Supervised Machine Learning algorithm using Sentiment Analysis based on customer feedback for smart phone product

K Gurumoorthy[1], Dr.P Suresh[2]
[1]Research Scholar, Department of Computer Science, Periyar University, Salem, Tamilnadu, India,
dr.gurumoorthy_k@rediffmail.com.
[2]Research Supervisor, Head Department of Computer Science, Salem Sowdeswari College[Govt-Aided], Salem,
Tamilnadu, India

ABSTRACT

Nowadays, there are several e-commerce giants namely Amazon, Flipkart and e Bay which have been built on forefront with Machine Learning (ML) that certainly manages the world. This is particularly seen currently in the sector of e-commerce and provided its consequences. However, the ML application in general and in its advancement has produced deep impact in our life style. In the e-commerce field, the technique of ML is implemented and accomplished with better results. Therefore, 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. Sentiment Analysis (SA) is a technique used for analyzing data that has expressed in terms of text and even for discovering sentimental and emotional content from the text. SA is a process of deep learning method which is an advancement of ML, utilized for finding information with positive, neutral and negative from the review of customers. The identification process and analysis of customer feedback is important using Natural Language Processing (NLP), is said to be sentiment analysis. In this research, the dataset has considered about smart phone products with 4000 customer's feedback and ratings for Prod_ID as input. It is utilized for the analysis based on related categories namely Prod_ID, Prod_name, Brd_name, Rating, Rvew and Rvew_vote. The performance evaluation of classifier can be measured in terms of the accuracy. Based on the results it is observed that Neural Network (NN) has produced better accuracy than other classification algorithm with 98%.

Key words: Machine learning algorithm, Sentiment Analysis (SA), smart phone.

1. INTRODUCTION

The web trading or electronic trading is also mentioned as E-commerce business which refers to buying and selling of items from merchandise. Moreover, e-commerce has managed their web utilization by cash exchange and even provides data for executing these exchanges whereas the consequent utilization in referring the physical product sales over the web is said to be internet business. In order to challenge the retailer business, the utilization of e-commerce has improved day-to-day [1]. Moreover, the data which is in the form of text are complaints, tweets in social media, feedback, reviews from customer, discussion etc. whereas there are several data that has been generated daily with incremental of human interaction over social media. Nowadays, people purchase products through online, booking hotels, tickets for travel and cab online which generates data, based on customer reviews. However, the SA assists for identifying the review as positive, neutral or negative. Hence, this type of analyzing data may assist the business by recognizing the perspective of customer for brand strategies [2].SA can observe right decision basis on user feedback. People will buy some commodity available on the internet and then they use it. They write a review of the specific product afterwards if it is good or poor, which helps others to purchase or not after reading that post. So, these reviews increase product sales in the market. Researchers are constantly working on ML to solve the certain problems immediately in a minimum of time. Under computer science, ML, deep learning, and artificial intelligence arrives to train systems and develop computing to work smart [2]. In supervised ML technique there are two types of datasets, namely the training and testing data set. Supervised ML can be used for classification purposes to rate emotions in positive and negative. Training data sets to train a model and evaluating data sets for test the model according to previous exercise. ML field has several issues which remain to be solved in this area, such as text mining, sentiment classification, image processing or classification problem. SA grouping there are three type, namely aspect-based sentiments, document-based

sentiments, and sentence-based sentiments. Initially, document based SA is used to figure out whether it is positive or pessimistic about the entire text. For example, a review relating to a mobile product which contains twenty lines, and we do not analyses a single sentence or words throughout the evaluation. Second, SA based on sentence to figure out opinion in a whole analysis, one by one paragraph. And sixth, aspect-based SA to search for product features analysis [3]. The general classifiers are discovered that the classification component used as train data set in the document and the classification accuracy can be calculated using the document dataset. There are few ML available that can classify the model exactly to help better in the classification of sentiments. For this study, the dataset has considered about smart phone product with 4000 customer's feedback has been collected from Amazon.com. It was screened to eliminate noisy data and was pre-processed using supervised learning to test the analysis of sentiments. This research work focuses classification methods for ML such as Naïve Bayes (NB) Logistic Regression (LR), Support Vector Machine (SVM) and Neural Network (NN) have been tested to determine the best classifier for this determination.

