A Comprehensive Analysis of Real Time Twitter Data to Draw Intelligence for Business and Research Specific Applications

Abhishek Dhillon

G D Goenka Public School, Sector-22, Rohini, New Delhi, India

ABSTRACT

Nowadays, imparting insights and communicating feelings through interpersonal interaction sites has become extremely normal. The primary focus of sentiment analysis is the identification and classification of opinions or feelings expressed in a text. This paper proposes a method for classifying a tweet as positive, negative, or neutral, as well as a method for extracting sentiments from the tweet. Any organization mentioned or tagged in a tweet can benefit greatly from this strategy.

Since tweets typically have an unstructured format, we must first convert them into a structured format. In this paper, tweets are resolved during the pre-processing phase, and the Twitter API is used to access tweets through libraries. We also offer additional comparisons and extract alternatives. Tests, apprenticeships, and so on., are compared to determine higher overall performance, and various scoring criteria for various techniques have been developed.

INTRODUCTION

The internet has changed how people communicate their thoughts. Every day, a significant amount of data is generated due to the ever-increasing popularity of social networking, microblogging, and blogging websites. User-generated content constitutes a significant part of these social networking websites. Before making a purchase, most people typically peruse many websites to learn more about the item in question. Before making a purchase, they consider the ratings and reviews for these products on these websites. It would be impossible for a normal person to use naive methods to analyze an enormous amount of information.

As a result, various sentiment analysis methods are utilized to streamline and automate this procedure. Such models are typically created using symbolic, knowledge-based, and machine-learning techniques. The recognition of content assumption extremeness is the goal of feeling investigation. It is possible to interpret assumption examination as an order issue. The process of separating the content into positive, negative, or neutral conclusions is thought as sentiment analysis. one in every of the solid models for traditional language preparation is that the profound neural system and therefore the mathematician mix model. customers researching a product or service or marketers researching belief of their company or product will like conviction analysis. However, it's a difficult task to research tweets that convey human emotions. The projected analysis is administered to avoid this case and increase the college's revenue as a result of the various difficulties exhibit by the tone, polarity, lexicon, and synchronic linguistics of the Twitter sentiment analysis employing a bag of words tweets.

METHODOLOGY

Twitter is used for conclusion analysis as a persistent data bank with much information available. Twitter is great for research because it has a lot of messages, many of which can be found for free, and it's easy to get them, unlike searching websites on the internet. Utilizing the Twitter API, information about Twitter is gathered for research purposes. Two extensively used techniques used for the identical are AI and Word reference Based system. We dissect the concepts of information posted by various clients using a dictionary-based approach. After that, this

information is thoroughly arranged. Tweets collected from examinations, for instance, are categorized as Positive, Negative, or Neutral.
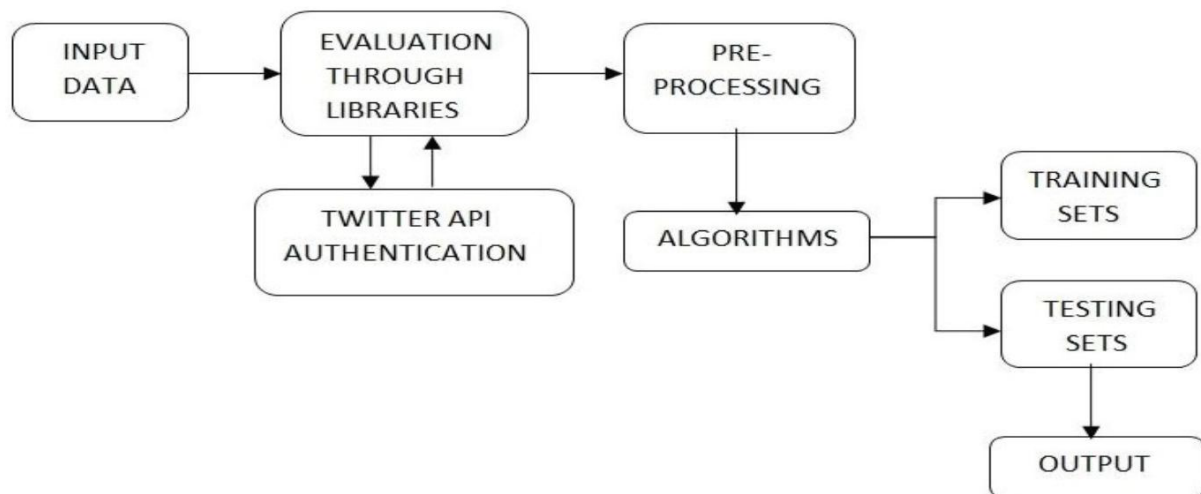

Fig 1: Workflow

A. Procedures of Feeling Investigation

Can utilize the semantic ideas of substances removed from tweets to quantify the general relationship of a gathering of elements with a given opinion extremity. The most basic form of polarity is whether a sentence or text is positive or negative. However, sentiment analysis has methods for determining polarity, including:

NLP is a subfield of computer science that makes computers use human language and input to understand the real world.

