Turkish Journal of Computer and Mathematics Education Vol.12 No.10 (2021), 5515-5523

Research Article

Mining Of Customer Review Feedback Using Sentiment Analysis For Smart Phone Product

[1]Dr.P.Suresh, [2]K.Gurumoorthy

[1]Head Department of Computer Science , Salem Sowdeswari College[Govt-Aided], Salem, Tamilnadu, India. Email: sur_bhoo71@rediffmail.com
[2]Research Scholar, Department of Computer Science, Periyar University, Salem, Tamilnadu, India. Email: masterguru84@gmail.com

Article History: Received: 10 January 2021; Revised: 12 February 2021; Accepted: 27 March 2021; Published online: 28 April 2021

Abstract: With the fast growth of e-commerce, large number of products is sold online, and a lot more people are purchasing products online. People also give feedback of product purchased in form of reviews. The user generated reviews for products and services are largely available on internet. To extract the valuable understanding, classification of reviews is required from a huge set of feedback which has converted into positive and negative sentiments. The process of Sentiment Analysis (SA) has mined the attitude, opinions and emotions spontaneously from text, speech and database via Natural Language Process (NLP).It contains feedback review about product, product features or some sentiment emotional views on the product given by the customer. In this research work, feedback from the customer which associated with smart phones is taken from Amazon.com in order to predict the rating of the product given by the user feedback using SA. Feedback review of the customers has been collected fromAmazon.com and this research work had nearly 4000customer feedback reviews based on related categories namely ID of the product, name of the product, name of the brand, Rating, review of the product and vote based on review. This kind of analysis will be helpful for the customers to identify the better product with quick analysis and identify the implicit product perhaps the e-commerce business to improve the sales based on providing offers for particular implicit products.

Keywords: Customer review, feedback, Sentiment Analysis, smart phone, product

1. Introduction

Nowadays huge information, reviews or opinions are getting stored in the websites of social media or e-services in the form of raw data. In recent years, customers prefer to get the product through online. Therefore the prospective buyers choose the right products, large number of data and collected in the form of feedback from the customer. These research work soften provide opinionated words which assist the e-commerce business to recognize areas that needs to be improved. In order to implement with the proper methods, the raw data is required and various methods are either related to adverbs, nouns, verbs or adjectives. However, a recent study shown in SA has the combination of adjectives and adverbs are stronger than adjectives alone but none of the research has focused on all the possible combinations of adverbs, adjectives and verbs. This paper presents the theoretical analysis of some well-known methods or proposal of SA. The data is really helpful, as well as knowledge for businesses looking to understand suggestions on their products or services. In fact, it is beneficial for consumers to be helpful with companies, the ratings and opinion stripped from them. For instance reviews of hotels in a city that helps a consumer search for a good hotel to stay a city. Similarly, product ratings help other people to determine whether the phone is worth buying or not. Similarly, phone reviews facilitate different users choose whether or not the mobile phone is worth for money or not[1][2].This methodology incorporates various algorithms for evaluating and making sense out of the corpus of data. SA uses the manipulation of natural language to remove the particular knowledge from the data [3] [4]. The key component takes consumer reviews as input in NLP system, and then separated by tokenizer into token. A series of characters combined in a text is called a symbol, while a semantic unit for processing is identified. The tokenizer contains punctuation marks, icons, words etc. it can turn a phrase into word level tokens that has been executed to create rules to produce word counts and even ranking is achieved by consumers who assist in an additional degree of quality segregation for the commodity. In this research work, feedback from the customer which associated with smart phones is taken from Amazon.com in order to predict the rating of the product given by the user feedback using SA.

The organizations of this paper is described as follows, section 2 defines the associated review about method based on SA, section 3 defines the proposed methodology based on customer review data collection, data preprocessing, SA and frequency of review rating , and section 4, discusses the conclusions.

2. Literature Review

In this section, many researches have been done using SA. The area of text based classification which was not much research work done to classify the sentences or word related to feedback rating.

