iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 **Issue:** VII **Month of publication:** July 2021

DOI: https://doi.org/10.22214/ijraset.2021.36552

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: 7.429
Volume 9 Issue VII July 2021- Available at www.ijraset.com

Sentiment Analysis: Classifier and Summarizer

Ayushi Aggarwal[1], Nishima Arora[2], Shikha Deep[3], Ms. Krista Chaudhary[4]
[1, 23]*Student, Computer Science and Engineering, Krishna Engineering College, Ghaziabad, India*
[4]*Assistant Professor, Computer Science and Engineering, Krishna Engineering College, Ghaziabad, India*

Abstract: ***Sentiment Analysis and summarization has a large number of application that are useful for determining the sentiment of the text and summarizing a big text into a small paragraph of few lines. Thus it has become an important topic to work onand for fulling the requirements of the customer. It has also become an important topic for researchers to focus on, as it is highly demanded and beneficial in different fields of product, services and growth of the business. At present when 89.9% of people are using social media platform, they express their reviews, feelings, emotions and share their comments and some exciting activities of their life through social media platform, so it becomes very important to analyses them and classify them as positive or negative, this can be done with the help of sentiment analysis. Also, to find the summary of a big document with large amount of data summarizer isvery useful as we can get the summary of a document in the favorable number of line. The basic model of sentimental analysis classifythe word as positive as negative with the help of some machine learning approaches, which will help in improving the quality of product and providing the service to the customer for building up a healthy competition in market and keeping the goodwill of the business . It also displays the output in the form of graph whose data is taken from social media platform. Sentimental analysis also helps in getting the summary of the document by picking the lines containing the words having maximum repetitions. It has been found that sentiment analysis able to classify the positive sentence by giving the output as 1 and negative sentences as 0. The model which is being built also graphically represents the classifications of positive and negative words picked from the dataset and it's also useful in summarizing a document Thus sentiment analysis comes out to be very important for classify the unstructured data on social media platform and so there is always a scope of building a better model which is more accurate and efficient.***

Index Terms: Logistic regression, Machine learning, Naïve Bayes, Natural language processing, Sentiment analysis

I. INTRODUCTION

Now a days, Social media is an integral part of our everyday life. Around the world billions of people and millions of companies are actively using the social media. People expresses their views, emotions, feelings and share comments and happenings in their life with the help of different social media platforms such as Twitter, Instagram, Facebook etc. Since different people have different mode of thinking and understanding, therefore the data is scattered and unstructured. It has been analyzed that 80% of data is unstructured all over the world and therefore it needs to have a proper structure and classification for accurate analysis and to gain valuable insights. Sentiment analysis, also known as opinion mining or text mining is the analysis of text message and knowing its polarity that iswhether the text is positive or negative. It uses information extraction and natural language processing technique to classify the data. Sentiment analysis is an enormous model whose applications can be endlessly applied in any company and on any field of technology from agriculture and cultivation to finance and healthcare. A lot of work is already done on sentiment analysis such as movie reviews ,quality of product, uber services, condition of stock markets, brand monitoring analysis, market research competitor analysis etc. And still there is a wide scope to work on it and make it more efficient and realistic model for analyzing the sentiments. Sentiment analysis uses artificial intelligence, machine learning processing and natural language processing technologies and as these approaches are booming in the world all around the importance of sentiment analysis will be groomed in the coming time with a demand of more accurate and efficient model possessing the latest features. The use of supervised and unsupervised learning approaches will help in accurately classifying the text. Bags of words, support vector machine, logistic regression naïve bayes theoremand statistics are some common techniques to achieve a model classifying the data according to its polarity. Although deep learning is a relatively new approach in sentiment analysis, it is already promising in polarity analysis In this model, a dataset is used to train the model so that it analyze the sentiments of the sentences given by the users as input. Themodel will also fetch the data from the social media website to check the polarity of comments as positive or negative. The result would be shown as a graphical representation depicting the polarity of the data, which will help the companies in analyzing the reviews of their products and services and also for the researchers to classify the data they are working on. This model gives a summarization of the URLs or text in the number of lines required by the user. The best lines from the text will be chosen according tothe score of words in the lines.

A. Sentiment Analysis,Classifier And Example

1) *Sentiment Analysis:* Identify feelings, categorizing and expressed that are written in form of sentences, text, document, paragraph or clause especially to determine that whether it is positive, negative. It can be done by various machine learning techniques.

2) *Classifier:* Sentiment analysis classifies the text into positive and negative polarity by using natural language processing which process the human language into the language which can machine recognize. Many tasks are included such as sentiment classification, sentiment extraction, sentiment detection, etc.

3) *Sentiment Analysis Example:* Sentiment analysis can classify the text as positive or negative using different approach

a) "The rooms were very beautiful" OUTPUT- Positive/ [1]

b) "The clothes of the person was very dirty"OUTPUT- Negative/ [0]

c) " The service of the room was not good"OUTPUT- Negative/ [0]

The model could give the result as a normal text, pie chart, graph, chart table depending on the model we are dealing with inthe model and the algorithm used.

B. Text Summmarizer

In this day to day life it is difficult to go through the vast amount of content and extract information from it. Natural language processing is a method used in machine learning and make machine to understand the human languages. Text analyses is one of the common field for researchers. Data from large documents, URLs are extracted and manipulate that data and create the shorter version of the large data obtained. This technique is known as text summarization..

In this paper, section [2] consist of a literature review of paper on sentiment analysis is depicted in Table [1]. Different methodology used in sentiment analysis mentioned in section [3], consisting of machine learning approaches is explained and its classified diagram is shown in Figure [1]. The procedure for sentiment analysis is explained through a Flowchart [1]. The proposed basic model consisting of classification through text and graphs and summarizer is explained in Section [5] where image [1] [2] [3] shows the result of how the text is classified as positive or negative and graph[1] [2] [3] displace the output in the form graph . The summarizer is explained in flowchart [6] and its result is shown in image [4]. The paper also contains applications, advantages and disadvantages of sentiment analysis as mentioned in section [6]. The conclusion of the model and its result are mentions in the section [7] and section [9] respectively. The paper also give future reference and directions for a scope of making the model more accurate in future as shown in section [10].

II. LITERATURE SURVEY

S.no.	Author	Year	Paper Name	Remark
1	B. Lakshmi DeviV. Varaswathi BaiSomulaRamasubbareddy	2020	Sentiment Analysis on MovieReviews	This indicates the classification of good or bad comments on movie and rating it which gives a scope of improvement in next movie and knowing the likes and dislikes of people.
2	Yash Indulkar	2020	Sentiment Analysis of Uber & Olausing Deep Learning	This indicates a comparative analysis of ola cab and uber cab on different parameters using deeplearning approaches
3	LI YANG YING LI , JINWANG, R. SIMONSHERRATT 3	2020	Sentiment Analysis for E-Commerce Product Reviews in Chinese Based onSentiment Lexicon and Deep Learning	E-commerce product are used by many people and knowing the reviews for customer helps in predicting the manufacturing required and theimprovement needed.
4	Abdullah Alsaeedi, Mohammad Zubair Khan21	2019	A Study on Sentiment AnalysisTechniques of Twitter Data	People share their emotions, complains, feelings and thoughts on twitter and to classify the comments as positive or negative sentiment analysis plays a vital role.
5	Saeed-Ul Hassana, , Naif R. Aljohanib , Nimra Idreesa , Raheem Sarwara , Raheel Nawazc ,Eugenio MartínezCámarad	2019	Predicting Literature's Early Impact with Sentiment Analysis inTwitter	Correct citation of words in different categoriesis important as positive comments and neutral comments if not classified or classified as a negative comment could have a great impact andmay lead to misconceptions.

	, Sebastián Venturae,b , Francisco Herrerad,b			
6	Pratima Deshpande1, Purva Joshi, Diptee Madekar, Pratiksha Pawar, Prof. M.D. Salunke	2019	A Survey On: Classification of Twitter data Using Sentiment Analysis	This is review paper showing the surveys on twitter data and how sentiment analysis plays an important role in classifying the result of surveysas positive, negative or neutral
7	Chaitanya Bhagat Deepak T. Mane RajarshiShahu	2019	Survey On Text Categorization onSentiment Analysis	In this the text is classified as positive and negative using sentiment analysis. Many machine learning approaches are used for the same.
8	Klaićeva 7, Zagreb, Croatia	2019	Sentiment Analysis of Text Documents	Classification of whole document as positive , negative is done using sentiment analysis on documents with the help of support vector machine and corpus based approach
9	Siddharth bansal, A.M.J.Muthu kumaran	2019	Word Vector Representations tobuild an Emojifier using LSTM	The text can be converted into suitable emojirather than the big text.
10	Doaa Mohey El-Din MohamedHussein	2018	A survey on sentiment analysischallenges	A brief discussion on what challenges are faced in sentiment analysis is depicted leading to a scope of further improvement in the model.
11	Akshi Kumar ArunimaJaiswal	2018	Systematic literature review ofsentiment analysis on Twitter using soft computing techniques	A review paper depicting all the soft computing techniques used in classifying the comments on twitter
12	Anima C Anima Chandranhandran	2018	Sentiment Analysis of Stock Market Related Tweets	The rise and fall of the stocks and its prediction of growth or fall is done using sentiment analysisby classifying each review of stock on twitter as positive or negative.
13	Sushma R ,Nishkala L K Rakshitha H Prakshitha K S 1 ; Mrs. Shruthi T R2	2017	Political Tendency Identification in Twitter Using Naive BayesClassification	Sentiment analysis plays a vital role in classifying the comment on twitter in the area of politics leading to opting a better politician whose reviews are positive
14	ChhayaChauhanSmritiSehgal2	2017	Sentiment Analysis On ProductReviews	Customers review on a product plays a vital role in knowing about the likes and dislikes of theproduct which help in its improvement.
15	Anna Baj-Rogowska	2017	Sentiment Analysis of FacebookPosts: the Uber case	Facebook is used by 90% of people. People post their activities and outgoings on it which helps indetermining the reviews of their travelling experience and mishappenings.
16	Xing Fang and Justin	2015	Sentiment analysis using product	Knowing the reviews of the customer about the
	Zhan		review data	product helps in analyzing the requirements of acustomer which are lacking in the product and knowing the demand of the product.
17	MSc. Majlinda Axhiu PhD. Florida VeljanoskaPhD. Biljana Ciglovska MSc. Mirlinda Husejni	2014	The Usage of Sentiment Analysisfor Hearing the Voice of the Customer and Improving Businesses	Customer's reviews matter a lot in knowing the reviews of the product or services of a business which helps to make improvements according to the requirement of the customer.
18	Rao, Tushar and Srivastava, Saket	2012	Analyzing stock market movements using Twitter sentiment analysis	65% people invest in stock market and so toknow which stock marketing will be beneficial to invest in, sentiment analysis plays a vital role.
19	G.Vinodhini* RM.Chandrasekaran	2012	Sentiment Analysis and OpinionMining: A Survey	A survey report on the various applications of sentiment analysis, how sentiment analysis is used in depicting the sentiments or views of customers and how opinion mining can impact a business or company is depicted.
20	Viswa Mani Kiran Peddinti, Prakriti Chintalapoodi	2011	Domain adaptation in sentimentanalysis of Twitter Share on	Adapting data from other domains and then performing sentiment analysis on the comments shared on twitter is the main motive.