1) NLP, or natural language processing: NLP methods are based on statistical and machine learning. A corpus of data and a large sample are used in conjunction with a general learning algorithm to learn the rules. At various granularity levels, sentiment analysis has been handled as a Natural Language Processing (NLP) task. It started as a classification task at the document level and has since been handled at the sentence level and, more recently, at the phrase level.

2) SVM or Support Vector Machine: Using a Support Vector Machine can identify tweet sentiments. Gathered training data from three different Twitter sentiment detection websites, most of which primarily employ pre-built sentiment lexicons to classify each tweet as positive or negative. They were able to achieve an accuracy of 81.3% in sentiment classification with SVM trained on these noisy labelled data. In addition, SVM can extract and analyse data with up to 70.3 per cent to 81.3 per cent accuracy on the test set.

B. Application Programming Interface (API) 1) The Alchemy API outperforms the others in terms of both the quantity and quality of the entities that can extract. The Python Twitter Application Programming Interface (API) was developed over time from a collection of tweets. Because the Python Twitter API only included Twitter messages for the last six days, it collected the data that needed to be stored in a different database. Python can automatically calculate the frequency of messages being retweeted every 100 seconds, sort the top 200 messages based on their frequency, and store them in the designated database.

2) Python H: Guido Van Rossum discovered Python in the Netherlands in 1989, and was made public in 1991. Python is a programming language that can solve a computer problem by making it easy to write a solution down.

Mentioned that Python is a scripting language that can use. Also mentioned is that Python is a great programming language for writing prototypes because it takes less time and comes with a working prototype, unlike other languages. As mentioned, Python is suitable for starting up social networks or media streaming projects, most web-based and driving big data. Numerous researchers have stated that Python is efficient, particularly for complex projects. It explained that Python could handle and control the amount of memory used. In addition, Python develops a generator that enables the program to acquire source data one item at a time and proceed through the entire processing chain, allowing for an iterative process.

RESULT AND DISCUSSION

A. Twitter Retrieved The developer must agree to the terms and conditions of the development Twitter platform that has been provided to obtain authorization to access data to associate with the Twitter API.   will save the JSON file containing the process's output. This is because JSON (JavaScript Object Notation) is a light data exchange format that is simple to write and read by humans. Machines can easily generate and parse JSON. Although JSON is a language-independent text format, programmers of the C family of languages, including Python and many others, are familiar with its convention. However, the time it takes to retrieve tweets from Twitter affects the size of outputs. Nevertheless, the output will be divided into encoded and unencoded categories. Some of the output will be displayed as an ID, such as a string ID, depending on the security issue for accessing the data.

Analysis of Emotions According to the lexicon dictionary, the tweets will be assigned a value for each word and categorized as positive or negative. The output will be available in .txt,.csv, and.html formats. Catchphrase extraction is troublesome on Twitter due to incorrect spellings and shoptalk words. Consequently, before feature extraction, a pre-processing step is carried out to prevent this. Getting rid of URLs, spelling mistakes, and slang words is all part of the pre-processing steps. By substituting two occurrences for frequently occurring characters, misspellings are avoided. Slang terms significantly influence a tweet's emotional content. Therefore, cannot simply remove them. As a result, a slang word dictionary is kept up to replace tweet-associated slang words with their meanings. Area data

contributes a lot to the development of shoptalk word reference. In addition, we employ a method by which, if the tweet's overall sentiment is determined, we can examine the new term's relative position in the sentence to determine its sentiment score.

B. Will use the lexicon dictionary to match the tweets from the JSON file for sentiment analysis to determine the value of each word. as a limitation on the number of words in the lexicon dictionary, which cannot value each tweet word. However, Python, a scientific programming language, can determine whether a tweet is positive or negative to produce a result.

