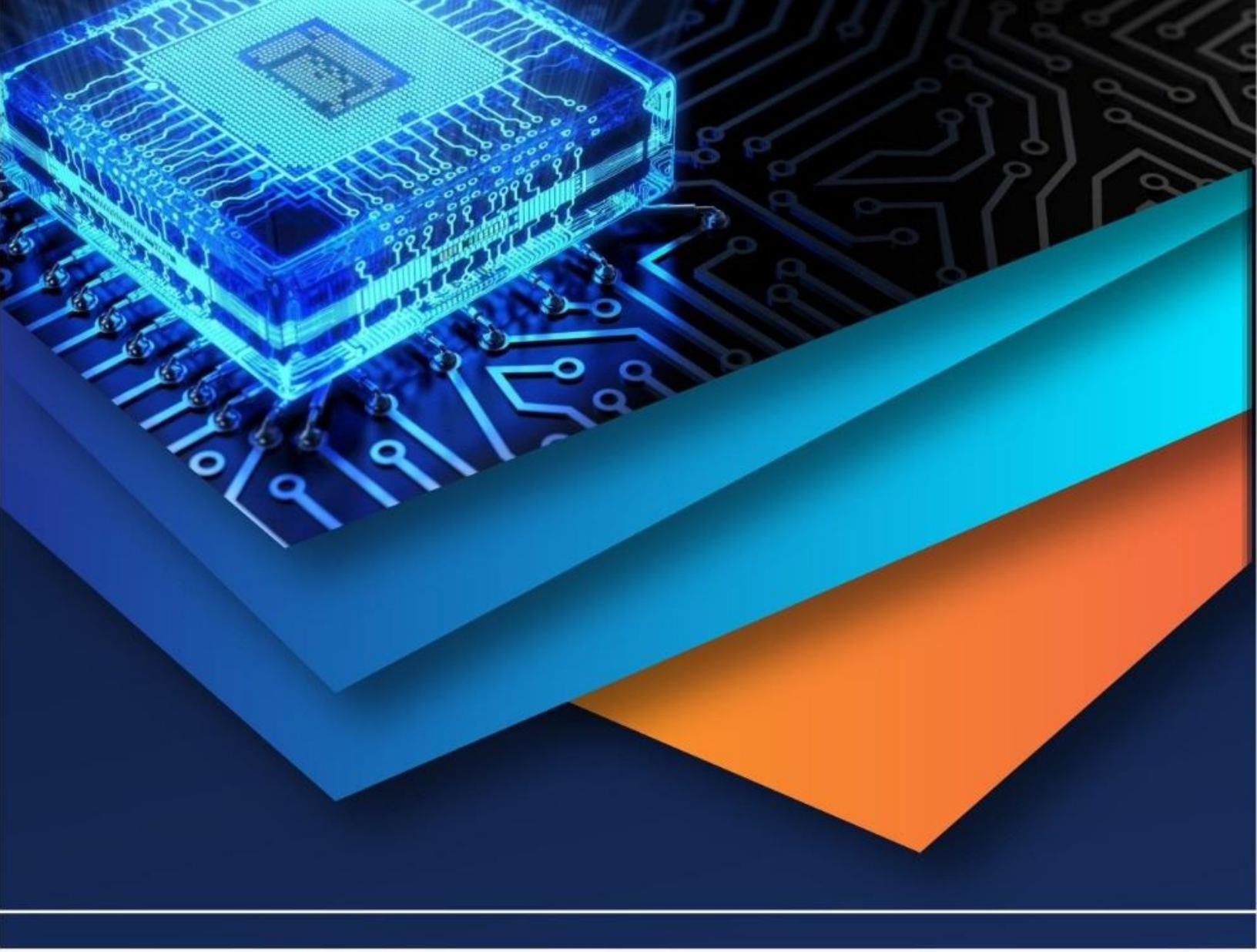
The authors would like to acknowledge the support and guidance provided by management and guides of SKN Sinhgad Institute of Technology and Science, Lonavala for providing the necessary support and guidance in carrying out this work.

REFERENCES

[1] A. Sarlan, C. Nadam and S. Basri, "Twitter sentiment analysis," Proceedings of the 6th International Conference on Information Technology and Multimedia, 2014.
[2] C. Kariya and P. Khodke, "Twitter Sentiment Analysis," 2020 International Conference for Emerging Technology (INCET), 2020..
[3] S. A. El Rahman, F. A. AlOtaibi and W. A. AlShehri, "Sentiment Analysis of Twitter Data," 2019 International Conference on Computer and Information Sciences (ICCIS), 2019.

[4] V. Pandya, A. Somthankar, S. S. Shrivastava and M. Patil, "Twitter Sentiment Analysis using Machine Learning and Deep Learning Techniques," 2021 2nd International Conference on Communication, Computing and Industry 4.0 (C2I4), 2021.
[5] A. Ikram, M. Kumar and G. Munjal, "Twitter Sentiment Analysis using Machine Learning," 2022 12th International Conference on Cloud Computing, Data Science & Engineering (Confluence), 2022.
[6] J. F. Raisa, M. Ulfat, A. Al Mueed and S. M. S. Reza, "A Review on Twitter Sentiment Analysis Approaches," 2021 International Conference on Information and Communication Technology for Sustainable Development (ICICT4SD), 2021.
[7] A. Roy and M. Ojha, "Twitter sentiment analysis using deep learning models," 2020 IEEE 17th India Council International Conference (INDICON), 2020.
[8] R. Wagh and P. Punde, "Survey on Sentiment Analysis using Twitter Dataset," 2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA), 2018.
[9] V. Prakruthi, D. Sindhu and D. S. Anupama Kumar, "Real Time Sentiment Analysis Of Twitter Posts," 2018 3rd International Conference on Computational Systems and Information Technology for Sustainable Solutions (CSITSS), 2018.
[10] N. Yadav, O. Kudale, S. Gupta, A. Rao and A. Shitole, "Twitter Sentiment Analysis Using Machine Learning For Product Evaluation," 2020 International Conference on Inventive Computation Technologies (ICICT), 2020.

©IJRASET: All Rights are Reserved



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)