Table 1: Literature Review of papers on sentiment analysis

The literature, models and work done on sentiment analysis in past is discussed in this section depicting the work done in each model and the improvements made to make it more efficient and accurate as shown in table 1. Sentiment has got a lot of applications and so there is always a scope of making improvement in the model which could fulfill the requirements of people in the coming future. Since it has made the life of people very easy, many researchers, engineers, scholars and students have worked on this model. R. Sharma, S. Nigam, and R. Jain, "Opinion mining of movie reviews at document level," arXiv preprint arXiv:1408.3829, 2014[17] depicts the use of sentiment analysis in knowing the reviews of the movie. Hsu, Y.W., Chen, K.H., Yang, J.J., Jaw, F. in their paper [5] Smartphone-based fall detection algorithm using feature extraction helped in predicting the downfall of service using smartphone and helped in knowing the reviews of a particular service. Hakim, A., Huq, M.S., Shanta, S., Ibrahim, B. [4] in his paper Smartphone based data mining for fall detection: analysis and design also predicted the reviews using sentiment analysis.Wang, Bo, and Min Liu in the paper [10] "Deep learning for aspect-based sentiment analysis." Elaborated the deep learning techniques and how deep learning is useful in sentiment analysis and the same was used in building a Sentiment analysis for uber and ola review and identifying which cab provides a better service. Ji, H.-Y. Zhang, and J.-Q Wang, ''A fuzzy decision support model with sentiment analysis for items comparison in E-commerce [12] in his paper wrote about different models that will help in analyzing the items and comparing different E- commerce. This helped in clustering the reviews of different people as positive and negative and using this information identifying how the item on E commerce is and what is its ranking. Sharma, S. Nigam, and R. Jain, "Opinion mining of movie reviews at document level," arXiv preprint arXiv: 1408.3829, 2014[17] depicts the use of sentiment analysis in knowing the reviews of the movie Medha Khuran, et al. "Sentiment Analysis Framework of Twitter Data using Classification", 5th IEEE International Conference, 2018 [26] [27][28][29][30] and Go, Alec, Lei Huang, and Richa Bhayani. "Twitter sentiment analysis." Entropy 17 (2009): 252[6] tells the proper framework of sentiment analysis and its use in classifying the twitter comments.M. Bouazizi and T. Ohtsuki, ―A Pattern- Based Approach for Multi-class Sentiment Analysis in Twitter,‖ in Proc. IEEE ACCESS, pp. 20617-20639, 2017.[31][32][33][34] gives a servey of classification of twitter data and how sentiment analysis and its approaches are used to classify tweets as positive, negative or neutral in tweets. Kumar A, Sebastian TM. Sentiment analysis on Twitter. Int J Comput Sci Issues. 2012; 9(4):372- 378[44] again tells us about how sentiment analysis is used in classifying twitter comments. Aggarwal CC, Zhai C. A survey of text classification algorithms. In: Mining Text Data. Boston, MA: Springer Science Business Media; 2012[42] tells about all the technique used to classify the text. All the machine learning techniques are explained in elaborated form. B. Liu, "Opinion Mining and Sentiment Analysis", in: Web Data Mining, Data-Centric Systems and Applications, Springer Berlin Heidelberg 2011, ISBN 978-3- 642-19460-3, pp. 459-526.[52] helped in jow to start with basic model of analyzing the text as positive , negative or neutral. Xia, C. Zong, and S. Li, "Ensemble of feature sets and classification algorithms for sentiment classification," Information Sciences, vol. 181, no. 6, pp. 1138-1152, 2011/03/15/ 2011.[16] essembled different feature set to give the output as an emoji rather than a text. Alexander, H., Paul, V.I., Bas, H., Flavius, F., Uzay, K., 2011. Determining negation scope and strength in sentiment analysis. In: Proceedings of the 2011 IEEE International Conference on Systems, Man, and Cybernetics (SMC 2011). IEEE Computer Society [38] made a broder view in determining the challenges while classifying the model more accurately. Rajman M, Besancon R. Text mining- knowledge extraction from unstructured textual data. Advances in Data Science and Classification: Proceedings of the 6th Conference of the International Federation of Classification Societies (IFCS-98) Università ''La Sapienza'', Rome, 21-24 July, 1998. Berlin, Germany[43] This paper clearly classify the soft computing techniques which would help in classifying the text as positive negative and neutral for twitter comments. Bollen, H. Mao, and X.-J. Zeng, "Twitter mood predicts the stock market," Computer, vol. 1010, no.3003v1, pp. 1–8,2010.[64] this paper is clear prediction of sentiment analysis text which would depict the uprisal and downfall of stock market.Ankur Goel, Jyoti Gautam, Sitesh Kumar, ―Real Time Sentiment Analysis of Tweets Using Naive Bayes,‖ in 2nd International Conference on Next Generation Computing Technologies, Dehradun, 2016[34] sentiment analysis of tweets on political application is done on twitter comments and proper classification is done accordinglyL. Banić, A. Mihanović, M. Brakus, "Using Big Data and Sentiment Analysis in Product Evaluation", Information & Communication Technology Electronics & Microelectronics (MIPRO), 36th International Convention 2013[54] [55] product review using sentiment analysis is done.Delua, J. 2011. Big Data Meets Sentiment Analysis! informatics, [blog] 27th June, Available at: http://blogs.informatica.com/perspectives/2011/06/27/big-data- meets-sentiment-analysis[57] Customer's reviews matter a lot in knowing the reviews of the product or services of a business which helps to make improvements according to the requirement of the customer. Bing Xu, tie-jun zhao, de-quan zheng, shan-yu wang, Product features mining based on conditional random fields model " , Proceedings of the Ninth International Conference on Machine Learning and Cybernetics, Qingdao, 11-14 July2010 [68] paper clarifies the different domains on which sentiment analysis plays a vital role and the impact of classifying the text as positive, negative and which helps in taking accurate decision.

Thus there is already a lot of work done on sentiment analysis and still there is high future scope for a better model which is efficient and more accurate.

III. METHODOLOGY OF SENTIMENT ANALYSIS

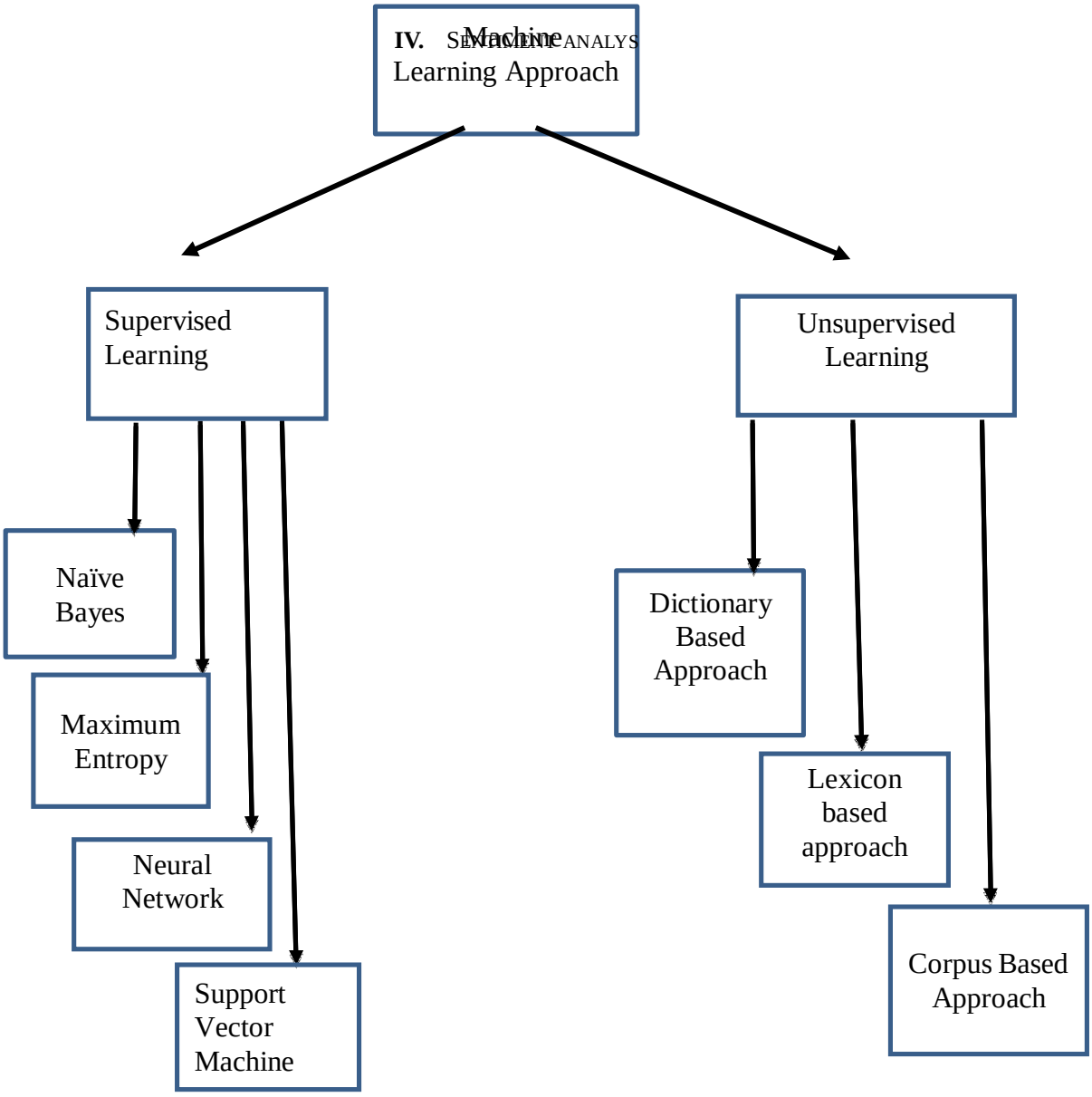


Figure 1: Classification of Machine Learning Approaches using Sentiment Analysis

The different methodologies of sentiment analysis is explained by classifying different machine learning approaches as is shown infigure 1. Each approach has different use which is explained below:

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VII July 2021- Available at www.ijraset.com

IV. MACHINE LEARNING APPROACHES

Machine learning is an approach to an artificial intelligence, text is classified into classes with the use of classification techniques. Pastexperience is used by machine to learn the data. There are two types of machine such as supervised and unsupervised.