The organizations of this paper are as follows. Section 2 describes the related survey regarding technique based methodological contributions from existing work, Section 3 describes the proposed methodology based on SA using various ML, Section 4, discusses evaluation based on confusion matrix, Section 5 concludes the evaluation work.

2. LITERATURE REVIEW

There is numerous research in the area of SA which has been done. However, there was not much research done in area of text based classification which can classify the sentence associated to positive, negative and neutral. Chawla et.al [4] provides for dividing positive, negative and neutral using SA based on various mobile phone reviews. This can be posted by various number of users being obtained by studying the various post considering about areas in which it can categorize the smart phones. Abudalfa et.al [5] discusses labeled micro-blogs that has been used in higher number in supervised learning than the semi supervised learning techniques. Dong et al. [6] integrated to exploit the power of deep learning by use of target information with recursive NN. Quan and Ren [7] discusses to provide high fine grained SA using similarity based method. Vo and Zhang [8] developed word2vec features which are appropriate for classification based on target-dependent sentiments whereas the features are used by reported efficiency when applying SVM classification with sentiment. Tang et al. [9] suggested target-dependent on Long Short Term Memory (LSTM) framework where they have been registered an increased accuracy by 0.4 percent and macro-F1 performance did not improve. Singh et al. [10] suggested a technique and mainly

focused on the OneR ML algorithm, as well as the BF tree, NB, and J48 algorithms used in the training and testing of imdb and Amazon reviews. OneR obtained the highest proportion of experimental findings from F-score 97.0 %t, BF tree received F-score 72.1%, NB 81.2 %, and J48 91.7 %. Elmurngi et al. [11] described using three distinct Amazon review datasets, including reviews of apparel, shoes and jewellery, baby reviews, and pet reviews which are characterized the feeling using four supervised ML algorithms. The results indicate the accuracy to detect unfair feedback using 4 sentiment classification algorithms. Jain and Dandannavar [12] explains the sentimental analysis on twitter data using decision tree and multinomial NB algorithms. Results concluded that decision tree algorithm performs better using parameter accuracy, recall, precision, and F1-Score.Guadalupe et al, [13] implemented three-kernel SVM algorithm namely linear, polynomial and radial to determine a positive, negative and neutral classification of comments. As a measure of assessment we measured responsiveness, accuracy, and predictive values. Appel et al. [14] introduced a hybrid method to the SA which is the combination of approaches based on the lexicon and the learning. Such strategies will be allocated on a polarity of opinion and will define the specific emotion. This research achieved better performance on a film review dataset of about 75% precision. Kumar et al. [15] analyze the various datasets used to classify text and sentiments and described for classifying sentiment sand texts using SVM classification technique based on smartphone product review. The obtained result on the basis of mobile feedback which approves higher accuracy. Shahid et al., [16] conducted by classifying the sentiment for opinion mining purposes of activated smartphone feedback. The result obtained by using evaluation parameters namely, recall, precision, accuracy and F-measure with the comparative study of 8 separate classifiers. The Random Forest Classifiers gives highly reliable forecasts but improved precision is also given by LSTM and CNN. Priyavrat and Singh [17] discusses on the basis of various output parameters for SA using the comparative study of four controlled ML approaches. In this comparative analysis, SVM is evaluated with greater success than other three supervised ML. Wahyudi and Kristyanti[18] address work using the selection features of the merger process, including Particle Swarm Optimization to improve the precision of SVM classifications. This research produces text classifications in the review of smartphone products as positive or negative. Kalaivani and Thenmozhi [19] provide DL models aid in resolving different issues of SA. The classification methods' performance is contrasted with respect of accuracy. Rana and et.al [20] used SVM and Naïve Bayes techniques to conduct a Proportional Analysis of Emotional Location. They used customer feedback of video datasets. They made a difference between people's views. The outcome was that Naïve Bayes, utilizing linear SVM and

synthetic word method which has been provided the highest precision. Kalavathi and Suresh[21] this paper mainly focuses on addressing ambiguity in applications that include measuring equipment control, data uncertainty, incompleteness, delay or loss of data updates and privacy tolerance issues, and also reduces overhead counts. This section has discussed various analysis of ML technique in order to identify the polarity of the words and also study about SA concept to the tokenized sentence and words

