Journal of Information Systems and Telecommunication
Vol.10, No.1, January-March 2022, 1-10
http://jist.acecr.org
ISSN 2322-1437 / EISSN:2345-2773

A Hybrid Machine Learning Approach for Sentiment Analysis of Beauty Products Reviews

Kanika Jindal[1*], Rajni Aron[2]

[1].Lovely Professional University, Punjab, India
[2].SVKM's Narsee Monjee Institute of Management Studies, Mumbai, India

Received: 22 Sep 2020 / Revised: 25 Aug 2021/ Accepted: 25 Sep 2021
DOI: https://doi.org/10.52547/jist.15586.10.37.1

Abstract

Nowadays, social media platforms have become a mirror that imitates opinions and feelings about any specific product or event. These product reviews are capable of enhancing communication among entrepreneurs and their customers. These reviews need to be extracted and analyzed to predict the sentiment polarity, i.e., whether the review is positive or negative. This paper aims to predict the human sentiments expressed for beauty product reviews extracted from Amazon and improve the classification accuracy. The three phases instigated in our work are data pre-processing, feature extraction using the Bag-of-Words (BoW) method, and sentiment classification using Machine Learning (ML) techniques. A Global Optimization-based Neural Network (GONN) is proposed for the sentimental classification. Then an empirical study is conducted to analyze the performance of the proposed GONN and compare it with the other machine learning algorithms, such as Random Forest (RF), Naive Bayes (NB), and Support Vector Machine (SVM). We dig further to cross-validate these techniques by ten folds to evaluate the most accurate classifier. These models have also been investigated on the Precision-Recall (PR) curve to assess and test the best technique. Experimental results demonstrate that the proposed method is the most appropriate method to predict the classification accuracy for our defined dataset. Specifically, we exhibit that our work is adept at training the textual sentiment classifiers better, thereby enhancing the accuracy of sentiment prediction.

Keywords: Sentiment Analysis; Machine Learning; Beauty Products; Feature Extraction; Social Media.

1- Introduction

Sentiment Analysis (SA) is a systematic study of the collection and classification of product reviews on various e-commerce platforms [1]. As the online business has become more popular these days, both sellers and customers are interested in asking and providing feedback on e-commerce platforms simultaneously. These opinions and reviews are a kind of verbal communication that includes personalized suggestions and product ratings. These reviews are a guiding tool for companies to improve their product quality and services. They are very beneficial for consumers to help in making decisions regarding the specific product [2]. Presently, the communication conduct of this digital era's customers has been customized towards the beauty industry that developed as a highly competitive business market [3]. Various social media and e-commerce platforms provide reviews and ratings of different types and brands of cosmetics products to consumers.

Amazon is one of the popular e-commerce platforms that is used to make online purchases. The customers can also provide and review feedbacks regarding any purchase or product available on the website [4]. Although it is very beneficial for consumers and vendors, the increasing number of reviews about a product confuses customers to make the right decision [5]. Therefore, a need arises to analyze these online reviews by classifying them as positive or negative, improving the decision-making process [6]. The customers also tend to express their views in their natural language, so extracting and classifying these language-based reviews using sentiment analysis is necessary. Sentiment analysis is a branch of Natural Language Processing (NLP) that can address the above-discussed problem [7]. Machine Learning techniques are used in sentiment analysis tasks to classify these reviews as positive, negative, and neutral [8]. These trained classifiers are processed to attain reasonable accuracy and require ascertaining the textual data pertinent to the current potentials [9].
This paper presented an empirical study of sentiment classification of textual data using the Bag-of-Words

✉ **Kanika Jindal**
kanikajindal11@gmail.com

technique and implemented three machine learning models. In our study, the unstructured data of beauty product reviews are extracted from Amazon. This work involves three steps, i.e., data pre-processing, feature extraction, and sentiment classification. For this, the unstructured reviews are pre-processed in the first step, and the features are extracted using the Bag-of-Words (BoW) model in the next step. A Global Optimization-based Neural Network (GONN) is proposed for the sentimental classification. Then an empirical study is conducted to analyze the performance of the proposed method with other machine learning classification methods, i.e., Naive Bayes, Random Forest, and Support Vector Machine, and K fold cross-validation is performed to evaluate the accuracy of the system. The other parameters such as precision, recall, and F1 score are also evaluated for all the models. It is concluded that the proposed GONN method outperforms all the other classifiers for the classification of the Amazon beauty products dataset and achieves the best accuracy.

The details of this work will be discussed in the following sections. Section 2 describes a review of related work. Section 3 elaborates on the proposed framework and methodology. Section 4 demonstrates the results and performance evaluation from the experimental work. Finally, section 5 summarizes the conclusion and future works.

2- Related Work

The term 'sentiment analysis' has attained extensive growth and attention in recent years [10]. The primary purpose of this technique is to understand the human emotions expresses in the form of sentiments on social media. It plays a significant role in various organizations concerning education, health, the stock market, and numerous products and services. The research work done in this direction is discussed in this section.

The work [11] is a sentiment analysis approach applied to Twitter data collected from disaster responses. The primary purpose is to understand the needs of the affected people so that rescue responders can help better. For this, the sentiments for the humanitarian reliefs obtained by affected people during and after the disaster are analyzed using machine learning methods. The paper [12] analyses public opinions regarding the demonetization policy implemented by the Indian government on November 8, 2016. The data is collected from Twitter for the two weeks after the policy declaration, and state-based analysis is performed on it. It concluded that almost all the states supported this policy after tackling some minor hindrances for some time. The article [13] is about the application development for cosmetics product reviews gathered from a popular website. It scrutinizes both positive and negative reviews about numerous products using Parts of Speech (PoS) tagger and Naive Bayes classifier. The author endorses using both types of comments in an equal ratio to achieve higher accuracy and efficiency.