A. Unsupervised Machine Learning

Unlabelled data is used to train unsupervised learning algorithm. To find the structure and pattern from the input data is the main aim of the unsupervised learning. No supervision is required for this. Patterns are found from the data by its own.

1) *Dictionary Based Approach:* This is a computational approach, the reader conveys a text of which feeling are measured. It is based on the manipulation of manually collected and annotated terms. Positive and negative is a binary classification, but multiple dimensions such as happy, sad, angry, etc. can be extended. WordNet dictionary is an example, which develops the SentiWordNet.
2) *Lexicon Based Approach:* To determine the polarity a sentiment dictionary which contains opinion words is used and matched with the data. Positive and negative words are describe by assigning a score to the sentimental words
3) *Corpus Based Approach:* Related to a specific domain corpus based approach have aim of providing dictionary. The use of statistical and semantic techniques of related words are grown through the search, a set of seed opinion words generates the dictionary.

B. Supervised Machine Learning

Labelled data is used to train supervised learning algorithm. The mapping function is required to map the input without the variable(X), with the output variable(Y)

$Y = f(X)$

Supervision is required for this. Output is predictive for supervised learning.

1) *Neural Network:* They can be used to cluster raw input or interpretative sensible data the signification computational resources are required neural network. Behind there prediction they interpret the logic that's why they are also called black works algorithm
2) *Support Vector Machine (SVM):* Support vector machine is the most popular algorithm of supervised learning which is used for classification problem andalso regression. It is a discriminative classifier, as it separate the hyperplane examples are pattern recognition and multimediainformation retrieval.
3) *Naïve Bayes:* Naïve Bayes algorithm is supervised learning algorithm, which is used for solving classification problems. This algorithm is based on Bayes theorem which is used in text classification. Text classification uses a high dimensional training data set-to train the model. It is a populistic classifier which is simple and effective classification algorithm. Quick predictions are made as naïve Bayes builds the fast machine learning models. Examples are: Sentiment analysis, spam filtration, etc. Formula for Bayes theorem is:

   $P(A|B) = P(B|A)\, P(A)/P(B)$

   From the above equation:

   P (A|B) is the probability of hypothesis A on the observed event B.

   P (B|A) is the probability of evidence given that probability of hypothesis is true.P (A) is the probability of hypothesis before observing evidence.

   P (B) is the probability of Evidence.

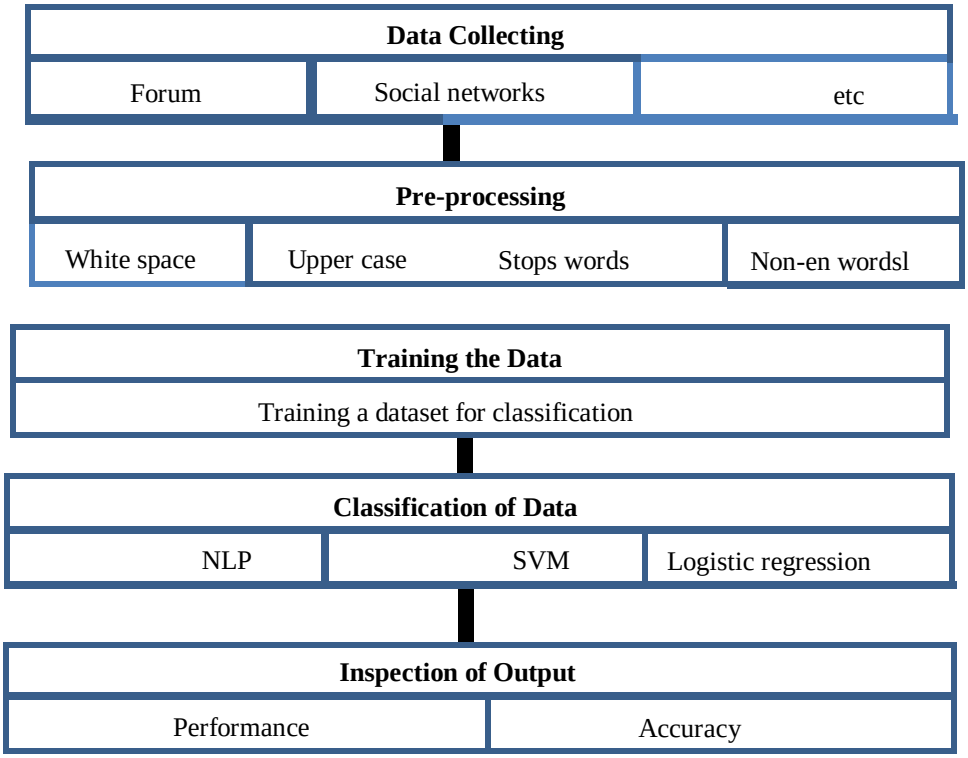
4) *Maximum Entropy:* In maximum entropy classify, no hypothesis are made on the relation between features extracted from the data set. This classifier is the probabilistic classifier belongs to exponential model. Varity of text classification problem such as sentimentalanalysis, topic classification and many more are solved by maximum entropy. As same as logistic method maximum entropy can also handle overlap feature that find distributions over classes. Unlike naive Bayes, no independent assumption for it feature made. Formula for maximum entropy is: $PME(c|d, \lambda) = \exp[\Sigma_i \lambda_i f_i(c, d)] \mid \Sigma_c \exp[\Sigma_i \lambda_i f_i(c, d)]$ Where c is the class, d is the tweet and λ_i is the weight.

V. LIBRARIES USED

1) NUMPY is a python library which is used for array processing packages. It is used to give high performance and accuracy in multidimensional array objects and tools that helps in proper working and functioning of these arrays. It also helps to integrate large variety of databases with high speed and accuracy.
2) REGULAR EXPRESSION (RE) is highly used for recognition of pattern and pattern matching using match function that helps to extract required matching strings, search a substring, replace a sub string and to match patterns against other patterns.
3) PYPLOT It is a plotting library used in python for graphs in two dimension. It has many applications such as shell, python scripts, web application servers, etc.
4) MATHPLOTLIB is basically numeric extension of numpy which is a cross platform that provides data visualization, maps, charts, plots and providing the data graphically. It is a compliment to pandas library in python.
5) BEAUTIFULSOUP is a library used in python for extracting data out of XML and HTML files. It is used to search, modify,and navigate the parser tree which saves hours of work of a programmer.
6) HEAPQ is useful for implementing items in queues having higher priority or weight in processing and the items having higher priority are displayed first.
7) SCIKIT-LEARN is most useful library in machine learning as it provides for classifying regression, classification, statistical modelling, dimensionality reduction, etc.
8) TWEEPY library is a python library which helps in communication of twitter platform with its API. It is open source and is hosted by GitHub.
9) PICKLE module is used in serializing and de- serializing object structure present in python. It is used to maintain the state ofthe program, transport the data and to store object in the form of byte stream in different file or data base.
10) UNPICKLE module is just the reverse of pickling module wherein objects are converted or converted back to the original form.
11) TfidfVectorizer is used to tokenize the data given in the document learn the vocabulary learn the data and helps to encode a new document containing words with different frequencies and weights**.**

VI.PROCEDURE OF BASIC PROPOSED MODEL

For preparing the basic model for sentiment analysis a procedure needs to be followed containing number of steps like data collection, pre-processing, training and classifying the data and inspecting the output as shown in flowchart1



Flowchart 1 : Steps involved in processor in sentimental Analysis

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VII July 2021- Available at www.ijraset.com

A. Data Collecting

People express their opinions on social media website through tweets, comments, blogs, decision boards, product review, forums and many more. These comments are expressed in different ways such as content of writing, use of short form, vocabulary, slang words etc. By different users. In this survey, a raw Cornell dataset is used which is already manually classified as positive and negative polarity. Because of the manually classified polarity the dataset is efficient and accurate but this dataset contains upper case, white space, wrong spellings, non-English words, punctuations, stopwords, URLs, mentions, etc. So this dataset requires the text pre- processing.