3. RESEARCH METHODOLOGY

This research work customer feedback about smart phone product has been collected from Amazon.com. The customer also wants real comments to the service company before charging the money because at the moment it is not possible to read all the suggestions provided by the customers from the website. Thus, any product analysis or feature of the company provides the new information. Subsequently any significant input provided by the customer is likely to be missed for resolving the aforementioned challenge; the frequency of the review rating has to be established. The extraction of all tokenized words is calculated by word count using SA based on customer feedback that has been attained by deep learning. The best way to read feedback by SA is to figure out the rating of responses along with the word count. Hence the rating may be accepted by the customer depending on the feedback. After obtaining the SA, the user has made a faster and reduced effort to interpret the input as the decision. For this work, the accompanying rating scale from 1 to 5 based on the frequency of sentimental words. Very positive, positive, neutral, negative and very negative rating are the following groups. Table.1 and figure.1 shows the overall rating with corresponding Prod ID and rating scale.

Table 1: Classifying the Reviews based on SA

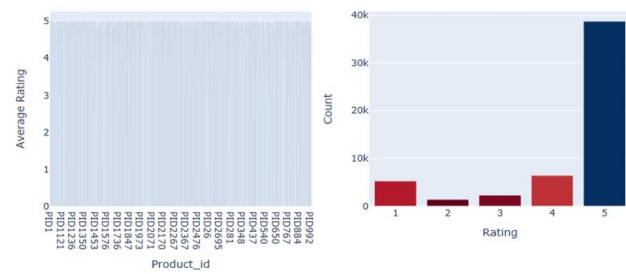| Prod_ID	Average Rating	User Count
PID992	2	30k
PID884	5	80k
PID767	3	35k
PID650	3	30k
PID540	4	19k
PID281	4	19k
PID2071	4	19k
PID1847	4	80k
PID1453	3	90k
PID1350	3	90k
PID1121	4	19k
PID1236	2	4k

Figure 1: Frequency Review Rating

3.1 Sentiment Analysis using various ML algorithms

One of the outstanding techniques for identifying the review rating is SA in which the customer has ability to analyze the product by viewing the rating but it doesn't provide very accurate prediction of the product but the outcome of SA will be able to make decision as quick as possible without reading the entire feedback review. According to the supervised ML technique, there are two different random sampled datasets are provided as train dataset and test dataset with the segregation of 80% and 20% whereas the classification factors of SA present over train dataset can be learned through automated classifier model by its ROC curve. Similarly, validating the model of SA test dataset using confusion matrix parameters namely Sensitivity, Specificity and accuracy of the classification models. Hence, the major four ML algorithms used here are NB, LR, NN and SVM. S.Muthukumaran and Dr. P.Suresh [22]

A. Logistic Regression (LR)

LR is a method to type f learning functions: A ->B, or P(b) where B is discreetly defined and A = {A1 ... Bn} is any vector that includes continuous or discrete variables. Consider the case of A being a Boolean function. LR takes a parametric form for the P (B) distribution and then calculates the parameters openly from the training data. When Y is Boolean in equation 1, the parametric method presumed by LR,

$$P(B - 1|A) = \frac{1}{1 + \exp\left(W_0 + \sum_{i=1}^{n} W_i A_i\right)} \tag{1}$$

$$P(B - 0|B) = \frac{\exp\left(W_0 + \sum_{i=1}^{n} W_i A_i\right)}{1 + \exp\left(W_0 + \sum_{i=1}^{n} W_i A_i\right)} \tag{2}$$