The authors of this work [14] proposed a framework in which the support vector machine method and three feature selection methods are used. The dataset comprises 200 reviews extracted from www.amazon.com. All three techniques, i.e., Genetic Algorithm (GA), Particle Swarm Optimization (PSO), and Principal Component Analysis (PCA), are compared, and it is concluded that the PSO technique resulted in the best accuracy with SVM. According to the authors [15], the application to automatically analyze the sentiments regarding skincare products can be an effective tool these days. It can be beneficial for both consumers and entrepreneurs. This work has been implemented on a web application to analyze skincare-based tweets by applying data pre-processing and classification methods. The performance results were evaluated to be more than 80%.

This paper [16] is based on stock market forecasting by amalgamating the financial market data with the sentiment features. The data was collected from two financial websites, and machine learning methods SVM is used in this work. The day-of-week effect has been contemplated in this study to improve prediction accuracy. Thus, this approach can help to make better investment decisions in the financial market. This work [17] analyses the cosmetics product reviews written in the Thai language by using the Naive Bayes algorithm. The authors have used various techniques to evaluate the significant phrases, such as cosine similarity, page rank, and Hopfield network algorithm. The paper concludes that the results were not very accurate due to highly unstructured social media data and inadequate management of synonyms. In this paper [18], a framework analyses the laptop reviews based on the product's design, performance, and features. The work consists of three phases, i.e., subjective extraction, calculating the frequency of words, and sentiment classification. It can help people to make effective decisions before buying laptops. The future suggestion is to incorporate the system for other domains. This work [4] has evaluated the textual data by considering the aspect level detection as well as bipolar words for the analysis of sentiments. Amazon data has been pre-processed for extracting information and positive or negative sentiments are generated by utilizing the proposed approach. The future work suggests including other challenges like sarcasm and negation related to sentiment analysis work.

3- Proposed Framework

Sentimental analysis of beauty product reviews in social media is the motivating research in this paper. The influence of social media reviews on beauty products has a positive impact on choosing the right product [3]. But to lead the marketplace, the brands may influence the marketing in the review comments. So, finding the sentiment of the review is the most essential to show the effective review to the consumers. Hence, a hybrid machine learning approach is proposed to effectively predict the sentiment of the social media review on beauty products.

The proposed framework has been segmented into three phases, i.e., data collection and pre-processing, feature extraction using the Bag-of-Words (BoW) method, and sentiment classification using Machine Learning (ML) methods. An empirical analysis of all these techniques has been performed to find the performance evaluation metrics, i.e., accuracy, precision, recall, and F1 score. A brief description of the proposed methodology has been conferred in this section. The process flow of the proposed technique is given in Fig. 1.

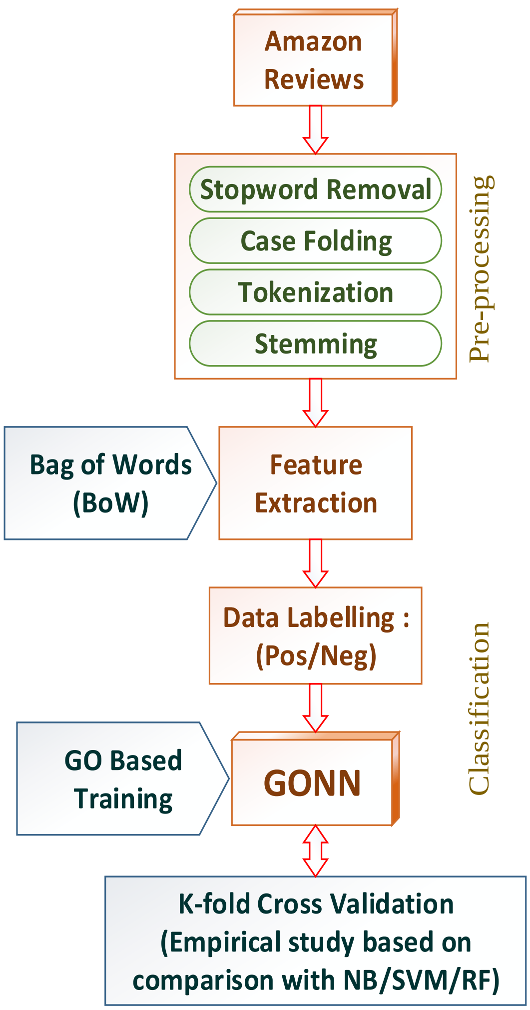

Fig.1 The framework of beauty products review analysis

3-1- Data Collection

The first module involves the collection and pre-processing of data. The dataset used in our work has been accumulated from the gigantic e-commerce platform Amazon.com [19]. It contains an abundant number of reviews based on each product category. This dataset is for various beauty products having 5269 reviews and JavaScript Object Notation (JSON) formats. The various features of each review of the dataset are elucidated in Table 1. An example of the unprocessed dataset is described below.