Chawla et.al [5] describes various mobile phone reviews based on SA which can be obtained by learning various post given from various numbers of users that can classify the smart phones. Sariet.al [6] discusses to collect feedback on Tokopedia's quality of service on-line analysis over several months of observations. Because of its high-level precision, the Naïve Bayes classification technique is applied, which facilitates large data processing. The outcome showed that the element of reliability and personalization needed more focus because they have a strong negative feeling. Moorjani and Sadath[7] suggest a Continuous Sentiment Analysis(CSA) system for repetitive study of customer emotions emphasizing the intent of one such effort to catch the tone of the message. This "Sentiment Analysis" approach is relatively a recent technique which using NLP to provide meaning to the plentiful data available at hand. Harjeet and Prabhjeet[8] explore a novel approach by trusting the comments on social media to build on a specific topic. The proposed solution includes a list of the words used to construct training dependent on knowledge of positive words and negative words. Originally data is obtained from web networks namely, Amazon, Flipkart, Ebayetc. Along with collecting special attributes from the information gathered and then applying them to vector and value set. This research study is carried out step-by-step, explaining the feedback, based on interpretation of SA. Sowmya et.al[9]proposed methodology used reviews from many customers who visits different hotels and book rooms and order food. This can be achieved using SVM algorithms, logistic regression and Naive Bayes.Bordoloi and Biswas[10]proposes a Machine Learning (ML)model for SA and compares some popular ML approaches in the context of sentiment classification. The classifier efficiency is calculated in terms of precision.Ganagavalliet.al[11]explores how text analysis methods can be used to investigate based on various tweet language patterns and message volumes on twitter into some of the details in a series of posts. The experimental tests reveal that the current classifiers for machine learning are more effective and accomplish better in terms of precision. Fang and Zhan[12] proposed with detailed process descriptions of sentiment polarity categorization. Experiments were done with positive findings on both sentence-level and review-level categorization. Thakur and Srivastava[13] discusses the Long Short Term Memory(LSTM) classifier provides the best results in classifying comments with POS tagged lexicon features into positive and negative review. Muthukumaran and Suresh[14] illustrate that mathematical approaches are frequently mixed with conventional linguistic laws and definitions. LamiaaMostafa[15] discusses the study of emotions relevant to the field of education and Gamification of learning. Naïve Bayes(NB) is the better classifier in which the results showed based on accuracy and also showed better results compared to the disagreement group when performing the 1000 students for testing, the agreement group in learning to use Gamification, that may improve student's evaluation in learning. Sharma and Mansotra [16] suggest to introduce a multimodal sentiment prediction framework from various modal sources namely, images, text and audio that can interpret the projected emotions and combine them to understand the student's community emotions in a classroom. This system includes a digital microphone device that records the student's live video and audio streams during a lecture. Hassan Saif et al.[17] had used lexicon-based approach from twitter posts to implement the SA mechanism. In this paper SentiCircles and lexicon-based methods are proposed which has been described primarily on the logical semantics that expresses the word-oriented sentiment. There are three separate databases which are Stanford Sentiment Gold Standard (STS-Gold), Obama-McCain Debate (OMD) and Health Care Reform (HCR) is tested by the proposed process. Bac Le et al. [18] suggested a twitter-data SA method. This paper describes the few methods to do text-based SA using lexicon-based methods. They dealt with three separate databases, Alchemy Rest, Open Calais and Zemanta. Apple et.al [19] describes the approach of hybrid model of SA which is based on learning and lexicon. These can define emotions and polarity of the opinion which can be obtained better accuracy of 75%. Mohan et al.,[20] suggested technique of sentiment analysis through the study of restaurant domain customer feedback. In addition, creating the rule base to classifier by predicting the polarity of the review used by priority based algorithm. For incremental instance counts the analysis performs well by K-Nearest Neighborhood (K-NN) create. Suresh and Gurumoorthy [21] suggested Apriori algorithm which is one of the standard algorithm for Association Rule Mining (ARM) that can used to mine frequent item sets and its associate rules. An enhanced apriori algorithm to prune the subset and identify the better frequent item set which identify the better selection for the smartphone that get explicit. Suresh and Gurumoorthy [22] has addressed the research of AI has attained an excellent level with sublevel of ML and deep learning application with a minimal method that is proceeding to concrete future business.

The above study has to identify polarity of words using the analysis of SA techniques and several NLP concept to linguistic the tokenized sentence and words.

