SENTIMENT ANALYSIS OF SOCIAL MEDIA USING R WITH DEEP AND MACHINE LEARNING: A DEEP DIVE

Shailja Bhardwaj[1], Deepak Arya[2], Sakendra Kumar[3], Rachan Karmakar[4], Gaurav Gupta[5]

[1]Invertis University, Bareilly, UP, India
[2,5]Department of CSE, Roorkee Institute of Technology, Roorkee, Uttarakhand, India
[3]Dr. Rajendra Prasad Central Agricultural University, Pusa, Samastipur, Bihar, India
[4]Dept. of Environmental Science, NIMS University, Jaipur, Rajasthan, India
Shailjab8@gmail.com[1]

*Abstract:*Social Media is one of the greatest stage for data sharing. Social data analysis is the examination of individuals' cooperation in social context [1]. The data investigated here is gathered through person to person communication site Twitter. This paper is about the Sentiment Analysis of tweets on point BARCELONA TERROR ATTACK [6]. Fundamental inspiration to utilize this subject is to notice, look at and dissect how individuals censure what is happening either by communicating their hostility against psychological militant or supporting the people in question, as we as a whole denounce such brutal exercises in our own specific manner. This paper is a general report on sentiment analysis by utilizing R language with deep and ML draws near.

INTRODUCTION

Nowadays, Social media is ending up being progressively better known since PDAs can get to casual local areas successfully from wherever. Thusly, Social media is transforming into a huge point for analysis in various fields. As quantities of people using relational association are creating bit by bit, to talk with their sidekicks so they can share their own tendency standards and viewpoints are made for gigantic scope[1]. Online life Monitoring or following is a most huge subject in the current circumstance. In today various associations have been using Social Media Marketing to broadcast their things or brands, so it gets crucial for them that they can have the choice to determine the accomplishment and accommodation of each and everything [2]. For Constructing a Social Media Monitoring, a different instrument has been required which incorporates two fragments: one to survey what number of client of their picture are pulled in due to their progression and second to find individuals' assessment of the particular brand.

To evaluate the assessment of the clients isn't so basic as it appears to all clients. For evaluating their mindset might expect to perform Sentiment Analysis, which is described as recognizing the limit of client lead, the theoretical and the sensations of explicit record or sentence. To handle this, we really want Machine Learning and Natural Language Processing methodologies and here most of the planners standing up to inconvenience when they are endeavoring to outline their own mechanical assemblies.

Sentiment analysis assessments and focuses the evaluations of people on unambiguous components, for instance, things, events...Etc. Nostalgic Analysis perceives the inclination in sort of happy which was given by client's or alternately maker's by then assessments it, hence the goal of the contemplative examination is to recognize sentiments which was conveyed and request them into the different speed of limit, for instance, a rating structure [3].

SENTIMENT ANALYSIS

Sentiment analysis alludes to the utilization of normal language handling to distinguish and remove uneven data in source materials or basically it alludes to the most common way of identifying the extremity of the text. It is likewise alluded as assessment mining, as it determines the assessment, or the demeanor of a client. A typical methodology of utilizing this is depicted individuals' thought process about a

specific theme. Opinion examination helps in deciding the contemplations of a speaker or an essayist as for some topic or the generally context oriented extremity of a report. The demeanor might be their choice or gauge, the close to home condition of the client while composing [4].

There was a review to investigate the opinions of individuals for the BARCELONA TERROR ATTACK [6]. The investigation will comprise of eight feelings and two opinions good and pessimistic. Reference diagram portrayal is utilized to picture the different feelings

behind tweets. It is very obvious that the negative bar is most elevated due to utilization of words like dread; kill which shows high regrettable tendency. The following bar is positive demonstrating the utilization of hopeful words like supplicate which were utilized by individuals for the casualties of the assault. The get_nrc opinion work from the bundle syuzhet is utilized which will contrast all the tokenized words and the wordsentinetEmoLex which contain an enormous number of words with various feelings. On the off chance that a word coordinates with the word present in the sentinet, the prelisted feelings for the word will be expanded by one. On adding every one of the qualities complete inclination and opinion can be determined [5].
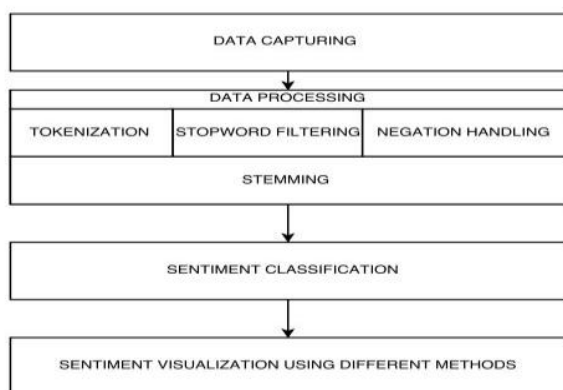
PROCESS OF SENTIMENT ANALYSIS

Fig.1

The Sentimental Analysis Flow Chart represented in Fig.1 is explained in brief below**[3]**: -

Data Acquisition: In this data acquisition, data are gathered from various pertinent sources, for example, web creeping, twitter tweets, online audit, newsfeeds, report filtering and so on.

Preprocessing: It is used to remove noisy, inconsistent and incomplete data. For doing the classification, Text preprocessing and feature extraction is a preliminary phase.

Preprocessing involves 3 steps:

Tokenization or segmentation: It is the method involved with parting a string of composed language into its words. Text data comprises of a block of characters alluded to as tokens. So the archives are being isolated as tokens and have been utilized for additional handling.

Removal of stop words: Stop words are the words which are required to have been filtered for example might be previously or after natural language processing. Stop words will be words which contain minimal information. Different apparatuses explicitly keep away

from to eliminate these stop words to help express pursuit. A few assortments of words can be picked as stop words for any reason. Some web search tools, eliminates the vast majority of the well-known words which incorporate lexical words, for example, "need" from a text to further develop execution. Web search tool or regular language handling might contain an assortment of stop words. It incorporates English stop words, for example, "and", "the", "a", "it", "you", "may", "that", "I", "an", "of" and so on which are considered as 'useful words' as they don't have meaning.

Specialists have shown that by eliminating prevent words from the record, you can get the advantage of diminished file size absent much by way of influencing the exactness of a client's. Yet, care ought to be thought about anyway to take the client's requirements. Generally, all web search tools helps in dispensing with the prevent words from their records. With the assistance of disposing of prevent words from the file, the record size can be diminished to around 33% for a word level list. While evaluating the substance of regular language handling, significance of word can be conveyed all the more obviously by eliminating the utilitarian word [8].

Stemming: It is the term which used to describe the process to reduce derived words to their origin word stem. Since 1960s, algorithms for stemming have been studied in the field of computer science. Different Stemming methods are commonly referred as stemming algorithms or stemmers. For English, the stemmer example is that, it should identify the string "cats", "catty" as based on the root word "cat", and also "walks", "walked", "walking" as based on the root word "walk" [9].

FORMATION OF WORDCLOUD

WORDCLOUD is the visual portrayal of words in the tweets. Here word cloud is utilized to picture fundamental data from the tweets.

The underneath word cloud shows that most often involved words in the tweets are kill, supplication, assault, fear, influence, individuals, etc. The various tones and size of the words show their recurrence for instance kill, supplication and assault have higher recurrence than different words. These words address the quick reaction and response of the people[7].
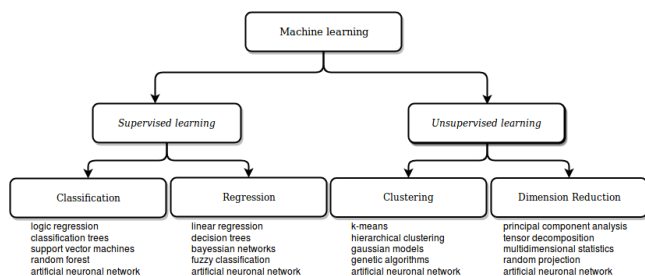
Fig 2: Most frequently used words

SENTIMENT ANALYSIS ARCHITECHTURE (by R language)

a) Twitter Application Authentication:
We really want to interface with Twitter API by utilizing the login certifications of the Twitter designer application. It is critical to validate, to associate R Studio with

Twitter for extricating the tweets. When the verification is finished, continue to further means.

b) Installing R Packages:
Installing of required bundles is vital to play out the analysis. The bundle comprises of different capacities which will be required in dissecting the sentiments.

c) Extraction of tweets:
It is to gather the information from the tweets on any subject utilizing hashtag "#".

d) Data Preprocessing and Data Cleaning:
The information is cleaned by eliminating undesirable articulations and words.

e) Data Modeling and Cleaning:
After recovery and it is changed to clean the information what's more, ready in an unmistakable organized configuration to recover opinions.

f) Retrieving Sentiments:
Analysis of opinions is performed.

g) Graphical Representation:
It is the last advance, where the opinions are plotted and are imagined by diagrams and word cloud[6].

DEEP LEARNING PARADIGM

Neural networks as of late have turned into an extremely famous subject of examination in the field of regular language handling, including feeling examination. Neural networks are demonstrating valuable in addressing practically any AI grouping issue.

The main change required is characterizing its design — the quantity of secret layers to be utilized,number of stowed away units to be available in each layer, initiation work for every hub, mistake limit for the information, the sort of between associations, and so on.

When a reasonable Neural network design is intended for the central concern, an answer for the grouping issue can acquired utilize profound learning models. The main interest for profound learning models is sufficient preparation information and enough time and assets to prepare the organization for characterization.

Obviously, a conventional ML calculation can be planned to utilize profound advancement yet not fundamentally the other way around. This is on the grounds that neural networks are fit for catching extremely complex attributes of information with next to no critical contribution to difficult work instead of machine learning frameworks. Profound learning utilizes profound neural organizations to learn great portrayals of the input information, which can then be utilized to perform explicit undertakings.

MACHINE LEARNING APPROACHES

Natural Language Processing (NLP) is a region which make machines, proficient to comprehend regular dialects for example English, Chinese and so forth. Feeling Analysis is subfield of NLP, it concentrates on that the way in which machines process text and perform grouping of text. This discipline utilizes Natural Language Processing calculations to remove highlights and AI calculations use to prepare machine for explicit datasets and for order and on the off chance that increment the quantity of elements, precision likewise increments. Ordinarily Sentence level opinion order need sentence names utilized managed realizing which is tedious and costly because of manual comment Fig. 3. Showing the ML draws near, sub classes and methods. Where Naïve Bayes, SVM, Neural Network are AI calculations.

Fig-3: Machine Learning Approaches

LIMITATIONS AND CHALLENGES

- The emotions which are presented through irony,humor sarcasm cannot be treated very well bypresent sentiment analysis tools.
- Emotions expressed through emoticons cannot bedecoded into proper sentiment.
- Use of abbreviations, slang words or some locallanguage derived words stand nowhere in thesentiment dictionary.
- Number of tweets used 4000 which is not muchcompared to the overall tweets posted by peopleacross the globe.
- Use of mixed language words that is transliteratedwords makes it difficult for analysis.
- Limitation of collection of words in wordnet which compares the sentiments.
- Only recent tweets maximum of one week andonly in text format can be analyzed, while otherforms of media and communication result canaffect the analysis if taken into considerations.

CONCLUSION

The Paper begins from the comprehension of feelings and assessments express by individuals on informal organizations and different web sources. Notions are the greatest source to break down the surveys of any item and administrations. Slant examination assumes a greatest job in finding those feelings and conclusions on World Wide Web. Estimation investigation is useful strategy for individuals from whom they can communicate their perspectives about any subject, test, substance or item. Supposition examination will dissect these statements of informal organizations by utilizing different apparatuses and methods. Opinion examination groupings are isolated into different levels. I)Document level can check the entire record ii) Sentence level investigate the sentence of a report iii)Aspect level consideration about individuals assessments of web based life about any item or any substance identified with

that item. iv)Sentiment vocabulary center around the feeling words. Here and there words have unbiased importance neither positive nor negative. Client administrations, Customer associations, Sentiment requests and huge hierarchical information are the wellsprings of information valuable to communicate different emotions and information. Remarks, unpredictable sentiments, counterfeit remarks, Sarcasm moves identified with online networking like twitter. Because of these procedures like opinion examination came in to light and delivers enormous outcomes by taking care of such issues. A few issues and difficulties are as yet unsolved in the field of AI, anyway when we talk about opinion examination its valuable procedures and approaches help a great deal to take care of different issue areas in the AI. These arrangements can be valuable to deal with up and coming difficulties in the field of estimation investigation. Momentum investigate in assessment examination concentrated on improving the effectiveness of calculations to discover supposition and surveys it prompts the decrease of human endeavors. Assumption arrangement includes different vocabulary and corpus of words.

REFERENCES

[1] Rajni Singh ,Rajdeep Kaur, "Sentiment Analysis on Social Media and Online Review" Department Of Computer Science and Engineering Lovely Professional University, Punjab

[2] Basant Agarwal, Narmita Mittal, Erik Cambria. 2013 "Enhancing Sentiment Classification Performance using Bi tagged Phrases", 13th International Conference on Data Mining Workshops, IEEE.

[3] KoneruLakshmaiah, MandavaGeetha Bhargava1*, DuvvadaRajeswara Rao2"Sentimental analysis on social media data using R programming",

[4] Vivek Sharma, Apoorv Agarwal, "Sentiments Mining and Classification of Music Lyrics using SentiWordNet", Symposium on Colossal Data Analysis and Networking 2016.

[5] Giuseppe Bruno, "Text Mining and Sentiment Extraction in Central Bank Documents", 2016 IEEE International Conference on Big Data.

[6] 1SONAL SINGH, 2SHYAM S CHOUDHARYSOCIAL MEDIA DATA ANALYSIS: TWITTER SENTIMENTAL ANALYSISUSING R LANGUAGE. Department of Computer Science, DR. RR & DR. SR VeltechTechnicalUniversity International Journal of Advances in Electronics and Computer Science, ISSN: 2393-2835 Volume-4, Issue-11, Nov.-2017

[7] RabiaBatool, Asad Masood Khattak, "Precise Tweet Classification and Sentiment Analysis."

[8] Stop words: http://en.wikipedia.org/wiki/Stop_words

[9] Stemming: http://en.wikipedia.org/wiki/Stemming

[10] O. R. L, "Using Machine Learning Techniques for Sentiment Analysis," FINAL PROJECT ON COMPUTER ENGINEERING, EE/UAB FG COMPUTER ENGINEERING, 2017.

[11] M. Vadivukarassi, N. Puviarasan and P. Aruna, "Sentimental Analysis of Tweets Using Naive Bayes Algorithm," World Applied Sciences Journal 35 (1), pp. 54-59, 2017.

[12] BinitaVerma, Ramjeevan Singh Thakur, "Sentiment Analysis Using Lexicon and Machine Learning-Based Approaches: A Survey", Proceedings of International Conference on Recent Advancement on Computer and Communication pp 441–447, 2018.