B. Pre-Prossesing Dataset

Variety of different tweets which is written in variety of different formats is data. The raw data highly inconsistent and contain repetitive words. This is the reason why pre-processing the raw data is integral part of preparing the dataset. So the first step towards the solution is to pre-process the data. The pre-processing includes:

1) *Cleaning The Raw Data:* The words which add no value to the meaning to the sentence are deleted and make use of valuable data. Cleaning steps are: Lower case, correct the spellings, removal of non-English tweets, removal of punctuation, removal of stopwords, removal of URLs removal of mention, removal of numbers,removal of whitespace, removal of hashtags.
2) *Tokenization:* Splitting of raw data which is phrase, sentence, paragraph into smaller chunks is called tokens and this process is called tokenization. Tokenization technique is used in almost every natural language processing task. To achieve sentiment analysis, it is important to understand the pattern of the text. For example, the text “He is running” can be tokenized into ‘He’, ’is’, ’running’.
3) *Stemming:* To obtain a root, removing and replacing suffix and prefix from token is done, this process is called stemming. Stemming is a normalization technique which reduce the number of computation required used in natural language processing. For example, the stem of words, ‘running’ and ’ran’ is ‘run’

C. Training The Dataset

Classification problems are solved by supervised learning algorithm, easily and effectively. It is better to train a classifier for future predications done on unknown data to get the reliable results. For making the model user friendly it is important to train the data manually and by machine also. Preparing a corpus the best way for training data. For the classifier this data will work as the fuel. Thealgorithm will used for the learning purpose of classifier. The model is trained to perform the analysis on the data provided by the user. To identify the text is positive and negative, scikit learn pipe line is used to train a machine learning model. Import regular expression library to provide the operations which make its convenient to handle the data. NLP convert the words into canonical formas for the formatting of the text required the library NLTK is imported. This training data set is passed through train validation and test. Method .train () is used to train the model and method .accuracy () is used to test the model on texting data. The training data set runs for 20 epoch. When the network saw the entire data once i.e. known as one epoch. The model will over fit the training data, whenthe number of epoch is increase as it tends to threat. It is better to stop at early stage to avoid the model from over fitting.
In this module, the user will add some text comment the polarity of the text whether it is positive or negative. The user can enter any comment from social media or can right there own text.

```
In [3]:
   ...: sample = ["the rooms are very disappointing"]
   ...: sample = tfidf.transform(sample).toarray()
   ...: print(clf.predict(sample))
[0]
```

Image 1: Negative text

```
In [7]:
   ...: sample = ["The pessimist sees difficulty in every opportunity"]
   ...: sample = tfidf.transform(sample).toarray()
   ...: print(clf.predict(sample))
[1]
```

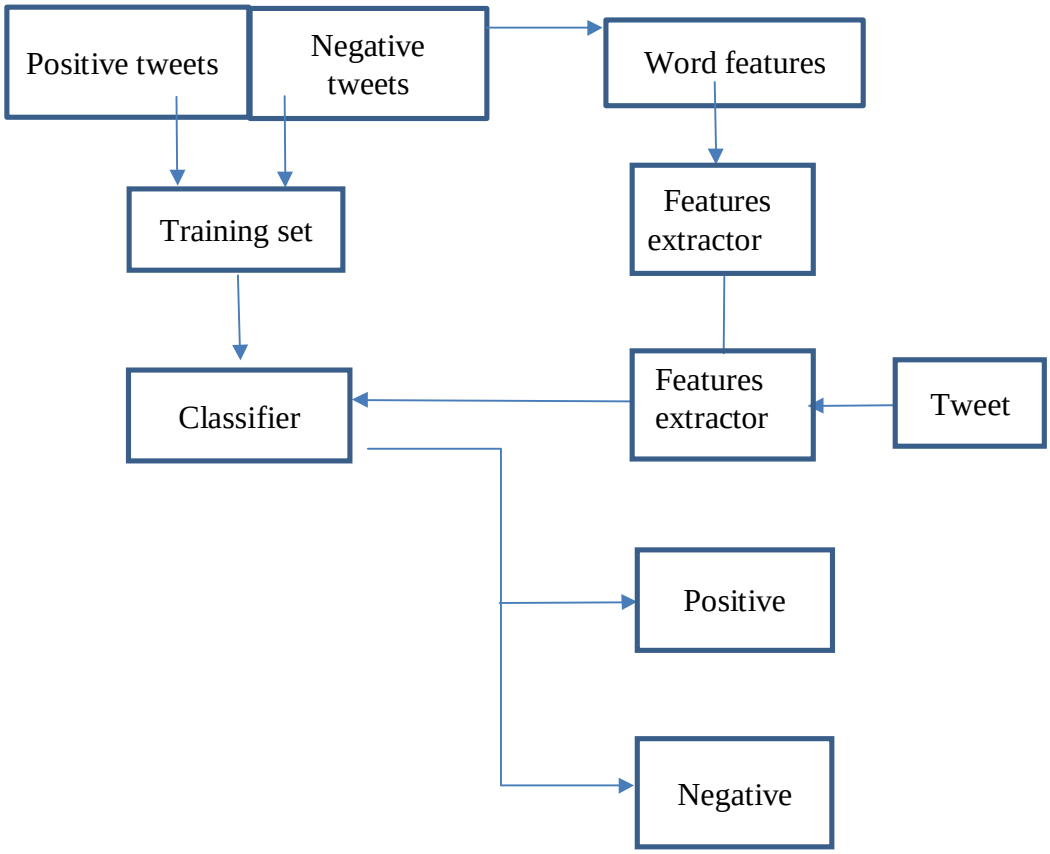
Image 2: Positive text

```
In [17]:
    ...: sample = ["a friend in need is a friend in deed"]
    ...: sample = tfidf.transform(sample).toarray()
    ...: print(clf.predict(sample))
[1]
```

Image 3: Positive text

D. Classification Of Data

Classification of data is a task which is used to classify the text as positive or negative with the help of training dataset asshown in flowchart 2.



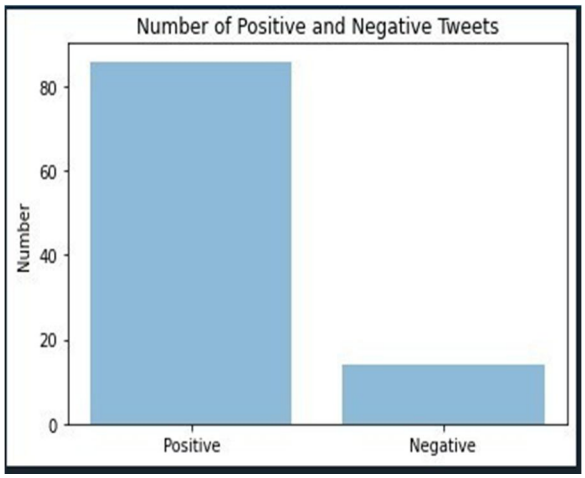
Flowchart 2: Sentiment analysis model to classify text

In this module, the comment are fetched from social networking sites as the data is unstructured. To check the polarity of thecomments there is a graphical representation to structure the data.

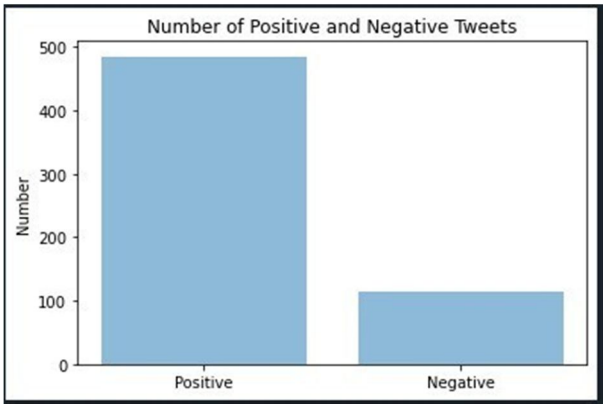
CATEGORY	POSITIVE DATASETS	NEGATIVE DATASETS	TOTAL DATASETS
Dataset 1	84	16	100
Dataset 2	543	107	650
Dataset 3	858	212	1070
Dataset 4	2038	602	2640

Table 2 : category with polarity distribution with datasets

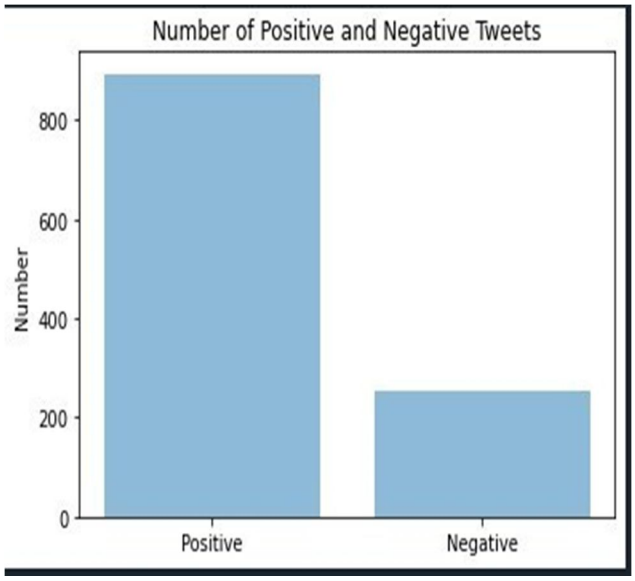
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VII July 2021- Available at www.ijraset.com



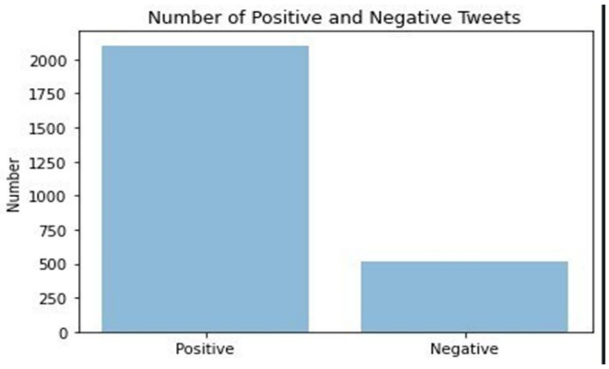
Graph 1: Showing result of Dataset1 containing 100 comments



Graph 2: Showing result of Dataset 2 containing 650 comments



Graph 3: Showing result of Dataset 3 containing 1070 comments



Graph 4: Showing result of Dataset 4 containing 2640 comments

Classification process is the heart of the entire technique. Positive and negative are the two groups in which sentiments are commonlyclassified. Each sentence is classified at this stage of methodology using:

1) *Natural Language Processing:* Natural language processing is a field of artificial intelligence which helps machine to understand text and spoken words translate andmanipulate human language. It is a discipline which derives real time computer programs that translate text from one language to another and rapidly summarize large document of text rapidly. Today NLP is booming as major text is automatic summarization, speech recognition, machine translation, conference resolution, etc.
2) *Natural Language Toolkit:* Natural Language Toolkit works with a human language data and is integral platform used to create python program. NLTK is bid in statistical and symbolling natural language processing. There are many text processing libraries bided for classification, tagging, stemming, tokenization, etc. NLTK is the most robust NLP library which understand the human language and responded with relevant response with some inbuilt packages which make machine understandable.
3) *Logistic Regression:* Logistic regression is a linear and classification algorithm and it works on predictive analysis. Its predict when the set of independent variable is binary. Data is described and a linear relationship is managed between the input and output variable. For highly accurate and effective model the linear relationship is revealed by data transforms input variable. There are two possible frame work where binary output comes - either the phenomena occurs [1], or it does not occurs [0]. As sometimes it's difficult interpret logistic regression a tool named intellects statistics allows easily to perform the analysis To extract sentiment from real time tweets and comments, the classifier is ready to use.

