JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue

JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)

An International Scholarly Open Access, Peer-reviewed, Refereed Journal

SMART SENTIMENT ANALYSIS USING MACHINELEARNING

K.Bansal[1], A. Jain[2],G.K.Singh[3], S.K.Jagtap[4], K.A.Pujari5,A.S.Pandit[6]

[#]Department Of E&TC Engg., Smt. Kashibai Navale College OF Engineering, SPPU , Pune Maharashtra

[1] krsnabnsl@gmail.com

[2] akshayjainptp@gmail.com

[3] gautamsngh900@gmail.com

[4] skjagtap.skncoe@sinhgad.edu

[5] kapujari.skncoe@sinhgad.edu

[6]anuradha.pandit_skncoe@sinhgad.edu

Abstract-. -. In today's era, Social media is a computer-based technology that facilitates sharing ideas, thoughts, and information through virtual networks and communities, including Instagram, Facebook, Twitter, Linkedin, and even Youtube. Commenting has become a new form of written communication as it allows people to avoid being interrupted and express their ideas. People write many comments on social media (on private accounts and beyond them), which can help brands. Tracking social activity can help us to serve customers better. Due to this extensive usage, Tries reflects public sentiment by analysing the ideas expressed in these comments. The proposed method consists of two phases called the testing and training phase. For extracting the features, the sentiwordnet is used. The distance-based classifier is used to classify the features. Finally, analyses of the sentiments are done by polarities classified by the SVM classifier. Sentiment analysis is a kind of datamining. It aims to develop a functional classifier for accurate and automatic sentiment classification of an unknown comment stream. Sentiment analysis classifying comments (sentiments) according to their positive, negative, or neutral opinion.

Keywords - Sentiment Analysis, Distance-Based Classifier, Emotion Detection.

I. INTRODUCTION

Sentiment Analysis (SA) is one of the most extensively researched Natural Language Processing (NLP) and Machine Learning implementations (ML). This field has exploded since the introduction of Web 2.0. Thanks to the Internet, people can now express their thoughts, ideas, and sentiments regarding products, people, and life. Consequently, the Internet has evolved into a large repository of textual data with strong opinions. Sentiment aims to retrieve crucial information about public opinion that may be utilised to make informed corporate decisions, election strategy, and product consumption. The sentiment (SA) is one of the most extensively researched Natural Language Processing (NLP) and Machine Learning implementations (ML). This field has exploded since the introduction of Web 2.0. Thanks to the Internet, people can now express their thoughts, ideas, and sentiments regarding products, people, and life. Consequently, the Internet has evolved into a large repository of textual data with strong opinions. Sentiment aims to retrieve crucial information about public opinion that may be utilised to make informed corporate decisions, election strategy, and product consumption.