Remember the equation 2 proceeds immediately from equation 1 because the sum of these dual possibilities must be equal to 1. One highly suitable property of this type for classification of P (B) is that it donates to a simple linear expression. The function y_k which maximizes $P(B = y_k)$ is allocated to identify any given A. If the following equation applies (Equation 3), give the label B = 0 otherwise.

$$1 < \frac{P(B - 0|A)}{P(B - 1|A)} \tag{3}$$

Substituting from equations (1) and (2), this becomes (Eq.4),

$$1 < \exp\left(W_0 + \sum_{i=1}^{n} W_i A_i\right) \quad (4)$$

and we have a linear classification rule that assigns label B = 0 if A follows the formula 5.

$$0 < \left(W_0 + \sum_{i=1}^{n} W_i A_i\right) \quad (5)$$

and assigns B = 1 otherwise. Therefore, in this study, the suggested proposed system customers review is evaluated in Table.2 with their prediction of sentiment performance with LR get analyzed.

Table 2: Prediction using LR based on sentiment

Prod_ID	Prd_Name	Brand Name	Rating	Reviews	Word Count	Review Votes	Sentiment	LR
PID2429	Samsung Galaxy S6 SMG920FD cellphone,32GB	Samsung	5	It's great.	{'great': 1, 'it': 1,}	0.0	True	0.99
PID118	Apple iPhone 5C 16GB	Apple	5	Excellent	{'Excellent': 1}	1.0	True	1.00

The model performance has to conclude using single metrics namely, Region of Curve (ROC) curve whereas the LR model has occurred around 0.97 has performed in figure.2
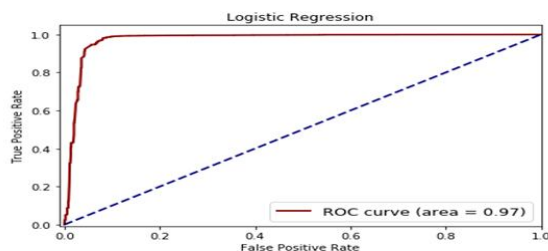

Figure 2: ROC curve area for LR

B. Support Vector Machine (SVM)

SVM is a commonly employed supervised classifier with a hypothetical basis that works more reliably in many implementations than most of other algorithms. Several researchers have claimed that SVM is possibly the maximum precise method of text classification SVM which is the linear model of training that seeks an ideal hyperplane for separating two groups. In a supervised method, SVM attempts to optimize the gap from either category to the closest testing point in order to attain improved quality of generalization on test data. In equation.6, it is possible to express a generalization-based linear SVM classifier(x), classifying x∈X diagrams denoted by a function vector (x1, x2) into classes C can be represented as {positive, negative}.

$$\gamma(x) = w.x + b \quad (6)$$

Where 'w' represents the perpendicular vector to the limit, and where 'b' represents the borderline offset which determines from the source. The versions of x would be marked as positive based on this formalization if $\gamma(x) <= 0$ or negative. Thus, the suggested proposed system customer's reviews are evaluated in table.3 with their prediction of sentiment performance with SVM get analyzed.

Table 3: Prediction using SVM based on sentiment

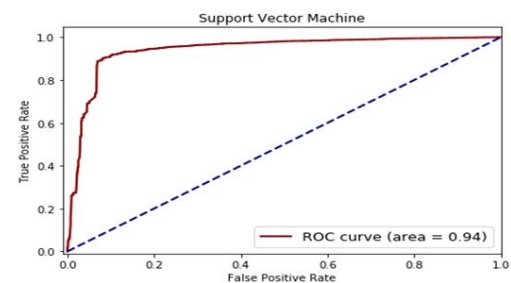| Prod_ID	Prd_Name	Brand Name	Rating	Reviews	Word Count	Review Votes	Sentiment	SVM
PID2429	Samsung Galaxy S6 SMG920FD cellphone, 32GB	Samsung	5	It's great.	{'great': 1, 'it': 1,}	0.0	True	1.07
PID118	Apple iPhone 5C 16GB	Apple	5	Excellent	{'Excellent': 1}	1.0	True	1.04

Figure 3: ROC curve area for SVM

The model performance has to conclude using ROC whereas the SVM model has occurred around 0.94 which is smaller than LR in figure.3.