3. Research Methodology

Most of the business establishments have done by "Market Basket Analysis" to evaluate the user input feedback on their mobile phone and buyer motive. According to an instance, a person is buying a mobile with best battery consumption feature in the basket. Later, he switches over to better front camera features instead of considering the battery life. Moreover, in the current techniques there is no consideration or intimation why the user or customer had switched over to another feature like front camera features instead of battery consumption feature. The existing techniques may not predict exactly whereas the Implicit Rule Interference algorithm is used to identify the kind of featured mobile that has been purchased by the person based on the basket data alone. To evaluate the explicit and implicit model, the present research work considers smart phone feedback analysis based rule mining with SA. In this work aims to progress a recommendation algorithm that is built from an explicit and implicit analysis based on laws of association. This paper discusses the use of NLP for deep learning as seen in Figure.1.

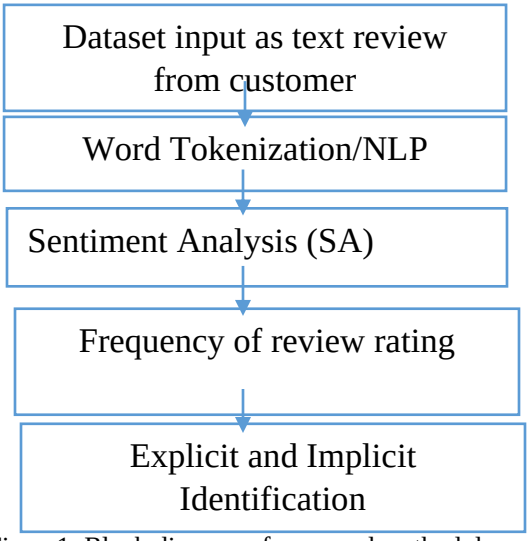

Figure1: Block diagram of proposed methodology

3.1 Input Dataset/Feedback

The data collection employed contains consumer ratings of smartphones obtained from amazon.com. The buyer agrees a recommendation on a scale from 1 to 5 and gives its individual opinion according to the overall experience about the product. For all scores the mean value is determined to attain at the final rating. Other visitors, based on their helpfulness, can also mark yes or no to a review which has been giving value to the review and reviewer. In this work, we examined more than the 4000 user experience reviews on mobile phones sold on Amazon.com. The dataset collected from "http:/www.kaggle.com" are outlined in table.1 and 2 with the following attributes from Amzon.com's based on the category Cell Phone.

Table 1: Features Involved in the Data Set

Feature	**Description**
Prod_ ID	Identification number of the product
Prod_ Name	Name of smart phone
Brd_ Name	Company name
Rating	Customer rating scale between 1 to 5
Rvew	Customer feedback provided for every smart phone
Rvew_vote	Number of people providing vote who found the review helpful

Table 2: Feedback and rating from online shopping customer based on Product ID

S.No	**Prod_ID**	**Prod_ Name**	**Brd_ Name**	**Rating**	**Rvew**	**Rvew_vote**
1	PID1	Galaxy SPHD700	Samsung	5	I feel so LUCKY to have found	1

2	PID2	Apple iPhone 5c 16 GB Green	Apple	5	Awesome description, condition and seller VERY PLEASED, THANK YOU.	0
3	PID3	Lenovo A850	Lenovo	4	Very good phone, excellent hardware, very good performance and high compatibility	6
4	PID4	Nokia Asha 302	Nokia	3	Shipped quickly and was exactly what I expected	2
5	PID5	BlackBerry Bold 9650	BlackBerry	4	I liked	0
6	PID6	Huawei Honor 6 Plus	Huawei	5	Excellent.	1
7	PID7	5530 XpressMusic	NOKIA	4	The Phone is pretty good	2
8	PID8	8330 BlackBerry	BlackBerry	4	Good.	0
9	PID9	ACER LIQUID E700 TRIO	Acer	4	It's work well	1
10	PID10	Acer Liquid M220	Acer	5	This is the best budget phone	0