C. Information Provided The outcome will be depicted in a pie chart representing a percentage of sentiment hashtags that are positive, negative, or null. The null hashtag represents the hashtags with zero values. On the other hand, this application can list the top ten positive and negative hashtags.
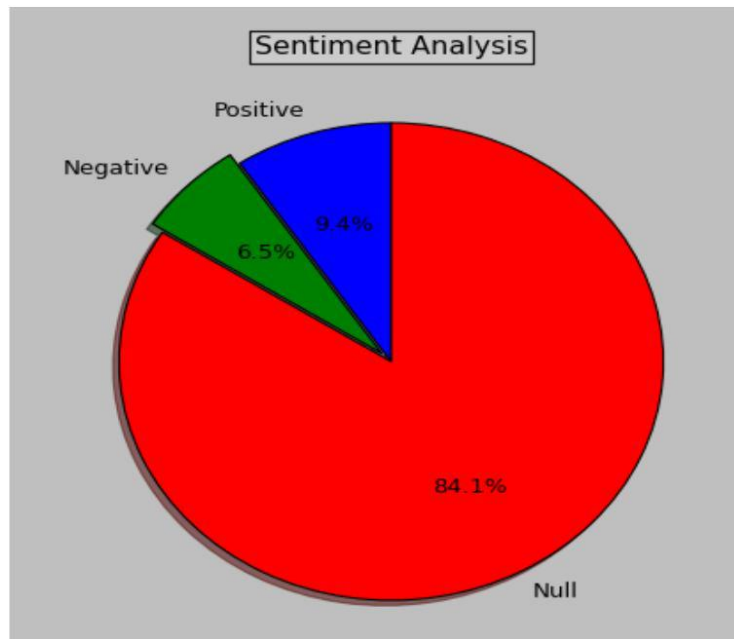
D. Data Pre-Processing, the process of removing redundant elements from the data is known as pre-processing. Reducing errors in statistics improves the precision of the outcomes. Can use one of the most crucial steps before a dataset for machine learning is the pre-processing of the data. The data from the real world need to be more consistent and complete.

Therefore, must clean it. Not pre-processing, like magic corrections, could cause the system to ignore crucial words. can use one of the most crucial steps before a dataset for machine learning is pre-processing and concentrating the effort of the data. The actual statistics need to be more vehement, lacking, and contradictory. Therefore, must clean it. Before the dataset can be used for machine learning, it must do this. The statistics from the real world need to be corrected and contradictory. Therefore, must clean it.
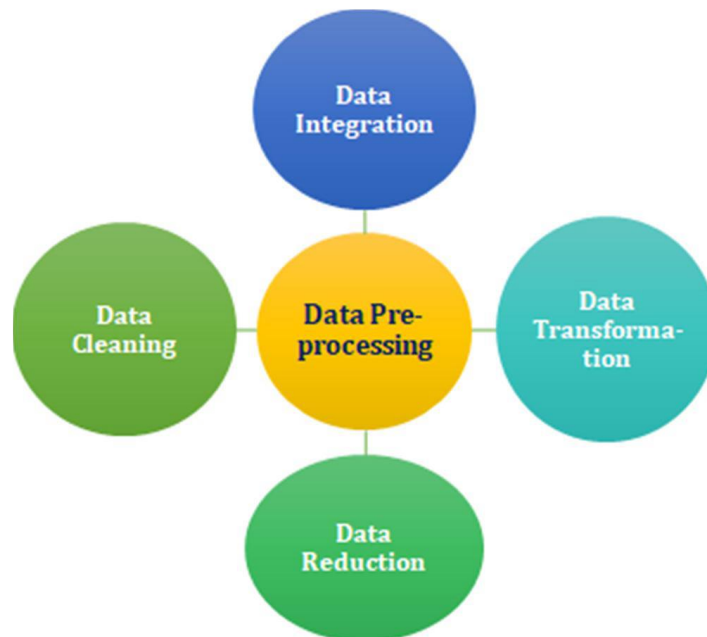

Fig 3. Data Pre-Processing

110

CONCLUSION

Various symbolic and machine learning methods for recognizing emotions in text exist. Symbolic methods are more complicated and inefficient than Machine Learning methods. To achieve 100% accuracy, can use these two methods in conjunction. In this paper, we tend to took the Sanders scientific dataset to dissect the tweets. Created the feature vector to gauge Twitter sentiments exploitation Machine Learning techniques when pre-processing the information. Consequently, the program's sentiment are divided into positive and negative classes and displayed in a very chart and hypertext mark-up language page. However, the program was originally meant to be developed as an online application thanks to Django's limitation that it will solely run on Linux servers or LAMP. As a result, it can't be done. As a result, this component should be improved further in subsequent research. The KNN algorithm is used to make sentiment analysis more effective. In contrast, the Naive Bayes algorithm makes sentiment analysis easy and effective by classifying tweets as positive, negative, or zero. Whenever a tweet is taken care of for feeling examination, it goes through different periods of opinion examination. Knowing a tweet's morph and elements is necessary for analysis. In this review paper, each of these phases and components of sentiment analysis is briefly discussed.

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.