{"overall": 5.0, "verified": true, "reviewTime": "01 31, 2018","reviewerID": A2IGYO5UYS44RW", "asin": "B00006L9LC", "style": {"Size:": " 281"}, "reviewerName": "Dawna Kern", "reviewText": "I love how soft this makes my skin and the scent is amazing. When my local stored are out I can always get it at Amazon", "summary": "BETTER THAN RAINBATH", "unixReviewTime": 1517356800}

Table 1: Features of the reviews in the dataset

Fields	Description
reviewerID	ID of the reviewer
asin	ID of the product
reviewerName	name of the reviewer
vote	helpful votes of the review
style	a dictionary of the product metadata
reviewText	text of the review
overall	rating of the product
summary	summary of the review
unixReviewTime	time of the review (unix time)
reviewTime	time of the review (raw)

The text and summary of the review and overall rating of the product have been considered for our work from these features. The overall rating contains a rating given by beauty product reviewers. These ratings are expressed in the form of 1-5 stars, with 1 being a bad review and 5 being good reviews. Fig.2 shows the rating distribution of beauty product reviews (1-5 stars). In Fig. 2, the number of reviewers rated the particular review in which the rating 1 is given by 200 reviewers, whereas rating 5 is given by 3750 reviewers.

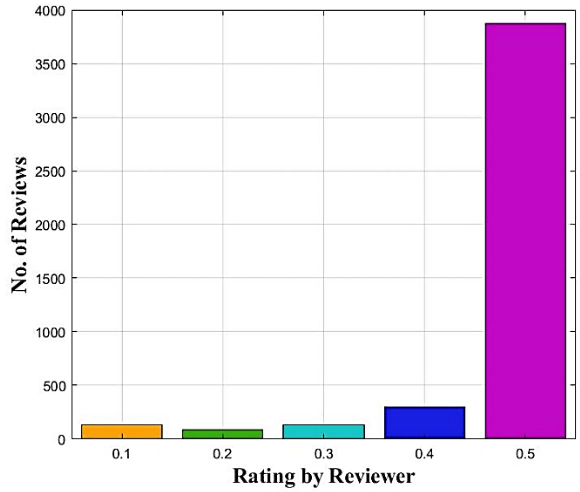

Fig.2 Rating distribution of the reviews in the Amazon dataset

3-2- Data Pre-processing

As the collected data is unstructured and noisy, the entire dataset is pre-processed to form a corpus [20]. The data needs to get clean as much as possible so that the machine learning model can easily understand it and predict whether the review is positive or negative. Therefore, in the data pre-processing process, the entire dataset passes through the following steps:

Stop Word Removal: All the non-relevant words are deleted in this step, like the, and, for, etc. These words do not help predict the polarity of the reviews.

Tokenization: All the relevant words are considered tokens, and all the punctuation marks and special symbols are omitted.

Case-folding: All the tokens are converted into lowercase to avoid repeating the same word in both uppercase and lowercase.

Stemming: Stemming means simplifying each word by its root that indicates enough about what that word means. All the conjugation of the verbs is removed in this step to reduce the redundancy and dimensionality of the sparse matrix.

3-3- Feature Extraction

The Bag-of-Words (BoW) model is used to extract features from the textual reviews collected and pre-processed in the previous phase. These extracted features can be effectively used in machine learning models in the next phase of our methodology. BoW model is used to represent text in the form of a vocabulary that contains the occurrence of the words in the whole document [21]. The frequency of occurrence of each word is assigned a unique number. The features are created by observing all the reviews discretely as an unordered corpus of words to be easily classified afterward. Finally, the textual reviews are fed into machine learning algorithms in the form of numerical vectors.

3-4- GONN -Based Sentiment Classification

The most crucial phase of the proposed work is to evaluate the sentiment prediction accuracy of the reviews expressed by the beauty product users. For this purpose, all the reviews are assigned by the Pos/Neg label to concoct a significant sentiment orientation. The labels are classified depending upon the ratings of the reviews specified by users. This labeled and classified dataset is divided into training (80%) and test (20%) data and implementing machine learning models. Machine Learning methods are best suited for the sentiment classification of these reviews because customers tend to express their suggestions and feedback in their natural language [22].

Hence the GONN is proposed for the effective prediction of the sentiment of the public review. In the proposed GONN, a global optimization technique with a swarm update rule is developed to train the neural network.

3-4-1 Mathematical Modeling of Feed-Forward Neural Network

The proposed neural network consists of three input neurons, one output neuron, and an 'M' hidden neuron. In this model, M is considered as 2. The three input represents three inputs such as word count, character count, and BoW feature. The output neuron represents the class label as positive or negative. The structure of the proposed neural network is given in Fig 3.

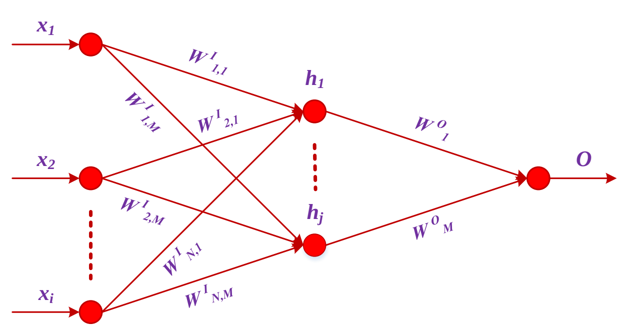

Fig.3 Structure of proposed neural network

Basis function at hidden layer: The basis function calculation is the first step in which the product of input with the weight of the respective link is calculated. The basis function for every node in the hidden layer is calculated as in Eq. (1).

$$b_j = \sum_{i=1}^{N} x_i W_{i,j}^{I} \qquad (1)$$

where 'b_j' is the basis function of jth hidden neuron; 'x_i' is the ith input value; '$W_{i,j}^{I}$' input weight between ith input neuron and jth hidden neuron, and 'N' is the total number of hidden neurons.