3.2 Data preprocessing-word Tokenization and NLP

Once the review text gets imported it is considered as each customer feedback which gets extracted in terms of required tokenization and produced a needed relationship by NLP. However, this process perform through NLP has assisted in comprehensively categorized as the controlled program of natural language which may apprehensively in connection among computer and human language from the computer science with deep learning. The large quantity of text has been analyzed and handled with predictive analysis using NLP. This is a part of deep learning technique with some characteristics such as stemming; chunking data and stop words removal get utilized. The beneficial of NLP in creating a sentiment words by segregating the words in term of noun and even the paragraph and sentences are tokenized and chunked in determining the sentences as positive and negative. Thus, the NLP also used for translator in translating one language to needed language. It may generate low noise which may lead to robust data. NLP assist in feeding customer feedback as an input and it get divided into each token using tokenizer. A sequence part of character has combined with organization involving punctuation marks, symbols, special characters, words, etc. that has added in modifying a sentence into various words based on word tokenization. This research has focused with Natural Language Tool Kit (NLTK) is measured and applied with python which get assisted and interpreted to predefine the structure of sentences along its meaning. According to this proposed method, the research need to be modifying the customer feedback represented in the attribute of review text along with text of unstructured to the structured data. At first, the data from part of speech is used in all NLP task for finding noun, verb, adjective and root to each word over the sentence from review text. This proposed chunking NLP algorithm assists in identifying the sentiment words present in the review text such as adverb, noun and adjective that are utilized as a feature which may represent high accuracy.

Chunking NLP Algorithm for extracting the required terms
Step: 1 Get extracted in term of required tokenization using Defextract_NN
Step: 2 assign grmr = r “”
Step: 3 NBAR: # Adjectives and Nouns, Noun during terminated words
{<NN.*>*<NN.*>}
Step:4 NP:{<NBAR>}# connected with, above, in/of/etc..
Step:5 identifying the opinion words present in the review text such as adverb, noun and adjective {<NBAR><IN><NBAR>}
“”
Step:6 parsing the partial syntactic structure of a sentence Chkr as nltk.RegexParser(grmr)
om = set()
Step:7returing over tokenization of specified character cnk as tokrnizerfractory for this chunker
Step:8 for tree in cnk.subtrees(filt = lambda t:t.label() = = ‘NP’);
om.add(‘’.join([child in tree.leaves()for child[0]]))
returnom
Step:9 sub= []
forsentenc in data;
#extract predefine the structure of sentences along its meaning NN (sentenc)
Step:10concat method in the string class as Sub.append(extract_NN(sentenc))
print (sub)

3.3 Sentiment Analysis
Customer feedbacks are evaluated in this proposed work using SA which has been received from the website. When before charging the money, the customer needs feedback about the company. At the moment to read all the suggestions has not possible which was provided by the customer in the website. However all kind of product analysis or feature analysis present in the companies are available with new information. Therefore, all kind of essential inputs have been provided from the customers are possibly to be missed. Thus the organized review rating frequency has assisted to resolve previous challenges. Then word count has been calculated from the extraction of all tokenized words based on SA. These can be obtained by deep learning. The easiest way to interpret the reviews using an SA along with word count which is to figure out the feedback rating. Hence, the rating can be based on the reviews given by the customer. After the SA output has been received, the consumer should make a quicker and minimized attempt to read the feedback as the decision.The analysis terms are equipped using Document Frequency (DF) or Inverse Document Frequency (IDF) have been used for determining word count are displayed in table 3.
Algorithm for Sentiment Analysis
Step 1: CountVectorizer() converts a collection of text documents to a matrix of token counts
Step 2: assign a shorter name for analyze
Step 3: analyzer = vectorizer.build_analyzer() #which tokenizes the string
Step 4: tokenize the string and continue, if it is not empty
If analyzer(s):
d = {}
Step 5: Find counts of the vocabularies and transform to array
Step 6: item() transforms the dictionary’s (word, index) tuple pairs
Step 7: For k, v in vc.items()
D →index:word
For index, i in enumerate (w[0]);
C →word :count
Return C
Step 8: dF1 = dF→ document frequency
dF[‘Rating’] . value_counts(). To_frame()
Step 9: color dF1[Rating] #Rating 4 higher→ positive, Rating 2 lower→negative, Rating 3 → neutral