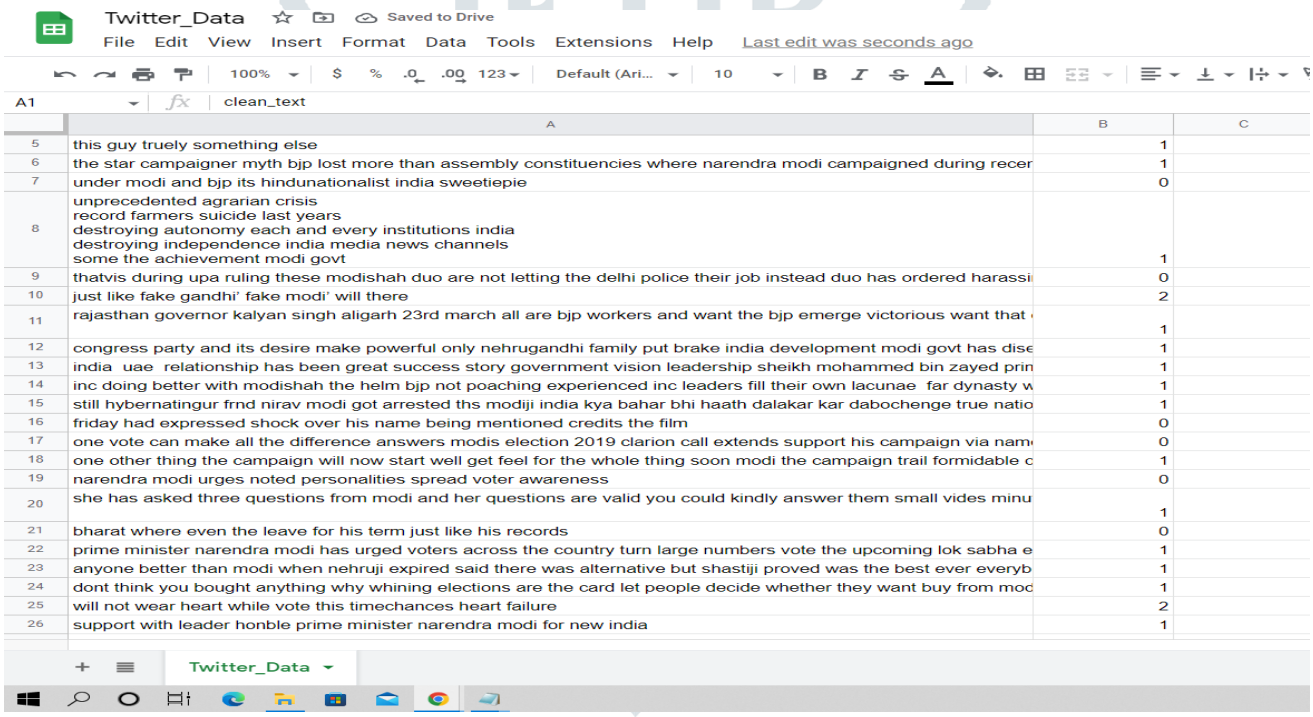
The Internet has recently spurred several government-related operations. Social media is being used to unite people to organise large-scale protests and oppose injustice. On the darker side, social media is used to rally individuals to support an ethnic group or a socioeconomic class, resulting in a significant loss of life. As a result, Sentiment Analysis software that can follow these tendencies and, if possible, interfere is required.

New types of networking, such as microblogging and instant messaging, have evolved and become prevalent in the previous decade. While the quantity of information transmitted by Flipkart and messages is limitless, these brief conversations are frequently used to discuss people's thoughts and feelings about what's going on globally. Flipkart and texts are much shorter than posts, each consisting of a sentence or a headline. Misspellings, jargon, new words, Url, even generic-specific vocabulary, and abbreviations like RT to "re-tweet" and # hashtags for Flipkart posts abound in the vocabulary. Another importance of social media data, such as that seen on Flipkart, would be that it offers detailed, required information on the individuals involved in the interaction.

For example, Facebook has a sizable emoji and textual collection. Twitter, for example, keeps track of who supports whom, while Flipkart and its tags provide debate information.

II. SENTIMENT TRAINING DATA SET

For Sentiment Analysis classification, Training is to be done for our system to understand whether the given text is Positive, Negative and Neutral. There is the text of the dataset attached below, which shows positive, negative, and neutral. There are 4500 texts each for testing and training and three categories: positive, neutral, and negative.



	A	B	C
5	this guy truely something else	1	
6	the star campaigner myth bjp lost more than assembly constituencies where narendra modi campaigned during recer	1	
7	under modi and bjp its hindunationalist india sweetiepie	0	
8	unprecedented agrarian crisis record farmers suicide last years destroying autonomy each and every institutions india destroying independence india media news channels some the achievement modi govt	1	
9	thatvis during upa ruling these modishah duo are not letting the delhi police their job instead duo has ordered harassi	0	
10	just like fake gandhi' fake modi' will there	2	
11	rajasthan governor kalyan singh aligarh 23rd march all are bjp workers and want the bjp emerge victorious want that	1	
12	congress party and its desire make powerful only nehrugandhi family put brake india development modi govt has dise	1	
13	india uae relationship has been great success story government vision leadership sheikh mohammed bin zayed prin	1	
14	inc doing better with modishah the helm bjp not poaching experienced inc leaders fill their own lacunae far dynasty w	1	
15	still hybernatingur frnd nirav modi got arrested ths modiji india kya bahar bhi haath dalakar kar dabochenge true natio	1	
16	friday had expressed shock over his name being mentioned credits the film	0	
17	one vote can make all the difference answers modis election 2019 clarion call extends support his campaign via nam	0	
18	one other thing the campaign will now start well get feel for the whole thing soon modi the campaign trail formidable c	1	
19	narendra modi urges noted personalities spread voter awareness	0	
20	she has asked three questions from modi and her questions are valid you could kindly answer them small vides minu	1	
21	bharat where even the leave for his term just like his records	0	
22	prime minister narendra modi has urged voters across the country turn large numbers vote the upcoming lok sabha e	1	
23	anyone better than modi when nehruji expired said there was alternative but shastiji proved was the best ever everyb	1	
24	dont think you bought anything why whining elections are the card let people decide whether they want buy from moc	1	
25	will not wear heart while vote this timechances heart failure	2	
26	support with leader honble prime minister narendra modi for new india	1	