Tansig activation function at the hidden layer: The activation function is considered the output of the hidden layer and the input to the output layer. Many functions are available for the activation function calculation, such as tansig, sim, dtansig, logsig. Among them, tansig is the most used and better technique for activation calculation. The activation for every node in the hidden layer is calculated as in Eq. (2).

$$h_j = \left[\frac{2}{1 + exp\left(-2 \times \sum_{i=1}^{N} x_i W_{i,j}^{I}\right)}\right] - 1 \qquad (2)$$

where 'h_j' is the activation function of jth hidden neuron.

Neural network output calculation: The output or the obtained output of the proposed neural network is the basis value of the output layer. It is the product of activation value with the respective link in between the hidden and output layer. The output of the neural network is calculated as in Eq. (3).

$$O = \sum_{j=1}^{M} \left\{\left[\frac{2}{1 + exp\left(-2 \times \sum_{i=1}^{N} x_i W_{i,j}^{I}\right)}\right] - 1\right\} W_j^{O} \qquad (3)$$

where 'O' is the calculated output of neural network; 'W_j^{O}' is the weight between jth hidden neuron and output neuron. Eq. (3) provides the output of the nth training data. After obtaining all the data in the training set, the mean square error (MSE) is calculated as in Eq. (4).

$$Fit = MSE = \frac{1}{T} \sum_{n=1}^{T} (O_n - C_n)^2 \qquad (4)$$

Global optimization based neural network training:
In the conventional neural network, the backpropagation algorithm was widely used for training. Any training algorithm intends to find all the weight values of the network. In a conventional algorithm, a random weight between 0 and 1 would be assigned. Then after calculating the error, its weights are updated. This process is time-consuming and overloading the system. So, the finding of weights value is formulated as an optimization, and global optimization is proposed to find the optimal weight with less mean square error. Hence the accuracy of the system can be improved. The step-by-step procedure of the proposed optimization algorithm is given as follows:

Initialization: In this step, a random set of solutions is generated. The dimension of the solution is the sum of weights required for the proposed model. The range of solution or upper and lower bound of the solution is 0 and 1, respectively. The initial population is represented as in Fig. 4.

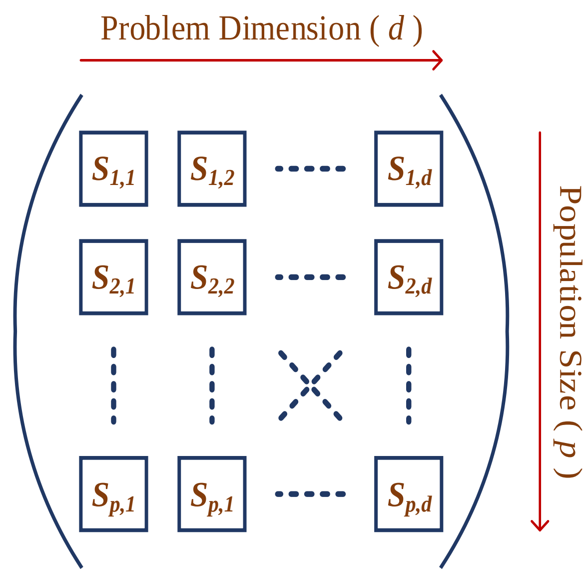

Fig.4 Initial population of proposed global optimization

In Fig. 3, the 'd x p' matrix is given, where 'd' is the dimension of the problem and 'p' is the population size. The population size is random can be any size. The large size of the population consumes execution time and converges at earlier iteration. But the dimension of the population is based on the number or required weights, which can calculate using Eq. (5).

$$d = (N \times M) + M \qquad (5)$$

Fitness Calculation: In this step, the fitness value for every solution set (singe row of the population) is calculated. The objective of this global optimization is to find the optimal weight for the neural network. So, the MSE has given in Eq. (4) is considered to evaluate fitness. The fitness evaluation is utilized to find the current best (C_{best}) and global (G_{best}) best values. The C_{best} is the best solution set among the population in the current iteration. The G_{best} is the over-best solution obtained among all the iterations as shown in Eq. (6).

$$G_{best}(iter+1) = \begin{cases} C_{best}(iter) & \begin{cases} iteration == 1 \\ G_{best}(iter) > C_{best}(iter) \end{cases} \\ G_{best}(iter) & otherwise \end{cases} \qquad (6)$$

Update Rule: After fitness evaluation, the solutions are updated based on a swarm rule. The swarm rule used here is referred to from [23].

$$pos(iter+1) = w \times pos(iter) + C_1 r_1 \left(C_{best} - sol(iter)\right) + C_2 r_2 \left(G_{best} - sol(iter)\right) \quad (7)$$

$$sol(iter+1) = sol(iter) + pos(iter+1) \quad (8)$$

In Eq. (7) and Eq. (8), 'pos' is the position value used, which is determined to find the new solution. The 'pos' of iteration 1 ($iter = 1$) is considered as 0, i.e.,$pos(1) = 0$. The parameters 'w, C_1, C_2, r_1, r_2' are probability values consider between 0 and 1.

Termination Criteria: The above steps are repeated for the maximum iteration. If the process meets maximum iteration, then the process is terminated by considering the G_{best} is the best solution or the optimal solution.