Table 3: Calculate word count

	PROD_ID	Product Name	Brand Name	Rating	Reviews	Review Votes	Word Counts
0	PID1	Galaxy SPHD700	Samsung	5	I feel so LUCKY to have found this used (phone...	1.0	{'after': 1, 'again': 1, 'all': 1, 'an': 1, 'a...
1	PID1	Galaxy SPHD700	Samsung	4	nice phone, nice up grade from my pantach revu...	0.0	{'an': 1, 'and': 2, 'android': 1, 'are': 1, 'b...
2	PID1	Galaxy SPHD700	Samsung	5	Very pleased .	0.0	{'pleased': 1, 'very': 1}
3	PID1	Galaxy SPHD700	Samsung	4	It works good but it goes slow sometimes but i...	0.0	{'but': 2, 'goes': 1, 'good': 2, 'it': 3, 'its...
4	PID1	Galaxy SPHD700	Samsung	4	Great phone to replace my lost phone The only...	0.0	{'adjust': 1, 'again': 1, 'am': 1, 'but': 1, '...

These reviewed word count to form a selected emotional words which has been vectorized and gets associated with particular customer. The positive and negative feedback based on selected words = [‘best’, ‘good’, ‘love’, ‘amazing’, ‘impressive’, ‘super’, ‘glad’, ‘fantastic’, ‘funny’, ‘wonderful’, ‘extraordinary’, ‘awesome’, ‘bad’, ‘boring’, ‘unhappy’, ‘never’, ‘upset’, ‘sad’, ‘terrible’, ‘disappointment’, ‘poor’, ‘confused’, ‘hard’, ‘hate’] are illustrated in figure 2.

```
Selected Words
great       8919
love        3305
happy       1150
awesome      921
bad          787
dtype: int64

ProductID
PID74     1448
PID420    1237
PID56     1069
PID37     1007
PID64      919
Name: PROD_ID, dtype: int64
```

Figure 2: Plot the frequency of sentimental words

In this work the frequency of the review rating from 1 to 5. The following ratinglevels namely extremely positive, positive, neutral, bad and very bad are used.For example, in a review very positive= 5 star and very bad = 1 star which are mapped onto 5 star ranking. It indicates the overall rating scale with corresponding product_IDare illustrated in figure.3

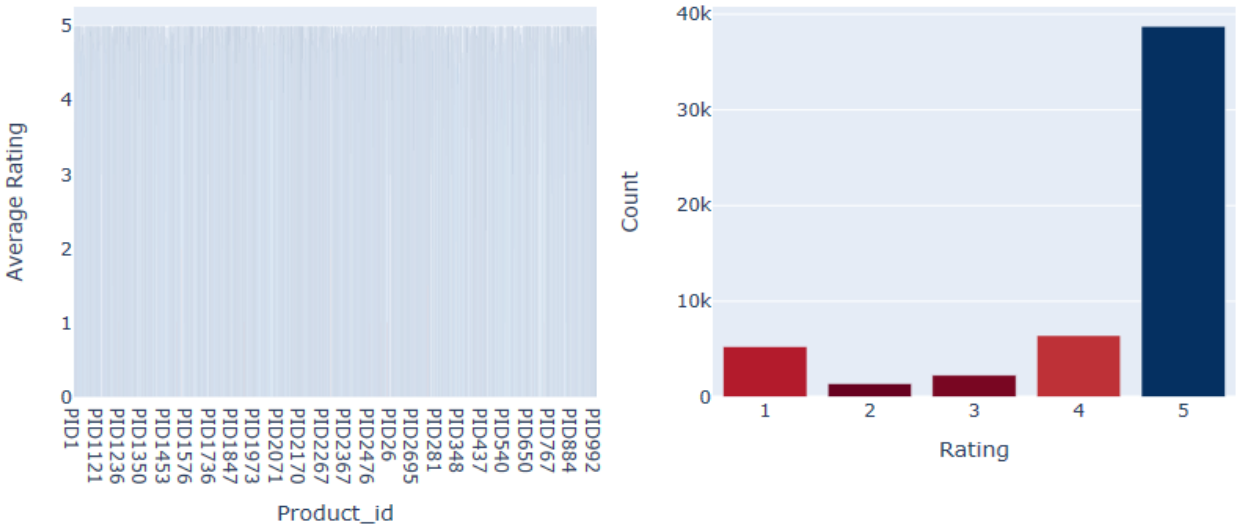

Figure 3: Average rating based on product_id

In this 3D map, the purchase rating level has illustrated the evaluation of both the explicit and implicit relationship from the feedback of customer opinion. The following figure.4 X axis represents the product_ID, Y axis represents rating, and Z axis represents number of purchases, based on the average value buyer purchase number of items being identified with combination of products and item infrequent.