C. Naive Bayes (NB)

NB classifier is a basic probabilistic system based on the rule of the Bayes together with a firm presumption of freedom. The concept of NB includes simplifying the presumption of contingent freedom. That is, the terms are conditionally distinct from each other provided a category positive or negative. This hypothesis does not greatly affect the reliability of text classification, but it creates the problem very importantly to fast classification algorithms. The maximum probability of a term going to a specific class is provided by the expression 7 in sentiment classification.

$$P(X_i | c) = \frac{Count\ of\ X_{i\ in\ documents\ of\ class\ C}}{Total\ no\ of\ words\ in\ document\ of\ class\ C} \quad (7)$$

During the training phase, frequency sums of terms are contained in hash tables. The probability of a specific text fitting to a group ci is given by formula 8, according to the Bayes law.

$$P(c_i | d) = \frac{P(d | c_i) * P(c_i)}{P(d)} \quad (8)$$

If we use the simplifying of conditional independence that the terms are conditionally separate from each other provided a category (positive or negative). The model is called "naive" because of this simplifying hypothesis formula 9.

$$P(c_i \mid d) = \frac{\pi P(X_i \mid c_j) * P(c_j)}{P(d)} \qquad (9)$$

Here the xi's are the document's single letters. With the highest posterior likelihood, the classifier produces the class. If the classifier finds a words that was not seen in the training set, the possibility of both categories would be null and there nothing to evaluate. The laplacian smoothing formula 10 should solve this problem,

$$P(X_i \mid c_j) = \frac{Count(X_i) + K}{(K+1) * (No\,of\,words\,in\,class\,c\,j)} \qquad (10)$$

K is normally picked as 1. Therefore, the probability of the new word being in any category is comparable. Thus, the suggested proposed system customers reviews are evaluated in Table 4 with their prediction of sentiment performance with NB get analyzed.

Table 4: Prediction using NB based on sentiment

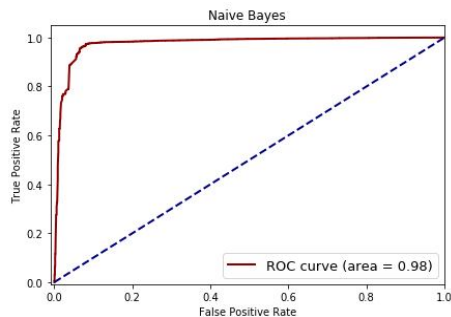| Prod_ID	Prd_Name	Brand Name	Rating	Reviews	Word Count	Review Votes	Sentiment	NB
PID 2429	Samsung Galaxy S6 SMG920FD cellphone,32 GB	Samsung	5	It's great.	{'great': 1, 'it': 1,}	0.0	True	0.99
PID118	Apple iPhone 5C 16GB	Apple	5	Excellent	{'Excellent': 1}	1.0	True	0.98

Figure 4: ROC curve area for NB

The model performance has to conclude using ROC whereas the NB model has occurred around 0.98 which is higher than LR and SVM in figure.4.

D. Neural Networks (NN)

NN is to extract features from linear input information combinations and then model the output of these features as a nonlinear purpose. The architecture of the best extensively utilized NN comprises of 3 layers namely source, an unknown, and a node output layer. A neuron is a scientific template that in two steps yields an output value. To obtain its output, the neuron first determines a biased total of its inputs and then relates a stimulation function to that sum. The activation function is usually a function of nonlinear which means that a nonlinear feature that is derived from the input data can be calculated by the entire network. The output of the network for an input x is expressed in equation (11):

$$S(x) = Wotanh(Whx + bh) + bo \qquad (11)$$

Where Wo is a n x g_n matrix, W_h is a $g_n \times |V|$ matrix, where 'b$_o$' represents n-dimensional vector, and 'b$_h$' represents g_c-dimensional vector. For a given example x, S(x) is the score vector for each class. The sentiment dataset is the input given to this model, while the performance has shown in table.5 based on the chosen sentiment terms and customer rating with NN.