3-4-2 Empirical Study to Analyze the Effectiveness

In this empirical study, a comparison-based analysis is performed. Here some conventional machine learning algorithms are considered for comparison. The machine learning algorithms used in our study are Naive Bayes, Random Forest, and Support Vector Machine. The entire dataset is fed into these classifiers, and empirical analysis is performed. After that, K fold cross-validation (K=10) is performed to evaluate the best classifier based on the predicted accuracy attained by the machine learning methods. The overview of our framework for beauty products review analysis has been diagrammatically represented in Fig. 1.

4- Experimental Results and Analysis

The experimented data has been collected from Amazon for beauty product reviews posted by reviewers. Amazon reviewers can provide a product rating from 1 (lowest) to 5 (highest) stars. In our work, the rating stars have been utilized for labeling the reviews. The reviews having 3-star ratings are discarded in our study because this rating is considered neutral (neither positive nor negative) usually.
Therefore, the dataset contains a positive (Pos) label for all those reviews that are 1- or 2- stars and a negative (Neg) label for 4- or 5-stars reviews. Table 2. shows an overview of the product reviews after assessing positive and negative labels based on ratings. The reviews having less than five words are also removed. So, the final pre-processed and labeled dataset, containing 4200 reviews, is being executed by all the three machine learning classifiers, i.e., Naive Bayes, Random Forest, and Support Vector Machine.

Table2: The polarity of the reviews

Review	Sentiment
As advertised. Reasonably priced	Pos
Like the order and the feel when I put it on...	Pos
I bought this to smell nice after I shave it ...	Neg
HEY!! I am an Aqua Veleva Man and abs...	Pos
If you ever want to feel pampered to a sha...	Pos
If you know the secret of Diva you'll LOVE...	Pos
Got this shampoo as a solution for my wife's...	Pos
No change my scalp still itches like crazy...	Neg
Too expensive for such poor quality. Ther...	Neg
It dries my hair doesn't help to reduce dand...	Neg
Outstanding! Tob organic shampoo!	Pos
So watered down I didn't feel like it was a...	Neg
10 stars night here. This product helped me...	Pos
First hair care product I've decided to purc...	Pos
Mad dandruff worse and irritated rest of s...	Neg
Worst shampoo I've ever used. Was mostly...	Neg
Made my hair brittle and dull-looking didn...	Neg
I received the shampoo because I was suff...	Pos

4-1- Evaluation Metrics for Performance Measurement

The evaluation metrics are the fundamental values to evaluate the performance of text classification [24]. The sentiments classified in positive and negative polarity are identified by creating a confusion matrix of true positive, false positive, true negative, and false negative. Accuracy, precision, recall, and F1-score are the significant measures that can be gauged from the confusion matrix based on mathematical rules. The aspects of a confusion matrix are shown below in Table 3.

Table 3: Confusion Matrix

		Predicted Values	
		Positive	Negative
Actual Values	Positive	True Positive (TP)	False Negative (FN)
	Negative	False Positive (FP)	True Negative (TN)

The parameters emphasized in the above table are described as:

- True Positive (TP) is the positive value that is correctly identified as positive.
- False Positive (FP) is the negative value that is incorrectly identified as positive.
- False Negative (FN) is the positive value that is incorrectly identified as negative.
- True Negative (TN) is the negative value that is correctly identified as negative.

Precision, recall, F1-score, and accuracy metrics have been computed using the derived values of these parameters. The precision determines the total number of reviews that are accurately classified as positive. Recall determines the total number of reviews that are accurately classified as negative. F1-score measures the weighted harmonic mean of both precision and recall and merges them in a single metric. Accuracy is the simplest metric used to measure the frequency of correct predictions rendered by machine learning models. These metrics are represented by Eq. (6), (7), (8), (9) described below.

$$Precision = \frac{TP}{TP + FP} \qquad (6)$$

$$Recall = \frac{TP}{TP + FN} \qquad (7)$$

$$F1\ Score = \frac{2 * Precision * Recall}{Precision + Recall} \qquad (8)$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \qquad (9)$$

4-2- K-fold Cross-Validation

K-fold cross-validation is an evaluation procedure used to attain the maximal efficiency of machine learning models [25]. In this work, the cross-validation method divides the dataset into k subsets that are reiterating k times. In every split of data, that k^{th} fold denotes the test data, and the rest k-1 denotes the training data. The machine learning algorithms used in our experimental work are Support Vector Machine (SVM), Random Forest (RF), and Naive Bayes (NB). So, the cross-validation method has been performed on all three models for ten folds (k=10) to attain the best accurate classifier. Table 4 illustrates the accuracy determined by the cross-validation method employed on all three algorithms. The performance of other evaluation metrics, such as precision, recall, and F1-score for NB, SVM, and RF, can also be seen in the table given below.