Fig 1: Text dataset

III. LITERATURE SURVEY

Isah, Haruna, P. Trundle and D. Neagu proposed social media analysis for Product Safety using Text Mining and Sentiment Analysis and introduced a framework for tracking and harnessing experiences and views using sentiment analysis and text mining A Naive Bayes classifier is used to get baseline results for measuring other classifiers. They have also provided details on work in growth, product comparison and initial brand results indicating the utility of the sentiment analysis and textmining on the social media data.[1].

A. Basant, S. Poria, N. Mittal, A. Gelbukh and Amir Hussain designed a concept-level Sentiment Analysis with Dependency-

based Semantic Parsing: A Novel Approach Project concentrated on fundamental issues of the sentiment analysis tasks. Mainly, they used concepts as features and introduced a concept extraction algorithm based on a new concept parser method to extract semantic features that use semantic relationships among words in natural language text. The experimental result expresses that their proposed method for sentiment analysis outperforms existing methods[2].

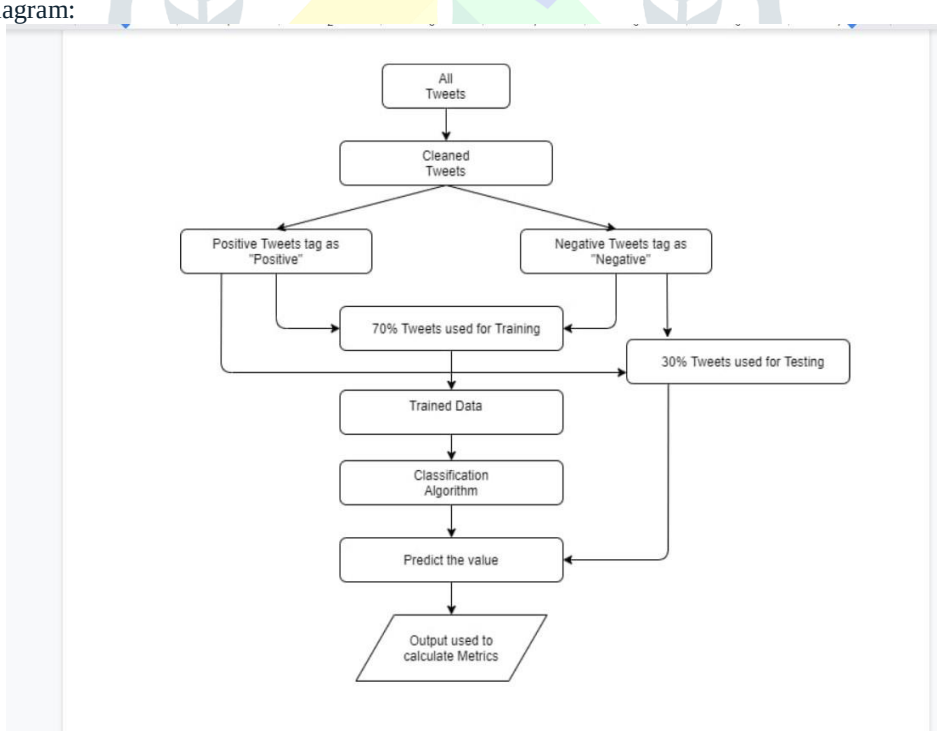
Haddi, Emma, Xiaohui Liu and Yong Shi proposed the role of Text Preprocessing in the Sentiment Analysis Project and considered sentiments of online movie reviews for analysis. They used the grouping of different pre-processing techniques to decrease the noise in text. They have also used the chi-squared technique to eliminate unrelated features that do not change their orientation. They have reported more experimental results, showing that suitable text pre-processing techniques, including filtering and data transformation, can insignificantly prove the classifier's overall performance[3].