E. Inspection Of Output

Mainly the sentiment analysis is used to change the unstructured data into structured and meaningful data. When analysis is completedthe result will be shown as graph which will indicate the total number of positive and negative comment in the model. To analyse, Comments can take from any social media website. Also the model will display the sentiment of individual's comment or any sentence inputted by the user. The **accuracy of the model is 84.75 %** which is not bad. Studies shows that machine learning and natural language technique are used in sentimental analysis. This is a wide and open field to study.

VII. EVOLUTION OF SENTIMENT ANALYSIS: CONFUSION MATRIX

A confusion matrix is a machine learning classification techniques used to measure the performance it is the N X N matrix, where N isnumber of target classes as shown in figure 2.Recall , precision accuracy are measured which is a useful machine learning method.

Accuracy = (TP+TN)/ (TP+TN+FP+FN)

Precision = TP / (TP+FP)Recall = TP/ (TP+FN)

F1 = (2xPrecisionxRecall)/ (Precision + Recall)

1) TP: True Positive; Predicted actual positive are predicted as true positive.
2) FP: Incorrectly predicted value are predicted as actual positive value I.e. some negative values are also predictive positive, Type 1error.
3) TN: Predicted actual negative are predicted as true negative
4) FN: Incorrectly predicted value are predictive as actual negative value I.e. some positive values are also predictive negative, Type 2error.

Accuracy test can be performed by:
Accuracy = TP + TN / TP + TN + FP + FN

Calculate confusion matrix
Download the data set which is manually classified

- Train the model with the data set
- The raw of data set are required to be predictive be by the model
- Calculate the output and manual prediction

- Total Number Of Each Correct Prediction
- Total number of incorrect prediction

These number must be classified in the :

- Predictive class should be linked to the every raw of the matrix
- An actual class must be correspond to the column of the matrix
- In the table, the total number of correct and incorrect prediction are launched.

The total number of true negative predictions are inserted in the matrix as [0] [0]
The total number of true positive predictions are inserted in the matrix as [1] [1]
The total number of false negative predictions are inserted in the matrix as [0] [1]
The total number of false positive predictions are inserted in the matrix as [1] [0]

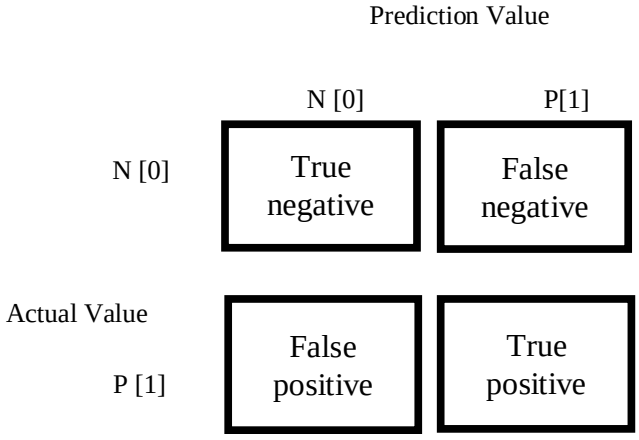
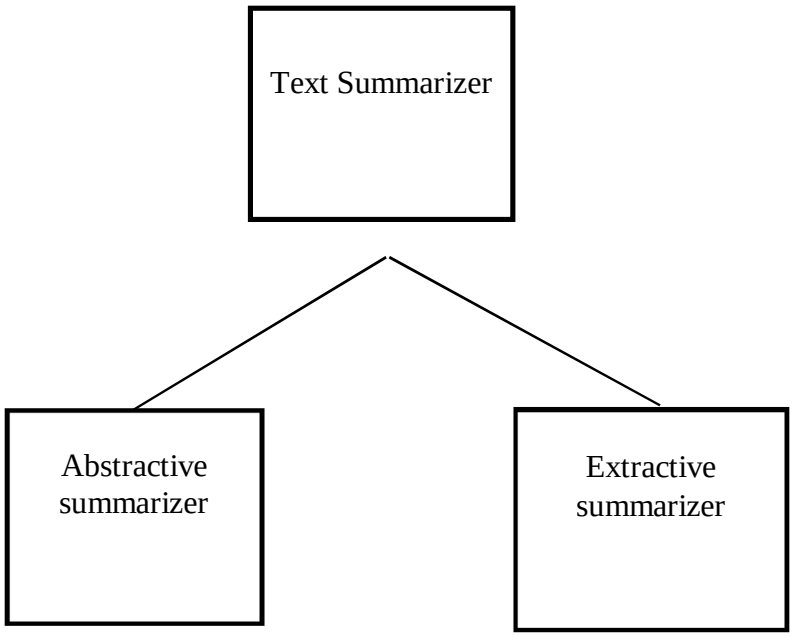


Figure 2: confusion matrix

VIII. TEXT SUMMARIZER

Now a days billions of websites and webpages are present on the internet. But in this day to day life it is difficult to go through thevast amount of content and extract information from it. So, today the technologies are reached to an different extent where all the task of human beings can be easily performed by a machine and machine learning is the field which makes such things happen. Natural language processing is a method used in machine learning and make machine to understand the human languages. Text analyses is oneof the common field for researchers. Data from large documents, URLs are extracted and manipulate that data and create the shorter version of the large data obtained. This technique is known as text summarization. Text summarizer is important to do the better research work, reduction of reading time, effective, reliable and many more.

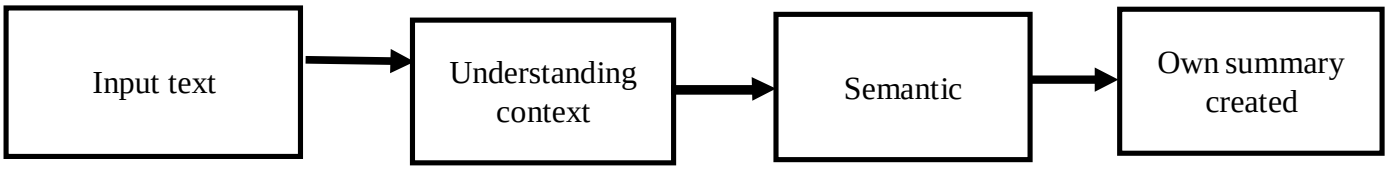
There are two types of text summarizer as shown in flowchart 3



Flowchart 3: Classification of summarizer

A. Abstractive Summarizer

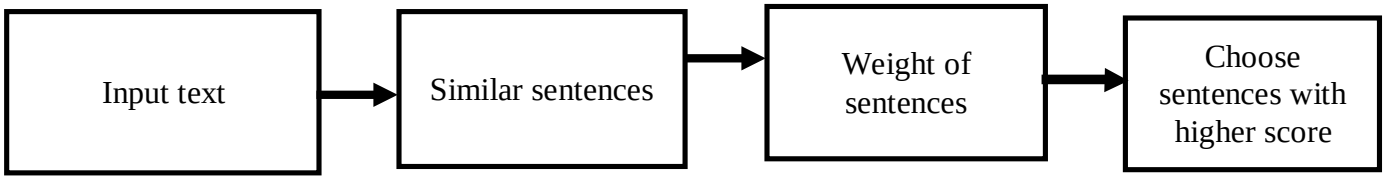
The input document is provided to the machine and machine focuses to interpret and understand the content and create its own summary as shown in flowchart 4.



Flowchart 4: Abstractive Summarizer

B. Extractive Summarization

The input document is provided to the machine. Sentence similarity is tested and await is provided to a sentence. Different techniques and methods are used to provide a score to the sentences. The higher the score, the best the sentences as shown in flowchart 5.

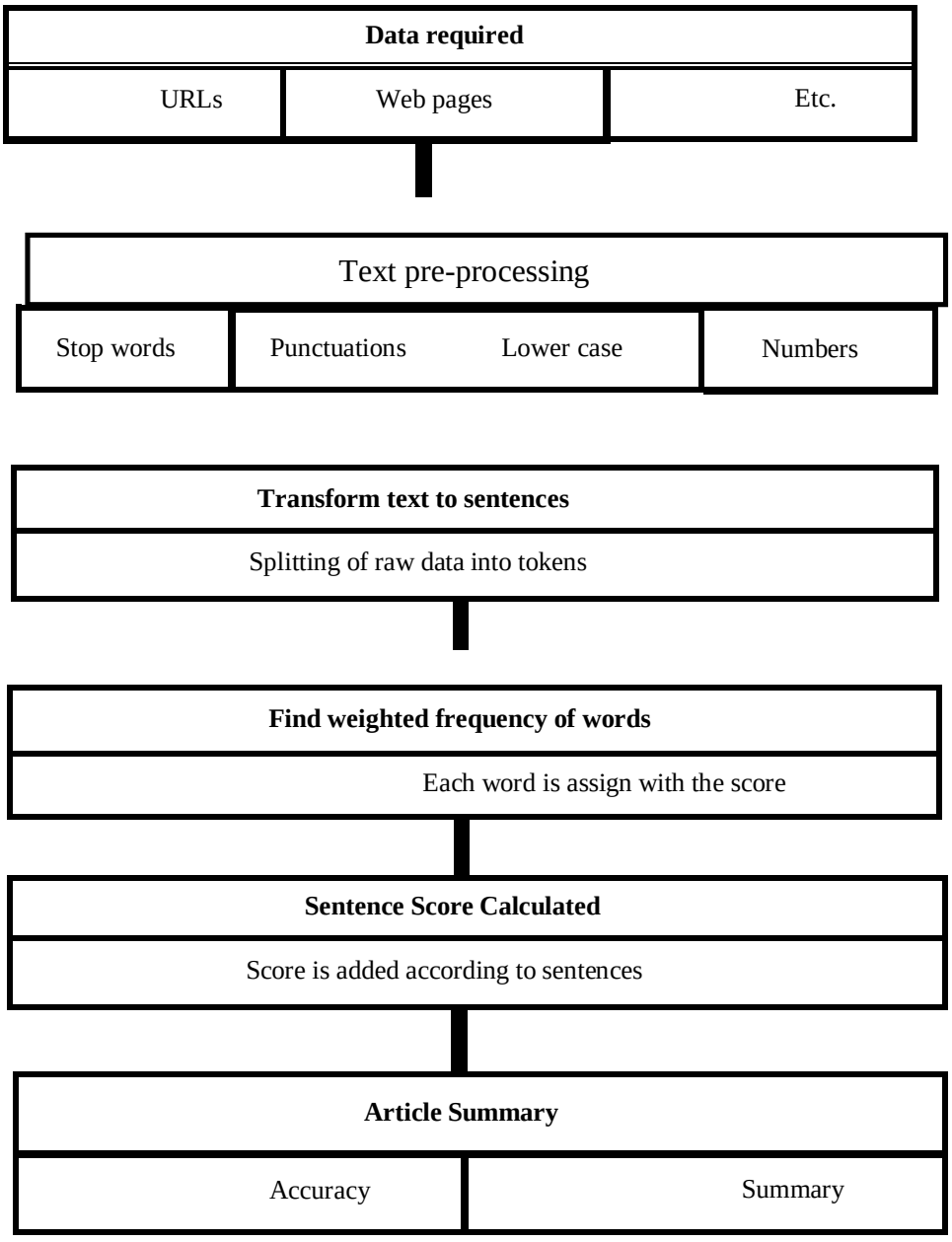


Flowchart 5: Extractive Summarizer

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VII July 2021- Available at www.ijraset.com

C. Procedure for Text Summarizer

Summarization of text involves a number of steps where in urls are taken and the output is given in the form of summary as shown inflowchart 6.