Table 4: Experimental results of evaluation metrics for all the machine learning methods on the dataset

Methods	Accuracy 10-fold	Precision	Recall	F1-score
Naive Bayes	82.96%	97.92%	83.04%	89.87%
Support Vector Machine	95.87%	97.37%	97.86%	97.61%
Random Forest	96.65%	97.14%	98.49%	97.81%
GONN	97.51%	96.07%	98.98%	97.51%

Based on the above table, it is concluded that the proposed GONN, along with Support Vector Machine and Random Forest, both achieve accuracies above 90%. The table shows that Naive Bayes determines the total number of accurately classified reviews as positive with the best precision value of 97.92%. It is found that the Random Forest offers the best F1-score that is used to measure the efficiency of sentiment analysis towards the beauty products dataset. The recall values of GONN are highest as compared to the other methods. The bar graph plotted in Fig.5 depicts that GONN outperforms the other two methods in terms of accuracy, i.e., 97.51%, and the Naïve Bayes model has the lowest predictive accuracy, i.e., 82.96%. The performance of the F1 score has been shown in Fig.6. It concludes that the proposed GONN is the most accurate classifier compared to Random Forest, Support Vector Machine, and Naive Bayes.

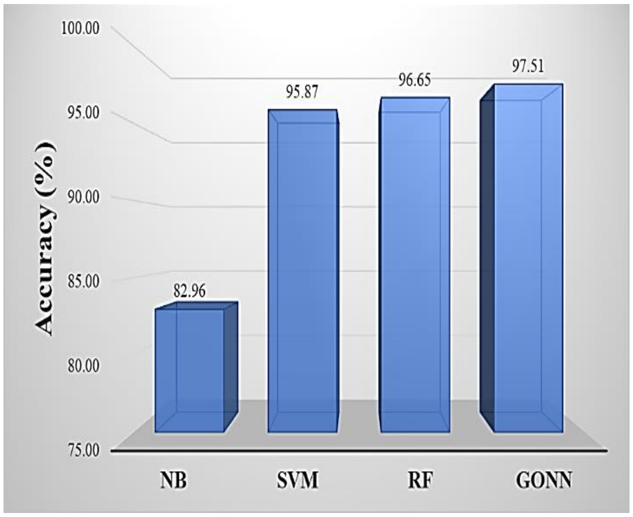

Fig.5 Performance comparison of the techniques regarding the accuracy

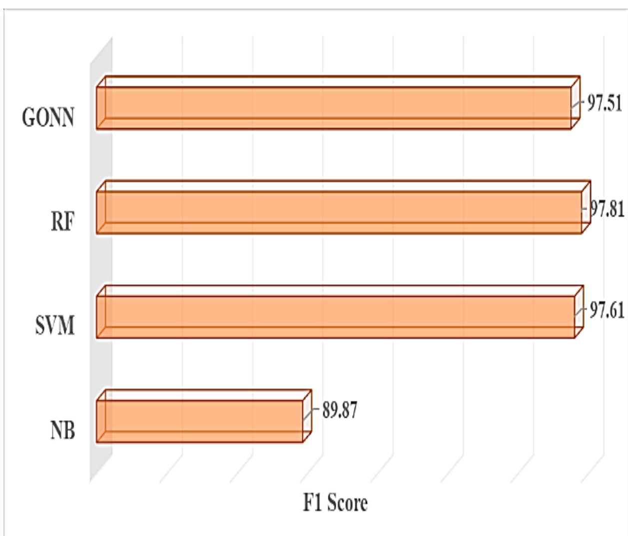

Fig.6 Performance comparison of the techniques regarding F1 score

Our work has also demonstrated the precision-recall curve for Naive Bayes, Support Vector Machine, and Random Forest diagrammatically. The precision-recall curve is a valuable measure to evaluate the performance of the model [26] visually. It has been effectively used in our work to overcome the limitations of an uneven dataset. The results shown in Fig.7 and Fig.8 illustrates that the average precision of SVM and RF comes out to be comparatively identical. Both are considered good classifiers to predict both the positive and negative classes. Although the average precision of Naive Bayes is slightly less, it is making minor prediction errors among the three methods (Fig. 9).