Duwairi, Rehab M and Islam Qarqaz designed an Arabic Sentiment Analysis proposed using Supervised Classification Paper says that sentiment analysis is a procedure during which a specified text's polarity (The Neutral, negative and positive) is determined. Usually, there are two techniques to solve this issue: lexicon-based method and machine learning technique. The author introduced the analysis of Arabic-review sentiments from machine learning perceptions. Total three classifiers are used on the dataset of comments and tweets. Classifiers like SVM, K-nearest neighbour and Naive Bayes classifiers are run on this dataset. The overall result expresses that the SVM provides higher precision while KNN (K=10) provides a higher recall[4].

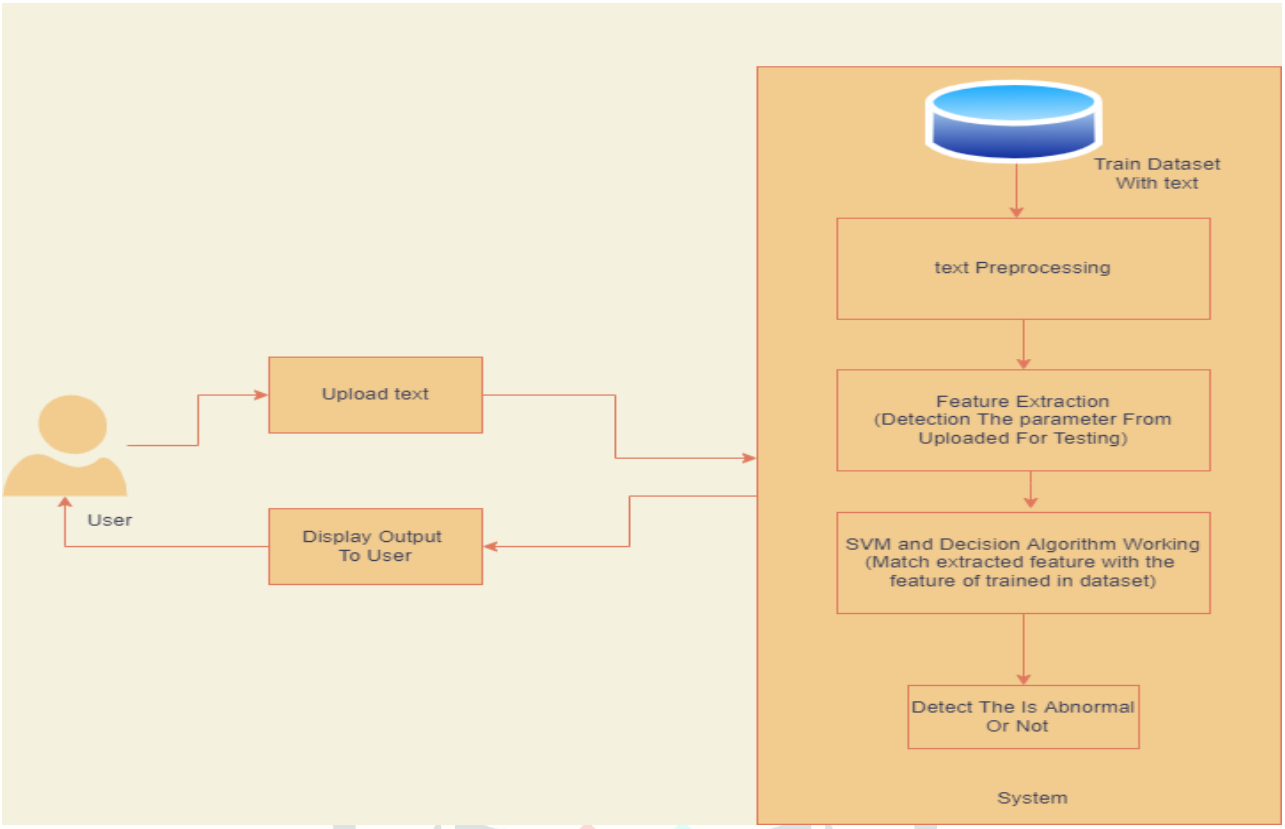
L. Bingwei, E. Blasch, Yu Chen, D. Shen and G. Chen, designed a Scalable Sentiment proposed a Classification for Big Data Analysis using Naive Bayes Classifier introduced a complete and straightforward method for analysing sentiments on the largest dataset using a Naive Bayes classifier and a Hadoop framework. Because of their simple setup, the average accuracy resides less than 82% in every case. An intelligent filter might help increase the overall accuracy[5].

IV. DESIGN AND IMPLEMENTATION

Sentiment detection and classification are based on several methods. It consists of some stages, which are shown below in the form of a block diagram:



In this diagram, it is flowchart that how's things going, from start to end.Here, we showing how all the dataset we imported and perform all the actions like training, testing the model with the help of SVM classifier and at last we giving output.



This diagram shows the pre-prosscessing of the data and the feature Extraction process of the data where the stop words are removed.And used to increase the accuracy since they may overcome the diversity and uncertainty of the vocabulary.

V.APPLICATION

Going forward, a computer will be able to detect human emotions through sensory means and then create an environment that contributes to the overall advancement of human life.

Mainly sentiment analysis is used to rate consumer products and services.

One of the uses of sentiment analysis is for social media monitoring.

Sentiment analysis can also be used for customer feedback.

Sentiment analysis can also use for brand monitoring and reputation management

VI. RESULT AND ANALYSIS

We have trained the Social media text dataset of more than 4000 sentences to give the sentiment of any text. In general, we can say that it gives a good accuracy of 85% and gives a good prediction the sentiment of the text is Positive, Negative or Neutral.

After Registration and login page the main analysis poage will open where the input is given.This diagram shows the positive output by giving positive input.

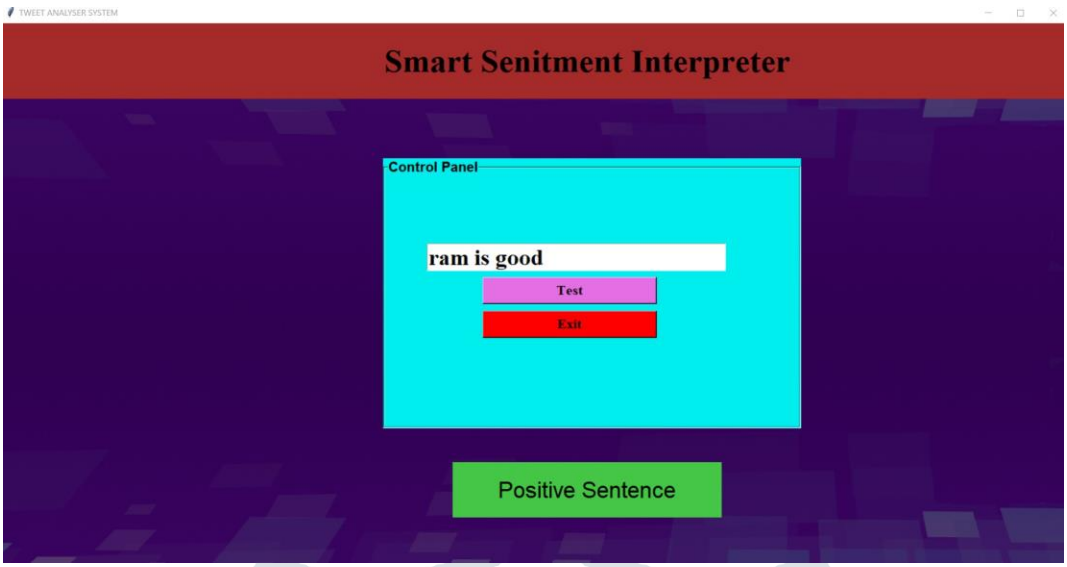


FIGURE 4: SENTIMENT OF A TEXT

This diagram shows the positive output by giving positive input. Showing the analysis of comment given.

FIGURE 5: SENTIMENT OF A TEXT

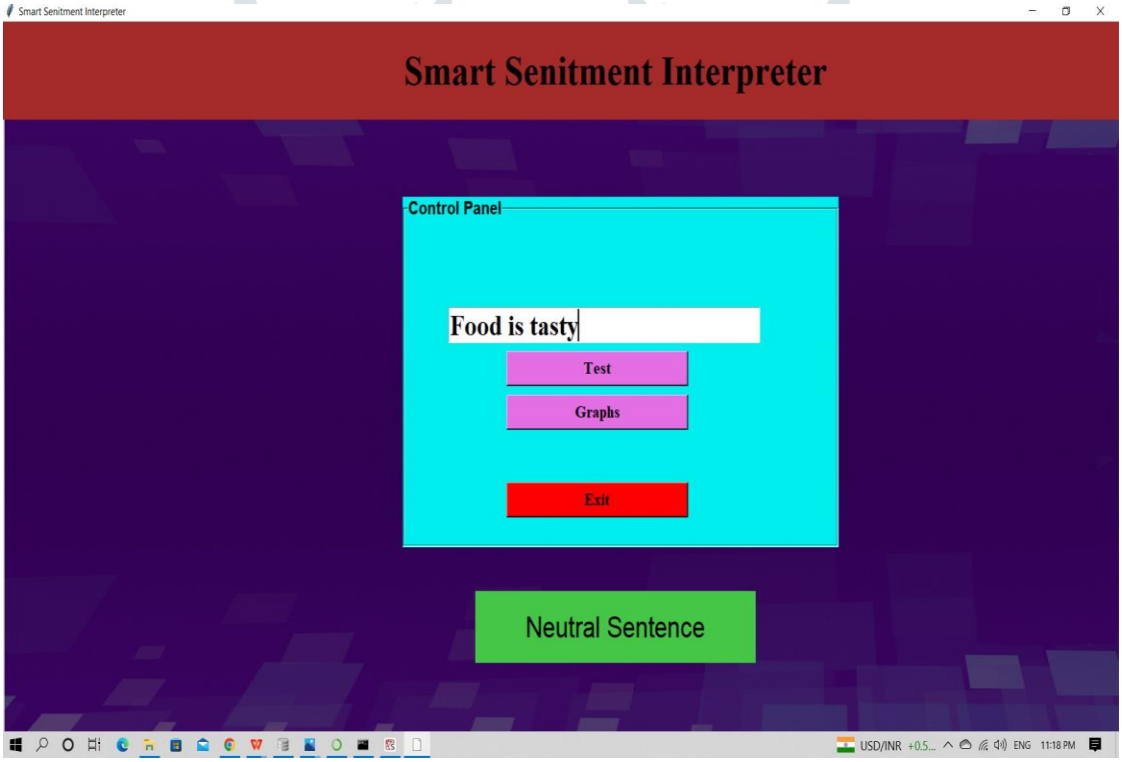


FIGURE 5: SENTIMENT OF A TEXT

The outcome of our model can be seen in the above images (Fig. 4,5). When we give an Input in the form of a text, then our model predicts whether the emotions or opinion of the text is Positive, Negative or Neutral. Here is the above Example of our model accurately predicting the sentiments of the given input.

VII. CONCLUSION

Interest in emotion categorisation has grown in tandem with the increase of social media. Individuals and businesses are increasingly unable to distinguish sentiment from email transparently and understandably. More extensive methodologies are necessary to eliminate misclassifications in the design of prediction models used to interpret feedback. The outcomes of various hybrid sentiment mining approaches are empirically examined on datasets of varying sizes in this study. The hybrid ensemble technique (HEM1) is the most resilient of the methodologies employed for balanced data models I, II, and III according to

different accuracy criteria. According to the findings, a compound blend of unigram, bigram, and trigram performs effectively in all prediction algorithms. Even though SVMs can cope with any amount of data imbalance, the findings imply that data imbalance can influence the employment of SVMs in real-time applications for class prediction. Although data rebalancing is a realistic option, both under and oversampling have drawbacks. The recommended revised bagging process has just been shown to be efficient and superior to many alternative approaches employing different data sampling methods in extensive experiments with benchmark and actual device datasets. PCA is a vital dimension reduction strategy for both balanced and unbalanced datasets when utilising mixed techniques, as per the findings. Other feature reduction methodologies, including such latent Dirichlet distribution, might be researched throughout the upcoming. More studies may be done to evaluate the influence of various domain and region-specific characteristics. Extending sentiment mining to new areas might result in some surprising results.

VIII. FUTURE SCOPE

In the future, further n-gram variants and attribute weighting might be offered to achieve a better level of accuracy. This study mainly focuses on dividing sentiment into two categories: positive and negative (binary classification). A multiclass emotion categorisation system including positive, negative, neutral, and other types might be implemented. This research aims to find Traits that appear as nouns or noun phrases in reviews. Future testing will discover inferred traits. Because ensemble learning procedures take a long time to decide, parallel computing technologies can be investigated as a solution. Because the knowledge obtained by ensembles is unavailable to humans, the inability to examine the findings of ensemble learning approaches is a crucial drawback. As a consequence, improving the interpretability of costumes is a prominent research topic. Future opinion-mining systems will require a more comprehensive and diversified knowledge and understanding of general and commonsense wisdom. This will provide a more precise representation of natural language perspectives as well as a stronger link between multimodal and machine-processable data. More microbe methodologies towards the configuration of intelligent opinion-mining systems capable of handling semantic information, making analogies, discovering new practical knowledge, and detecting, perceiving, and "feeling" emotions will result from combining theoretical theories of emotion with practical engineering goals of analysing feelings in natural language text.

IX. REFERENCES

[1] Isah, Haruna, Paul Trundle and Daniel Neagu, “Social Media Analysis for Product Safety using Text Mining and Sentiment Analysis”, In Computational Intelligence (UKCI) 14th UK Workshop onBayesWorkshop IEEE, pp. 1-7, 2014.

[2] Agarwal, Basant, Soujanya Poria, Namita Mittal, Alexander Gelbukh and Amir Hussain, “Concept-level Sentiment Analysis with Dependency-based Semantic Parsing: A Novel Approach”, Cognitive Computation. Springer, Vol. 7, No. 4, pp. 487-499, 2015.

[3] Haddi, Emma, Xiaohui Liu and Yong Shi, “The Role of Text Preprocessing in Sentiment Analysis”, Procedia Computer Science. Elsevier, Vol. 17, pp. 26-32, 2017.

[4] Duwairi, Rehab M and Islam Qarqaz, “Arabic Sentiment Analysis using Supervised Classification”, International Conference on Future Internet of Things and Cloud (FiCloud). IEEE, pp. 579-583, 2014.

[5] Liu, Bingwei, Erik Blasch, Yu Chen, Dan Shen and Genshe Chen, “Scalable Sentiment Classification for Big Data Analysis using Naive Bayes Classifier”, International Conference Big Data. IEEE, pp. 99- 104, 2013.

[6] Wang, Sida, and Christopher D. Manning. "Baselines and bigrams: simple, good sentiment and topic classification." Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers-Volume 2. Association for Computational Linguistics, 2012.

[7] Cambria, Erik, et al. "New avenues in opinion mining and sentiment analysis." IEEE Intelligent Systems 28.2 (2013): 15-21.

[8] Gautam, Geetika, and Divakar Yadav. "Sentiment analysis of Twitter data using machine learning approaches andsemantic analysis." Contemporary Computing (IC3), 2014 seventh international conference on IEEE, 2014.

[9] Niu, Teng, et al. "Sentiment analysis on multi-view social data." International Conference on Multimedia Modelling. Springer, Cham, 2016.

[10] Cliche, Mathieu. "BB_twtr at SemEval-2017 Task 4: Twitter Sentiment Analysis with CNNs and LSTMs." arXiv preprint arXiv:1704.06125 (2017)

[11] Liu, Bing, Minqing Hu, and Junsheng Cheng. "Opinion observer: analysing and comparing opinions on the web." Proceedings of the 14th international conference on the World Wide Web. ACM, 2005.

[12] McCallum, Andrew, and Kamal Nigam. "A comparison of event models for Naive Bayes text classification." AAAI-98 Workshop on learning for text categorisation. Vol. 752. 1998.

[13] Kotsiantis, Sotiris B., I. Zaharakis, and P. Pintelas. "Supervised machine learning: A review of classification techniques." (2007): 324.

[14] Zhu, Maoran, et al. "Identification of Opinion Leaders in Social Networks Based on Sentiment Analysis: Evidence from an Automotive Forum." (2016).

[15]