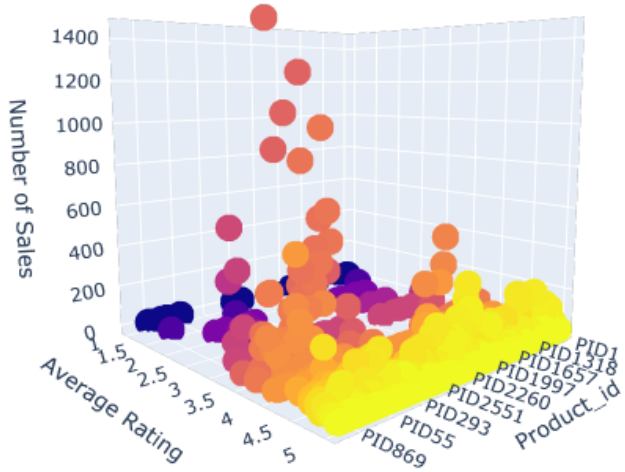

Figure 4: 3D plot of the number of sales, average rating and product_id

This kind of analysis may be helpful for e-commerce business to improve the sales and identify the implicit product based on providing offers for particular implicit products.

4. Conclusion

The major challenge for the customer is to select the right mobile phone while choosing online shopping due to product features that can't be able to justify. However, the customer feedback and rating may recommend the phone quality to the customer but there is a lag in identifying the exact product feature quality based on rating. Many businesses claim that their business success depends solely on customer satisfaction. Therefore, scientists are encouraged to find better solutions for SA. Consequently, this work has focused on addressing the needs of customer feedback with their review text using SA. This SA has deal with NLP which assists to tokenize for making word counts. Therefore, the word count is compared with the words of sentiments along with the customer rating based on product ID to determine the better smart phone. This method of research has to be boosting the sales by defining the indirect product and offering approach for the various implicit products. In future research work, proposed system is

made to evaluate the train and test dataset of SA with various classification techniques for justifying the accuracy level of qualified model.