Table 5: Prediction using NN based on sentiment

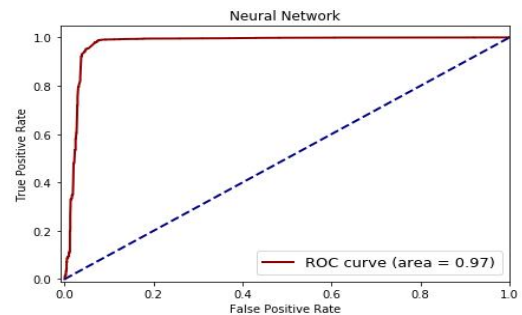| Prod_ID	Prd_Name	Brand Name	Rating	Reviews	Word Count	Review Votes	Sentiment	NN
PID 2429	Samsung Galaxy S6 SMG920FD cellphone,32 GB	Samsung	5	It's great.	{'great': 1, 'it': 1,}	0.0	True	0.9
PID118	Apple iPhone 5C 16GB	Apple	5	Excellent	{'Excellent': 1}	1.0	True	1.0

Figure 5: ROC curve area for NN

The model performance has to conclude using ROC whereas the NN model has occurred around 0.97 which is higher than SVM and lesser than NB in figure.5. In order to predict the feedback review text, the python logic functions are used.

4. EVALUATION OF SENTIMENT ANALYSIS USING VARIOUS ML ALGORITHMS

ML as an application to sentiment analysis has recently attracted attention. Two types of datasets, namely are required for supervised ML technique of dataset train and dataset test. The general classifier was discovered that the classification element in the research work is used as the collection of train data and the classification accuracy which can be calculated using the test data. This research work shows the success of the sentiment analysis using four ML are LR, NB, SVM and NN are made for comparison with sentiment analysis to discover the better performance of ML with sentiment analysis. In this research, work involves the test dataset as feedback from the customer whereas the prediction has been evaluated via confusion matrix parameters: sensitivity, specificity and accuracy [23]. Figure.6 illustrates the confusion matrix for the respective classifier namely LR, SVM, NN and NB given in table 6.
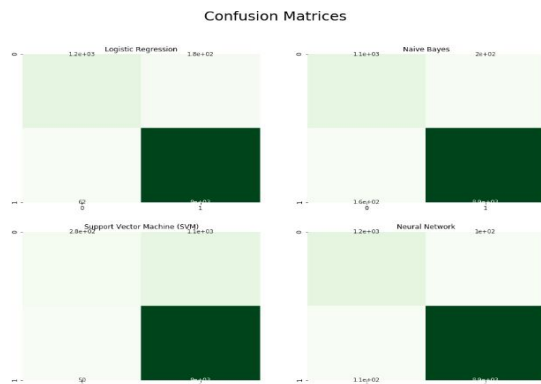
Confusion Matrices

Figure 6: Performance of classification model

Table 6: Performance evaluation of various ML algorithm

ML algorithm	Confusion Matrix value			
	True Positive (TP)	True Negative (TN)	False Positive (FP)	False Negative (FN)
LR	8000	1200	180	62
SVM	8000	280	1100	50
NB	8900	1100	200	160
NN	8900	1200	100	110

In this analysis the evaluating parameters are measured as a capability to determine the correct student review based on the value of sensitivity, specificity and accuracy are illustrated in table 7 and figure 7.

Table 7: Confusion matrix parameters for various ML algorithm

S.NO	ML algorithm	Sensitivity	Specificity	Accuracy
1.	LR	0.99	0.870	0.974
2.	NB	0.98	0.846	0.965
3.	SVM	0.99	0.203	0.878
4.	NN	0.99	0.923	0.980

According to sensitivity, NB has lesser accurate determination as 0.98 but other classification algorithm performs greater than NB. Consequently, in this research the LR, SVM and NN exhibits stronger responsiveness as per student analysis and rating. According to specificity, SVM has the lesser value as 0.203 compared to LR, NB and NN. According to accuracy, NN has the best performance value to determine SA as 0.980 which is the highest value of SA when compared to other classification algorithm [24]. Therefore, in this research NN performs better accuracy on customer review and rating. The customer feedback is good can be determined using SA gets evaluated.
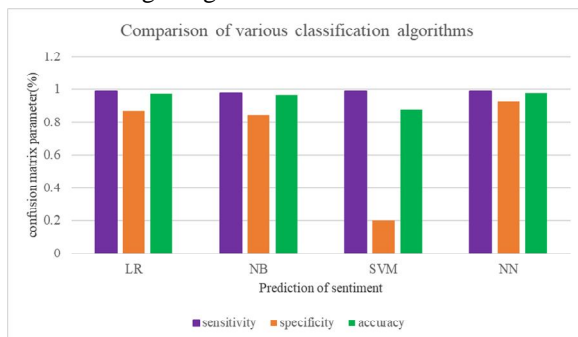

Figure 7: Confusion matrix parameters for various classification algorithm

5. CONCLUSION

The main aim of this research is to use ML algorithms to identify the sentences according to their sentiments. In this paper, to perform sentimental analysis, first the feedback has been collected, given to the trained model and the sentiments are classified. Thus, ML algorithms are used to test the SA datasets to determine the best results. Similarly, ROC predictions of the train data set with these four ML using the analysis. In order to predict the feedback review text, the python logic functions are used. Hence, the ROC curve area for NB value as 0.98 which is greater than other classification algorithm. The test data gets validated with the performance of confusion matrix parameters have illustrated that NN algorithm performed better results in accuracy. However, the NN has performed better accuracy value as 98% which is higher than LR, NB and SVM algorithm. Therefore, the proposed SA with NLP is performed better with NN to identify the better implicit of product. This method of research helps to define the indirect product to boost the sales by offering deals for the respective implicit products.

REFERENCES

1. T. Kumar and M. Trakru, "**The Colossal Impact Of Artificial Intelligence In E - Commerce : Statistics And Facts,**" no, pp. 570–572, May 2019.
2. S. Vanaja, "**Aspect-Level Sentiment Analysis on E-Commerce Data,**" 2018 Int. Conf. Inven. Res. Comput. Appl., no. Icirca, pp. 1275–1279, 2018.
3. Premnarayan Arya and Amit Bhagat, "**Performance Evaluation of Several Machine Learning Classification Algorithms with Combined Feature Selection Methods for Sentiment Analysis**", International Journal of Innovative Technology and Exploring Engineering, ISSN: 2278-3075, Volume-8 Issue-6, April 2019.
4. 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.
5. Shadi I. Abudalfa and Moataz A. Ahmed, "**Semi-Supervised Target-Dependent Sentiment Classification for Micro-Blogs**", Journal of Computer Science & Technology, Volume 19, Number 1, April 2019.
6. L. Dong, F. Wei, C. Tan, D. Tang, M. Zhou, and K. Xu, "**Adaptive Recursive Neural Network for Target-dependent Twitter Sentiment Classification**," Proc. of the 52nd Annual Meeting of the Association for Computational Linguistics (Short Papers), pages 49–54, Baltimore, Maryland, USA, June 23-25, 2014.

7. C. Quan, and F. Ren, "**Target Based Review Classification for Fine-Grained Sentiment Analysis**," International Journal of Innovative Computing, Information and Control, vol. 10, no. 1, February 2014.

8. D. Vo and Y. Zhang, "**Target-Dependent Twitter Sentiment Classification with Rich Automatic Features**," the Twenty-Fourth International Joint Conference on Artificial Intelligence (IJCAI), 2015.

9. D. Tang, B. Qin, X. Feng, and T. Liu, "**Effective LSTMs for Target-Dependent Sentiment Classification**," COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers, pp. 3298–3307, Osaka, Japan, December 11-17 2016.

10. Jaspreet Singh, Gurvinder Singh and Rajinder Singh, "**Optimization of sentiment analysis using machine learning classifiers**", Human-centric Computing and Information Sciences, Springer Open, Dec 2017.

11. Elshrif Ibrahim Elmurngi and Abdelouahed Gherbi, "**Unfair reviews detection on Amazon reviews using Sentiment Analysis with supervised learning techniques**", Journal of Computer Science, pp. 714-726, 2018.

12. A. P Jain, P.Dandannavar, "**Application of Machine Learning Techniques to Sentiment Analysis**", Proceeding of 2018 IEEE International Conference on Current Trends toward Converging Technologies, pp. 628-632, 2018.

13. Guadalupe Gutiérrez Esparza , Alejandro de-Luna, Alberto Ochoa Zezzatti, Alberto Hernandez , Julio Ponce , Marco Álvarez , Edgar Cossio and Jose de Jesus Nava, "**A Sentiment Analysis Model to Analyze Students Reviews of Teacher Performance Using Support Vector Machines**", Advances in Intelligent Systems and Computing, Springer International Publishing AG 2018, DOI 10.1007/978-3-319-62410-5_19.

14. 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.

15. UpmaKumari, ArvindK.Sharme and Dinesh Soni, "**Sentiment Analysis of smart phone product review using classification technique**", IEEE, 2017.

16. MominaShaheen, Shahid M. Awan, NisarHussain and Zaheer A. Gondal, "**Sentiment Analysis on Mobile Phone Reviews Using Supervised Learning Techniques**", I.J. Modern Education and Computer Science, Vol-7, pg 32-43, 2019.

17. A. J. Singh, "**Sentiment Analysis : A Comparative Study of Supervised Machine Learning Algorithms Using Rapid miner**", vol. 5, no. Xi, pp. 80–89, 2017.

18. MochamadWahyudi and Dinar AjengKristiyanti, "**Sentiment analysis of smart phone product review using SVM algorithm based particle swarm optimization**", Journal of Theoretical and applied information technology, 2016.

19. KalaivaniA andThenmozhi D, "**Sentimental Analysis using Deep Learning Techniques**", International Journal of Recent Technology and Engineering, Volume-7, Issue-6S5, April 2019.

20. S. Rana and A. Singh, "**Comparative analysis of sentiment orientation using SVM and Naive Bayes techniques**," 2nd International Conference on Next Generation Computing Technologies (NGCT), pp. 106–111, 2016.

21. M.Kalavathi and Dr. P.Suresh, "**An analysis of space query classifier indexing for mining uncertain data**", International Journal of Computational Intelligence and Informatics, ISSN:2349-6363, vol.6, no-1, june 2016, pp-66-72.

22. S.Muthukumaran and Dr. P.Suresh, "**A Unified Framework of Sentimental Analysis for Online Product Reviews Using Enhanced Ant Colony Optimization Algorithm**" International Journal of Pure and Applied Mathematics Volume: 114, Issue No: 14,PP: 489-486,June 2018 ISSN: 1314-3395,SCOPUS Indexed

23. M Mallikarjuna, R Prabhakara Rao, "**Application of Data Mining Techniques to Classify World Stock Markets**," International Journal of Emerging Trends in Engineering Research, Volume 8, No. 1, pp. 46-53, January 2020.

24. Haryo Pramanto, Candra Setiawan, Abba Suganda Girsang, Roni Jhonson Simamora, Melva Hermayanty Saragih "**Student Absence Attendance Fine using Data warehouse System**," International Journal of Emerging Trends in Engineering Research, Volume 8. No. 4, April 2020