Flowchart 6: Procedure of text summarizer

D. Data Required

Summarization any URL, web page or Wikipedia then first obtain the URL data from the internet and this processor is called web scraping. In python, the Beautiful soup library is require to be install for web scraping. All the web pages containing dataare fetched by this library.

To use this library first it is need to be install and then import which can be done by command ….Pip install beautifulsoup4

Import BS4 as BS

After scraping, data pre-processing has to be done on the extracted text

E. Text Pre- Processing

The raw data highly inconsistent and contain repetitive words. This is the reason why pre-processing the raw data is integral part of preparing the data. The words which add no value to the meaning to the sentence are deleted and make use of valuabledata. Some cleaning steps are lower case, correct the spelling, removal of punctuation, removal of stop words, removal of URLs, removal of numbers, removal of whitespace, removal of hashtags.

F. Transform Text To Sentences

Splitting of raw data which is phrase, sentence, paragraph into smaller chunks is called tokens and this process is called tokenization. Separate entities are require so the breakdown of text into sentences and sentences into word is important. Afterthe pre-processing, perform tokenisation as it can provide us structured and formatted data from unstructured data.

G. Find Weighted Frequency Of Words

According to the number of repetition of words a score is provided to the each word which makes some meaning. Initially the number allotted are in integer format. Score of the each word is divided with the highest score available in the data. The new values are obtained in the float.

H. Sentence Score Calculated

Now each word is assigned with a score which is unique. To calculate the score of the sentence, take a sentence and add the weighted frequency assigned to the different word. The higher the score is, the better the sentence.

I. Article Summary

The sentences are assigned with a priority in the decreasing order. Summary of the article can contain the N number ofsentence. To summarize the article the heapq library is used to select the N number of top sentences.
Now a days it is difficult to read the large document and files for an individual.so a text summarizer can become an integral part of lifefor user. This will summarize the whole document in the sufficient number of lines require by the user.
Data taken from: https://en.wikipedia.org/wiki/COVID-19 and the result is shown as a summary of 5 lines in Image 4.

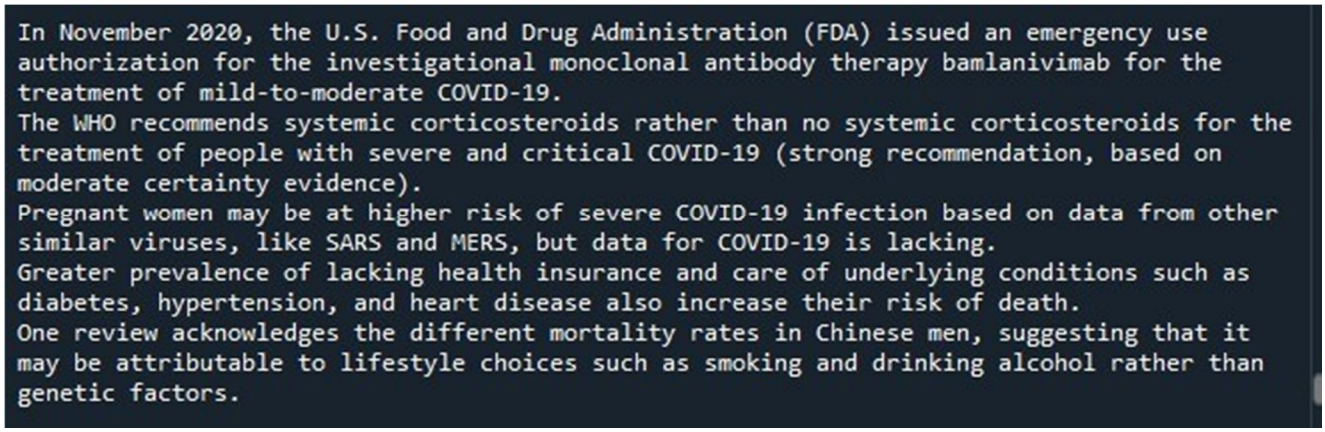


Image 4: Summary of URL in 5 lines

IX. APPLICATIONS,ADVANTAGES, DISADVANTAGES

A. Applications

1) *Brand Monitoring:* Sentiment analysis plays a vital role in helping the company to know about customer's reviews on brand of the product so that improvements can be made if needed and the product manufacturing can be done according to the demand and publicity.
2) *Improving Your Customer Support:* It has been analyzed that 39% people drop a brand, product or service if they have one bad experience. Experience of customer is shared on social media platform and the data needs to be collected at one place leading to improvised customer support.

3) *In Banking:* If a good service is not provided to the customer, the customer might switch the account to another bank. So sentiment analysis helps in knowing the reviews of the customer and knowing the issues faced by them so that it can be resolved.
4) *Competitive Analysis:* Sentiment analysis in business helps in fulfilling the gaps in the market leading to the better performance of business and keeping up the goodwill which would attract more customers and will eventually help in growth of company.
5) *To Know Market Trends:* Knowing the market trends of the company will lead to prediction of the upcoming requirements of market and knowing the future trends which will help in analyzing the field of marketing whose scope is going to be very high in future.
6) *In Stock Markets:* Sentiment analysis have helped in the growth of finance company giving them insights of the real time price analysis and security trading giving detail of every single up or down in stock market
7) *In Government Sector:* To know the issues faced by citizens sentiment analysis plays an important role. It helps government solve the issues faced by the citizen and resolve the similar problems which may create issues in their life. It also helps in being approachable to the government.
8) *In Healthcare:* To get insight of the huge amount of data and to keep a record of the medical responses to varies medicines and with various responses, sentiment analysis plays a very vital role that too without disrupting the daily work of staff and employs
9) *In call Centers:* To keep a record of customer care executive, their reviews, analysis, complains and stored information.
10) *Voice of Customer:* To have the knowledge of personalized customer experience and combine the customer's feedback from mails, surveys, chat, call center etc. to identify the topics and know the repeated concerns.
11) *Voice of Employee:* To reduce the turnovers and increase the productivity, sentiment analysis helps in knowing the subject opinion of all the employees in the absence of human being.
12) *Movie Review:* In knowing the review of a particular movie and all the positive, negative comments and ratings. Positive analysis plays a vital role.
13) *Ola And Uber Review Analysis:* rating on driver's experience, time and speed of reaching the destination and drive experience using sentiment analysis help the customer to choose better type of cab.

B. Advantages

1) Quality products are developed because the companies are able to analysis the large number of reviews of user in less time.
2) New marketing strategies are coming to the way of the user
3) Sentiment analysis is a wide application and it continuously improving
4) Human inter assumptions are not require as patterns and trends are easily identify
5) multi variety and multi -dimensional data is handled
6) customers get an idea about different products and services without reading a thousand number of comments
7) to gain customer's insight, it's cost is effective than convential methods
8) because of analysis, user makes the correct decision
9) Provide summary of urls, web pages, documents, text, etc.
10) create a summary by reducing a text, make text easy to read
11) saves a lot of time of user by providing the summary
12) In general manger good exposure is provided to the user
13) Confusion matrix works on divert dimension which is work on true positive and negative data.

C. Disadvantages

1) It does not understand multiple sense words.
2) To analysis tremendous data it takes some time
3) The comment contain the useful nuances which are unique inn themselves, as sentiment analysis is not a perfect replacement of reading the comments.
4) User use unsuitable perception strategy.
5) It does not save the data which is already decoded
6) Confusion is created among the employs by the dual accountability.
7) This require vertical and horizontal contribution, highly and effectively.
8)

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VII July 2021- Available at www.ijraset.com

X. RESULT

Sentiment Analysis is a technique which is used to classify whether the data is positive or negative. Dataset istrained with different machine learning approaches to extract the feature and tells the context polarity. Natural language processing, logistic regression, text analysis, computational linguistics and study of affective states and subjective text summarizer are used in this model. Different approaches of machine learning are used to make the base model of sentiment analysis. The result comes out in the form of the sentences or number (0 for negative and 1 for positive sentences) by checking the polarity that in turn gives the result by stating that the word, sentences or textis positive or negative. In graphical representation of sentiment analysis the graphs shows positive and negative result in the form of graph depicting the number of words which are positive and negative from a given dataset. The range of checking the polarity can be from 2 – 5000 comments (as 5000 comments are present in the dataset we tookto train our model). The summarizer is used in sentiment analysis by summarizing the text in the required number oflines demanded by the person but if user doesn't choose it then by default it is set to 15 lines (as coded in the code).Thus this model of sentiment analysis is a perfect model which is used as a classifier and summarizer on unstructured data.

XI. CONCLUSION

In today's era, there are millions of comments or text present on the social networking sites. All this data is unstructured and formless. Because of this people face many issues to gain appropriate and accurate information about any company,organizations, product or service. Social networking sites provides more accurate and effective results to people, when the data is more polished. So, this model sentiment analysis or opinion mining is used to classify the text as positive and negative and also provides a summary of whole text which will save some intervals of the user.

In this potential model, the text is fetched from different social networking sites and also the user can input their text accordingly to check the opinions. The model will classify the text into polarity as positive or negative by giving the result as 1 or 0 and the result could also be represented in the form of graph and also there is a text summarizer in whichuser can input the text and obtained with a summary according to the number of line required. This model is working with the accuracy 84.75%. All this can be achieved by using natural language processing, logistic regression and machine learning approaches.