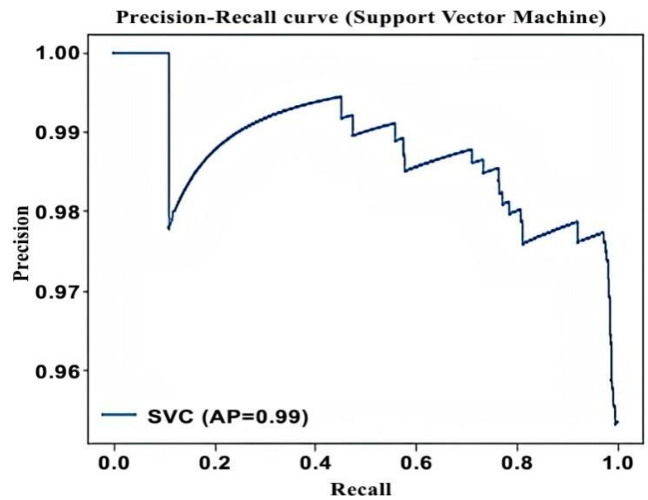

Fig.7 Precision-Recall Curve for Support Vector Machine

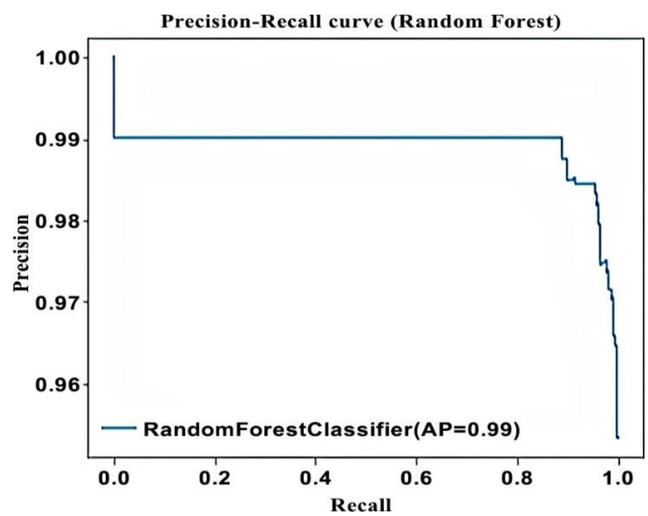

Fig.8 Precision-Recall Curve for Random Forest

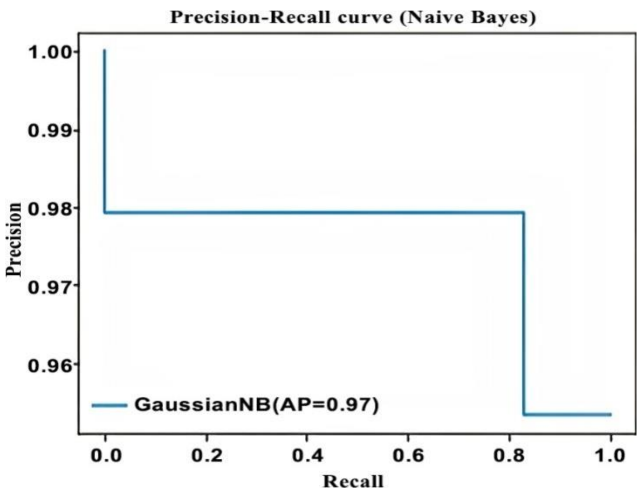

Fig.9 Precision-Recall Curve for Naive Bayes

Table 5: Classification performance comparison

Techniques	Accuracy	Precision	Recall	F1-score
[11]	82.32	84	78	76
[13]	94.17	92	92	92
[14]	83	93	73	81.79
[17]	82.04	77.93	82.4	80.1
[18]	96.36	93.87	95.47	94.66
GONN	97.51	96.07	98.98	97.51

Table 5 shows the comparison of classification performance in terms of accuracy, precision, recall, and F1-score of various techniques in the literature. The comparative analysis in Table 5 clearly shows that the proposed GONN techniques have better performance in all metrics. The best accuracy is achieved by the proposed GONN, which is 97.51%, whereas the second-best accuracy achieved by the technique used in [18] attained 96.36, which is almost 1.15% lesser than the proposed. So, it is evident that the proposed GONN has reasonable due to the effective learning mechanism using global optimization. Similarly, other metrics like precision, recall, and F1-score of proposed GONN is better than the other literary techniques. Based on these performance analyses, it is suggested that the proposed GONN is more suitable for the review analysis than the other techniques.

5- Conclusions and Future Scope

In this paper, we have exhibited the use of machine learning methods to extrapolate the sentiments over the Amazon dataset evoking the opinions and experiences of beauty product users. This empirical work has been carried out using data processing techniques, including stop word and punctuation removal, case folding, stemming, and tokenization in the first phase. Next, the feature extraction process has been implemented by using the Bag-of-Words model. A Global Optimization-based Neural Network (GONN) is proposed for the sentimental classification. Then an empirical study is conducted to analyze the consumers' sentiments towards our dataset by evaluating the performance of the proposed GONN and comparing it with the other machine learning algorithms, such as Random Forest (RF), Naive Bayes (NB), and Support Vector Machine (SVM). These methods categorized the reviews based on positive and negative polarity and cross-validated by ten folds. All the techniques used in our empirical work have been evaluated over the precision, recall, F1-score, and accuracy metrics, and the proposed method has offered the best accuracy results. In future work, we will expand our work by including neutral polarity reviews in the dataset and exploring its effect on the evaluation metrics. Future work will also consider the comparison of other machine learning classification methods and evaluate their performance. The framework implemented in this work will also be adapted for the reviews obtained from other domains.

References

[1] L. Yang, Y. Li, J. Wang and R. Sherratt, "Sentiment Analysis for E-Commerce Product Reviews in Chinese Based on Sentiment Lexicon and Deep Learning", IEEE Access, vol. 8, pp. 23522-23530, 2020. DOI: 10.1109/access.2020.2969854.

[2] T. U. Haque, N. N. Saber, and F. M. Shah, "Sentiment analysis on large scale Amazon product reviews," *2018 IEEE Int. Conf. Innov. Res. Dev. ICIRD 2018*, no. May, pp. 1–6, 2018, DOI: 10.1109/ICIRD.2018.8376299.

[3] J. Park, "Framework for Sentiment-Driven Evaluation of Customer Satisfaction With Cosmetics Brands", *IEEE Access*, vol. 8, pp. 98526-98538, 2020. DOI: 10.1109/access.2020.2997522.

[4] N. Nandal, R. Tanwar and J. Pruthi, "Machine learning based aspect level sentiment analysis for Amazon products", Spatial Information Research, vol. 28, no. 5, pp. 601-607, 2020. DOI: 10.1007/s41324-020-00320-2.

[5] M. Hu and B. Liu, "Mining and summarizing customer reviews," KDD-2004 - Proc. Tenth ACM SIGKDD Int. Conf. Knowl. Discov. Data Min., pp. 168–177, 2004, DOI: 10.1145/1014052.1014073.

[6] P. Jain, R. Pamula and G. Srivastava, "A systematic literature review on machine learning applications for consumer sentiment analysis using online reviews", Computer Science Review, vol. 41, p. 100413, 2021. DOI: 10.1016/j.cosrev.2021.100413.

[7] X. Fang and J. Zhan, "Sentiment analysis using product review data," J. Big Data, vol. 2, no. 1, 2015, DOI: 10.1186/s40537-015-0015-2.

[8] K. Jindal and R. Aron, "A systematic study of sentiment analysis for social media data", Materials Today: Proceedings, 2021. DOI: 10.1016/j.matpr.2021.01.048.

[9] W. Medhat, A. Hassan, and H. Korashy, "Sentiment analysis algorithms and applications: A survey," Ain Shams Eng. J.,

vol. 5, no. 4, pp. 1093–1113, 2014, DOI: 10.1016/j.asej.2014.04.011.

[10] Z. Liu, L. Liu, and H. Li, "An Empirical Study of Sentiment Analysis for Chinese Microblogging," Elev. Wuhan Int. Conf. E-bus., 2012.

[11] J. R. Ragini, P. M. R. Anand, and V. Bhaskar, "Big data analytics for disaster response and recovery through sentiment analysis," Int. J. Inf. Manage., vol. 42, no. September 2017, pp. 13–24, 2018, DOI: 10.1016/j.ijinfomgt.2018.05.004.

[12] P. Singh, R. S. Sawhney, and K. S. Kahlon, "Sentiment analysis of demonetization of 500 & 1000 rupee banknotes by Indian government," ICT Express, vol. 4, no. 3, pp. 124–129, 2018, DOI: 10.1016/j.icte.2017.03.001.

[13] P. Pugsee, P. Sombatsri, and R. Juntiwakul, "Satisfactory analysis for cosmetic product review comments," ACM Int. Conf. Proceeding Ser., vol. Part F1287, pp. 0–5, 2017, DOI: 10.1145/3089871.3089890.

[14] D. A. Kristiyanti and M. Wahyudi, "Feature selection based on Genetic algorithm, particle swarm optimization and principal component analysis for opinion mining cosmetic product review," 2017 5th Int. Conf. Cyber IT Serv. Manag. CITSM 2017, 2017, DOI: 10.1109/CITSM.2017.8089278.

[15] P. Pugsee, V. Nussiri, and W. Kittirungruang, Opinion mining for skin care products on twitter, vol. 937. Springer Singapore, 2019.

[16] R. Ren, D. D. Wu, and D. D. Wu, "Forecasting stock market movement direction using sentiment analysis and support vector machine," IEEE Syst. J., vol. 13, no. 1, pp. 760–770, 2019, DOI: 10.1109/JSYST.2018.2794462.

[17] N. Thessrimuang and O. Chaowalit, "Opinion representative of cosmetic products," 20th Int. Comput. Sci. Eng. Conf. Smart Ubiquitos Comput. Knowledge, ICSEC 2016, 2017, DOI: 10.1109/ICSEC.2016.7859945.

[18] T. Chatchaithanawat and P. Pugsee, "A framework for laptop review analysis," ICAICTA 2015 - 2015 Int. Conf. Adv. Informatics Concepts, Theory Appl., 2015, DOI: 10.1109/ICAICTA.2015.7335358.

[19] J. Ni, J. Li, and J. McAuley, "Justifying recommendations using distantly-labeled reviews and fine-grained aspects," *EMNLP-IJCNLP 2019 - 2019 Conf. Empir. Methods Nat. Lang. Process. 9th Int. Jt. Conf. Nat. Lang. Process. Proc. Conf.*, pp. 188–197, 2020, DOI: 10.18653/v1/d19-1018.

[20] E. Haddi, X. Liu, and Y. Shi, "The role of text pre-processing in sentiment analysis," Procedia Comput. Sci., vol. 17, pp. 26–32, 2013, DOI: 10.1016/j.procs.2013.05.005.

[21] Y. Zhang, R. Jin, and Z. H. Zhou, "Understanding bag-of-words model: A statistical framework," Int. J. Mach. Learn. Cybern., vol. 1, no. 1–4, pp. 43–52, 2010, DOI: 10.1007/s13042-010-0001-0.

[22] B. K. Bhavitha, A. P. Rodrigues, and N. N. Chiplunkar, "Comparative study of machine learning techniques in sentimental analysis," Proc. Int. Conf. Inven. Commun. Comput. Technol. ICICCT 2017, no. Icicct, pp. 216–221, 2017, DOI: 10.1109/ICICCT.2017.7975191.

[23] G. Tomassetti, and L. Cagnina, "Particle swarm algorithms to solve engineering problems: a comparison of performance," Journal of Engineering, vol. 2013, no. 1, pp. 1-13, 2013, DOI: 10.1155/2013/435104.

[24] H. Nguyen, R. Al, and K. Academy, "Comparative Study of Sentiment Analysis with Product Reviews Using Machine Learning and Lexicon-Based Approaches," SMU Data Sci. Rev., vol. 1, no. 4, 2018.

[25] J. D. Rodríguez, A. Pérez, and J. A. Lozano, "Sensitivity Analysis of k-Fold Cross Validation in Prediction Error Estimation," IEEE Trans. Pattern Anal. Mach. Intell., vol. 32, no. 3, pp. 569–575, 2010, DOI: 10.1109/TPAMI.2009.187.

[26] J. Keilwagen, I. Grosse, and J. Grau, "Area under precision-recall curves for weighted and unweighted data," PLoS One, vol. 9, no. 3, pp. 1–13, 2014, DOI: 10.1371/journal.pone.0092209.