Reference

1. U. Kumari, A. K. Sharma, and D. Soni, "Sentiment analysis of smart phone product review using SVM classification technique," *2017 Int. Conf. Energy, Commun. Data Anal. Soft Comput. ICECDS 2017*, pp. 1469–1474, 2018.
2. M. Shaheen, "Sentiment Analysis on Mobile Phone Reviews Using Supervised Learning Techniques," *Int. J. Mod. Educ. Comput. Sci.*, vol. 11, no. 7, pp. 32–43, 2019.
3. D. Kamalapurkar, N. Bagwe, R. Harikrishnan, S. Shahane, and G. Manisha, "Phone recommender: sentiment analysis of phone reviews," *Int. J. Eng. Sci. Res. Technol.*, vol. 6, no. 5, pp. 212–217, 2017.
4. P. Pankaj, P. Pandey, M. Muskan, and N. Soni, "Sentiment Analysis on Customer Feedback Data: Amazon Product Reviews," *Proc. Int. Conf. Mach. Learn. Big Data, Cloud Parallel Comput. Trends, Prespectives Prospect. Com. 2019*, pp. 320–322, 2019.
5. Shilpi Chawla, Gaurav Dubey and Ajay Rana, "Product Opinion Mining Using Sentiment Analysis on Smartphone Reviews", *International Conference on Reliability, Infocom Technologies and Optimization (ICRITO)*, Sep. 20-22, 2017.
6. P. K. Sari, A. Alamsyah, and S. Wibowo, "Measuring e-Commerce service quality from online customer review using sentiment analysis," *J. Phys. Conf. Ser.*, vol. 971, no. 1, 2018.
7. G. Moorjani and L. Sadath, "Sentiment analysis-A tool for data mining in big data analytics," *Int. J. Innov. Technol. Explor. Eng.*, vol. 8, no. 9, pp. 2125–2131, 2019.
8. H. Kaur and P. Kaur, "Dimensionality reduction in sentiment analysis using colony–support vector machine," *Int. J. Innov. Technol. Explor. Eng.*, vol. 8, no. 8, pp. 2791–2797, 2019.
9. K. Sowmya, K. Monika, M. Radha, and V. Vijay Kumar, "Customer review rating analysis using opinion mining," *Int. J. Innov. Technol. Explor. Eng.*, vol. 8, no. 7, pp. 2444–2447, 2019.
10. M. Bordoloi and S. . Biswas, "Sentiment Analysis of Product using Machine Learning Technique: A Comparison among NB, SVM and MaxEnt," *Int. J. Pure Appl. Math.*, vol. 118, no. July, pp. 71–83, 2018.
11. K. Ganagavalli, A. Mangayarkarasi, T. Nandhinisri, and E. Nandhini, "Sentiment analysis of twitter data using machine learning algorithm," *J. Comput. Theor. Nanosci.*, vol. 15, no. 5, pp. 1644–1648, 2018.
12. X. Fang and J. Zhan, "Sentiment analysis using product review data," *J. Big Data*, vol. 2, no. 1, 2015.
13. Priyanka Thakur and Dr. Rajiv Shrivastava, " Sentiment Analysis of Tourist Review using Supervised Long Short Term Memory Deep Learning Approach ", IJIRCCE, Vol. 7, Issue 2, 2019.
14. S.Muthukumaran, Dr.P.Suresh "Text Analysis for Product Reviews for Sentiment Analysis using NLP Methods", International Journal of Engineering Trends and Technology (IJETT), V47(8),474-480 May 2017.
15. Lamiaa Mostafa , "Student Sentiment Analysis Using Gamification for Education Context" , Springer Nature Switzerland AG 2020, AISC 1058, pp. 329–339, 2020. https://doi.org/10.1007/978-3-030-31129-2_30.
16. Archana Sharma and Vibhakar Mansotra, "Multimodal Decision-level Group Sentiment Prediction of Students in Classrooms", IJITEE, ISSN: 2278-3075, Volume-8 Issue-12, October 2019.
17. H. Saif, Y. He, M. Fernandez, and H. Alani, "Contextual semantics for sentiment analysis of Twitter" Information Processing & Management, 2016, pp. 5-19.
18. B. Le, and H. Nguyen, "Twitter sentiment analysis using machine learning techniques", In Advanced Computational Methods for Knowledge Engineering, pp. 279-289. Springer, Cham, 2015.
19. O. Appel, F. Chiclana, J. Carter, and H. Fujita. "A hybrid approach to sentiment analysis", In IEEE Congress on Evolutionary Computation (CEC), pp. 4950-4957. IEEE, 2016.
20. Aishwarya Mohan, Manisha.R, Vijayaa.B, Naren.J, "An Approach to Perform Aspect level SentimentAnalysis on Customer Reviews using SentiscoreAlgorithm and Priority Based Classification", (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (3) , 2014, 4145-4148.
21. Dr.P.Suresh and K.Gurumoorthy, "Identification of explicit smartphone features using apriori algorithm", International Journal of Advanced Science and Technology, vol-29, No-3, PP-856-869, 2020.
22. Dr.P.Suresh and K.Gurumoorthy, "Supervised Machine Learning algorithm using Sentiment Analysis based on customer feedback for smart phone", International Journal of Emerging Trends in Engineering Research, Volume-8, No.8,2020.

23. Asraf Yasmin, B., Latha, R., & Manikandan, R. (2019). Implementation of Affective Knowledge for any Geo Location Based on Emotional Intelligence using GPS. International Journal of Innovative Technology and Exploring Engineering, 8(11S), 764–769. https://doi.org/10.35940/ijitee.k1134.09811s19
24. Muruganantham Ponnusamy, Dr. A. Senthilkumar, & Dr.R.Manikandan. (2021). Detection of Selfish Nodes Through Reputation Model In Mobile Adhoc Network - MANET. Turkish Journal of Computer and Mathematics Education, 12(9), 2404–2410. https://turcomat.org/index.php/turkbilmat/article/view/3720