Apart from the model which is already being build there is obviously a scope for making a better model which will fulfil the demands of the customer in future and which will have a high accuracy and better performance. Some of the future trends and directions are already mentioned in the paper. Apart from that is always a need of working on neutral sentences because there is a problem in classifying the sentence containing both positive and negative words. So some work should be done on this. It is a wonderful model to do the analysis of text having different sentiments and thus help us in classifying the text according to the polarity.

XII. ACKNOWLEDGEMENT

We would like to pay our deep gratitude and thanks to Dr. Pramod Kumar (HOD) of CSE department as well as our mentor Ms. Krista Chaudhary, Krishna Engineering College, Ghaziabad for encouraging us to the highest peak and providing us the golden opportunity to make a project on the topic Sentimental Analysis, which also helped us in doing a lot of research and we came to known about so many things about the same.

XIII. FUTURE TRENDS AND DIRECTION

A. The main focus of this is to classify the words as positive and negative and also to plot a graph depicting a number of positive or negative words in a defined number of comments from the dataset. Sentiment also played a vital role in summarizing a large URL into required number of output as result.

B. This model has wide number of applications and it is a wonderful platform for a small business man to a big company to analyses and get the analysis of the crude information and unstructured data that needs to be analyzed and classified for better results. Since this model plays a vital role so there is an enormous future scope in sentiment analysis for making the model more efficient and realistic.

C. Many sentiment analysis model are already present in classifying the data as positive and negative but further classification of positive sentiments as happy, excited, surprised etc. and negative sentiments can be further classified as sad, depressed, angry, disappointing etc.

D. Slang language is used by 85% of people because it saves time and it is easy to write but it becomes important for model to convert those terms into machine level language so that model could analyze it.

E. Long short term memory can be used to give the result of sentiment analysis with the help of emoji.

F. Apart from positive and negative classification, a neutral classification could also be done for the sentences which contain both the positive and negative words.

G. A model of sentiment analysis could also be made for identifying sarcasm and irony words in a sentence and classify accordingly.

H. An extension of sentiment analysis can be made with the help of synonym and antonym inn the dataset which helps in classifying the text in a better and quick way

I. Lastly most of the models of sentiment analysis are fed with a small amount of data for training which does not give 100% accuracy and so the models should be trained with very large amount of data for better accuracy and an efficient model.

REFERENCES

[1] Belghith, A., Obaidat , M .S.: Wireless sensor networks applications to smart homes and cities. In: Obaidat, M.S., Nicopolitidis, P. (eds.) Smart Cites and Homes: Key Enabling Technologies,pp. 17–36. Elsevier, Cambridge (2016)

[2] R. Sharma, S. Nigam, and R. Jain, "Opinion mining of movie reviews at document level," arXiv preprint arXiv:1408.3829, 2014.[17] wrote about the analysis of different movie reviews by using sentiment analysis which helped in identifying the positive and negative comments which people post on social media sites

[3] Deen, M.J.: Information and communications technologies for elderly ubiquitous healthcare in a smart home. Pers. Ubiquit. Comput. 19, 573–599 (2015). https://doi.org/10.1007/s00779-015-0856-x

[4] Guelzim, T., Obaidat, M.S., Sadoun, B.: Introduction and overview of key enabling technologies for smart cities and homes. In: Obaidat, M.S., Nicopolitidis, P. (eds.) Smart Cites and Homes: Key Enabling Technologies, pp. 1–15. Elsevier, Cambridge (2016)

[5] Hakim, A., Huq, M.S., Shanta, S., Ibrahim, B.: Smartphone based data mMining for fall detection: analysis and design. Procedia Comput. Sci. 105, 46–51 (2017). https://doi.org/10.1016/j.procs.2017.01.188

[6] Hsu, Y.W., Chen, K.H., Yang, J.J., Jaw, F.: Smartphone-based fall detection algorithm using feature extraction. In: CISP- BMEI, pp. 1535–1540 (2016). Go, Alec, Lei Huang, and Richa Bhayani. “Twitter sentiment analysis.” Entropy 17 (2009): 252.

[7] Sharma, Anuj, and Shubhamoy Dey. “A comparative study of feature selection and machine learning techniques for sentiment analysis.” Proceedings of the 2012 ACM research in applied computation symposium. 2012.

[8] Kim, Yoon. “Convolutional neural networks for sentence classification.” arXiv preprint arXiv:1408.5882 (2014).

[9] Severyn, Aliaksei, and Alessandro Moschitti. “Twitter sentiment analysis with deep convolutional neural networks.” Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval. 2015.

[10] Wang, Bo, and Min Liu. “Deep learning for aspect-based sentiment analysis.” Stanford University report (2015).

[11] R. Liang and J.-Q. Wang, ‘‘A linguistic intuitionistic cloud decision support model with sentiment analysis for product selection in E-commerce,’’ Int. J. Fuzzy Syst., vol. 21, no. 3, pp. 963–977, Apr. 2019.

[12] P. Ji, H.-Y. Zhang, and J.-Q. Wang, ‘‘A fuzzy decision support model with sentiment analysis for items comparison in E- commerce: The case study of PConline.Com,’’ IEEE Trans. Syst., Man, Cybern. Syst., vol. 49, no. 10, pp. 1993–2004, Oct. 2019.

[13] D. Zeng, Y. Dai, F. Li, J. Wang, and A. K. Sangaiah, ‘‘Aspect based sentiment analysis by a linguistically regularized CNN with gated mechanism,’’ J. Intell. Fuzzy Syst., vol. 36, no. 5, pp. 3971–3980, May 2019.

[14] Y. Chen, J. Wang, S. Liu, X. Chen, J. Xiong, J. Xie, and K. Yang, ‘‘Multiscale fast correlation filtering tracking algorithm based on a feature fusion model,’’ Concurrency Comput., Pract. Exper. , to be published, doi: 10.1002/cpe.5533.

[15] Y. Chen, R. Xia, Z. Wang, J. Zhang, K. Yang, and Z. Cao, ‘‘The visual saliency detection algorithm research based on hierarchical principle component analysis method,’’ Multimedia Tools Appl., to be published.

[16] R. Xia, C. Zong, and S. Li, "Ensemble of feature sets and classification algorithms for sentiment classification," Information Sciences, vol. 181, no. 6, pp. 1138-1152, 2011/03/15/ 2011.

[17] R. Sharma, S. Nigam, and R. Jain, "Opinion mining of movie reviews at document level," arXiv preprint arXiv: 1408.3829, 2014.

[18] R. Sharma, S. Nigam, and R. Jain, "Polarity detection at sentence level," International Journal of Computer Applications, vol.86, no. 11, 2014.

[19] D. Factiva, "Quick Study: Direct Correction Established Between Social Meidia Engagement and Strong Financial Performance," PR News, 2009.

[20] S. R. Das and M. Y. Chen, "Yahoo! for Amazon: Sentiment extraction from small talk on the web," Management science, vol. 53, no. 9, pp. 1375-1388, 2007.

[21] Altmetric LLP (2017), Fetching detailed article level metrics for an article, https://help.altmetric.com/support/solutions/articles/6000086844-sample-api-response, Retrieved on Dec 19, 2017

[22] Ananiadou, S., Thompson, P., & Nawaz, R. (2013). Enhancing search: Events and their discourse context. In International Conference on Intelligent Text Processing and Computational Linguistics, pp. 318-334. Springer, Berlin, Heidelberg.

[23] Batista-Navarro, R. T., Kontonatsios, G., Mihăilă, C., Thompson, P., Rak, R., Nawaz, R., Korkontzelos, I. & Ananiadou, S. (2013). Facilitating the analysis of discourse phenomena in an interoperable NLP platform. In International Conference on Intelligent Text Processing and Computational Linguistics, pp. 559-571. Springer, Berlin, Heidelberg.

[24] Bonaccorsi, A., Cicero, T., Haddawy, P., & Hassan, S. U. (2017a). Explaining the transatlantic gap in research excellence. Scientometrics, 110(1), 217-241.

[25] Bonaccorsi, A., Haddawy, P., Cicero, T., & Hassan, S. U. (2017b). The solitude of stars. An analysis of the distributed excellence model of European universities. Journal of Informetrics, 11(2), 435-454.

[26] Medha Khuran, et al. “Sentiment Analysis Framework of Twitter Data using Classification”, 5th IEEE International Conference, 2018.

[27] P. Garg, H. Garg, and V Ranga, "Sentiment analysis of the Uri terror attack using Twitter," Computing, Communication and Automation (ICCCA), 2017

[28] K. Lavanya and C. Deisy. "Twitter sentiment analysis using multiclass SVM," Intelligent Computing and Control (I2C2), International Conference on. IEEE,

2017.

[29] Ahuja, Shreya, and G. Dubey, "Clustering and sentiment analysis on Twitter data," 2nd International Conference onTelecommunication and Networks (TEL-NET), IEEE, 2017.

[30] M. Trupthi, S. Pabboju, and G. Narasimha. "Sentiment analysis on twitter using streaming API," Advance Computing Conference (IACC), IEEE 7th International. IEEE, 2017.

[31] M. Bouazizi and T. Ohtsuki, ―A Pattern-Based Approach for Multi-class Sentiment Analysis in Twitter,‖ in Proc. IEEE ACCESS, pp. 20617-20639, 2017.

[32] S. M. Mohammad and S. Kiritchenko, ―Using Hashtags to Capture Fine Emotion Categories from Tweets,‖ in Computational Intelligence, vol. 31, no. 2, 2015, pp. 301– 326.

[33] B. Plank and D. Hovy, ―Personality Traits on Twitter or How to Get 1,500 Personality Tests in a Week,‖ in Proc. of the 6thWorkshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, 2015, pp. 92–98.

[34] Ankur Goel, Jyoti Gautam, Sitesh Kumar, ―Real Time Sentiment Analysis of Tweets Using Naive Bayes,‖ in 2nd International Conference on Next Generation Computing Technologies, Dehradun, 2016.

[35] Aparna Garimella and Rada Mihalcea, ―Zooming in on Gender Differences in Social Media,‖ in Proceedings of theWorkshop on Computational Modeling of People‘s Opinions, Personality, and Emotions in Social Media, 2016

[36] [36]Ahmed, A., Zhu, Z., David, Z., Hsinchun, C., Jay, F.N., 2010. Detecting fake websites: the contribution of statisticallearning theory. MIS Q. 34 (3), 435–461.

[37] Ainur, Y., Yisong, Y., Claire, C., 2010. Multi-level structured models for document-level sentiment classification. In: Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing. MIT, Massachusetts, Association for Computational Linguistics, USA, pp. 1046–1056

[38] Alexander, H., Paul, V.I., Bas, H., Flavius, F., Uzay, K., 2011. Determining negation scope and strength in sentiment analysis. In: Proceedings of the 2011 IEEE International Conference on Systems, Man, and Cybernetics (SMC 2011). IEEE Computer Society

[39] Alexandra, B., Ralf, S., 2009. Rethinking sentiment analysis in the news: from theory to practice and back. In: Troyano, Cruz, Dı´az (Eds.), WOMSA'09, pp. 1–12.

[40] Alexandra, B., Ralf, S., Mijail, K., Vanni, Z., Erik, V.D.G., Matina, H., Bruno, P., Jenya, B., 2013. Sentiment analysis in the news. In: Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)

[41] International Data Corporation. 2017. https://www.idc.com/about/about.jsp. Accessed April 22, 2017.

[42] Aggarwal CC, Zhai C. A survey of text classification algorithms. In: Mining Text Data. Boston, MA: Springer Science+Business Media; 2012.

[43] Rajman M, Besançon R. Text mining-knowledge extraction from unstructured textual data. Advances in Data Science and Classification: Proceedings of the 6th Conference of the International Federation of Classification Societies (IFCS-98) Università ''La Sapienza'', Rome, 21-24 July, 1998. Berlin, Germany: Springer-Verlag Berlin Heidelberg; 1998:473-480.

[44] Kumar A, Sebastian TM. Sentiment analysis on Twitter. Int J Comput Sci Issues. 2012;9(4):372-378.

[45] Pang B, Lee L. Opinion mining and sentiment analysis. Found Trends® Inf Retr. 2008;2(1-2):1-135.

[46] K Indhuja, Raj P C Reghu. "Fuzzy Logic Based Sentiment Analysis of Product Review Documents", IEEE FirstInternational Conference on Computational Systems and Communications, 2014.

[47] Tanvir Ahmad, Mohammad NajmudDoja. "Ranking System for Opinion Mining of Features from Review Documents", International Journal of Computer Science Issues, Vol. 9, Issue 4, No 1, July 2012.

[48] Weishu Hu, Zhiguo Gong, JingzhiGuo. "Mining Product Features from Online Reviews", 7th IEEE International Conferenceon E-Business Engineering, 2010.

[49] SiddharthAravindan and AsifEkbal. "Feature Extraction and Opinion Mining in Online Product Reviews", International Conference on Information Technology, 2014.

[50] JantimaPolpinij, Aditya K. Ghose. "An Ontology-based Sentiment Classification Methodology for Online Consumer Reviews", IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology,Vol. 1, 2008.

[51] B. Pang, L. Lee, "Opinion mining and sentiment analysis", Foundations and Trends in Information Retrieval Vol. 2, No 1-2,2008, pp. 1-135.

[52] B. Liu, "Opinion Mining and Sentiment Analysis", in: Web Data Mining, Data-Centric Systems and Applications, SpringerBerlin Heidelberg 2011, ISBN 978-3-642-19460-3, pp. 459-526.

[53] N. Tetsuya, J. Yi, "Sentiment analysis: capturing favourability using natural language processing", Proceedings of the KCAP03, 2nd International Conference on knowledge capture, 2003.

[54] N.A. Vidya, M.I. Fanany, I. Budi, "Twitter Sentiment to Analyze Net Brand Reputation of Mobile Phone Providers",Procedia Computer Science, Elsevier Volume 72, 2015, pp. 519-526.

[55] L. Banić, A. Mihanović, M. Brakus, "Using Big Data and Sentiment Analysis in Product Evaluation", Information &Communication Technology Electronics & Microelectronics (MIPRO), 36th International Convention 2013.

[56] Delo, C. 2011. Twitter Now Has 100 Million Active Monthly Users: CEO. [online] Available at: http://adage.com/article/digital/twitter-100-million-active-monthly-users-ceo/229688/.

[57] Delua, J. 2011. Big Data Meets Sentiment Analysis. informatica, [blog] 27th June, Available at: http://blogs.informatica.com/perspectives/2011/06/27/big-data-meets-sentiment-analysis/

[58] Deshp, E, M. and Sarkar, A. 2010. BI and Sentiment Analysis. 15 (2), p. 41

[59] Go, A., Huang, L. and Bhayani, R. 2009. Twitter sentiment analysis. Entropy, 17

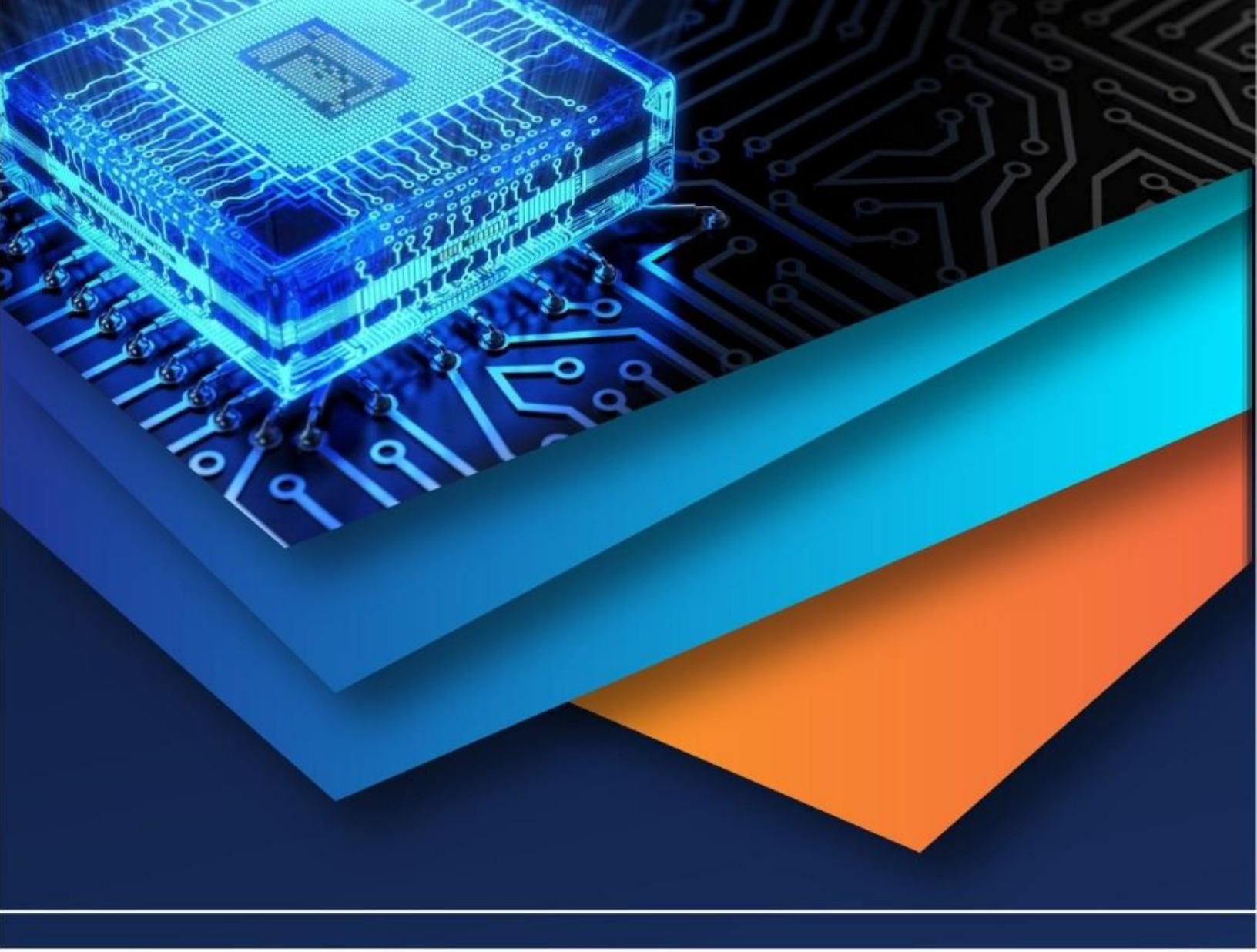
[60] Grimes, S. 2008. Sentiment Analysis: Opportunities and Challenges. [online] Available at:http://www.b-eye-network.com/view/67442.

[61] D. M. Boyd and N. B. Ellison, "Social network sites: Definition, history, and scholarship," Journal of Computer-MediatedCommunication, vol. 13, no. 1, pp. 210–230, 2007.

[62] J. S. Brown and P. Duguid, The Social Life of Information. Boston, MA, USA: Harvard Business School Press, 2002.

[63] H. R. Liangfei Qiu and A. Whinston, "A twitter-based prediction market: Social network approach," ICIS 2011 Proceedings.Paper 5, 2011.

[64] J. Bollen, H. Mao, and X.-J. Zeng, "Twitter mood predicts the stock market," Computer, vol. 1010, no. 3003v1, pp. 1–8, 2010.
[65] H. Mao, S. Counts, and J. Bollen, "Predicting financial markets: Comparing survey, news, twitter and search engine data," arXiv.org, Quantitative Finance Papers 1112.1051, Dec.2011.
[66] Andrea Esuli and Fabrizio Sebastiani, "Determining the semantic orientation of terms through gloss classification", Proceedings of 14th ACM International Conference on Information and Knowledge Management, pp.617-624, Bremen, Germany, 2005.
[67] Bai, and R. Padman, "Markov blankets and meta-heuristic search: Sentiment extraction from unstructured text," Lecture Notes in Computer Science, vol. 3932, pp. 167–187, 2006.
[68] Bing xu, tie-jun zhao, de-quan zheng, shan-yu wang, Product features mining based on conditional random fieldsmodel " ,Proceedings of the Ninth International Conferenceon Machine Learning and Cybernetics, Qingdao, 11-14 July2010
[69] Chaovalit, Lina Zhou, Movie Review Mining: a Comparison between Supervised and Unsupervised Classification Approaches, Proceedings of the 38th Hawaii InternationalConference on SystemSciences – 2005.
[70] Chau, M., & Xu, J. (2007). Mining communities and theirrelationships in blogs: A study of online hate groups. International Journal of Human – Computer Studies, 65(1), 57–70.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY