Comparative Analysis of Machine Learning Algorithms for Conducting Sentiment Analysis on Product Reviews Derived from Social Media Data

Asha Patel
Faculty of Computer Applications
Ganpat University
Mehsana, India
asha.patel@ganpatuniversity.ac.in

Bhavesh Patel
Faculty of Computer Applications
Ganpat University
Mehsana, India
bhavesh.patel@ganpatuniversity.ac.in

Helly Patel
Faculty of Computer Applications
Ganpat University
Mehsana, India
hbp04@ganpatuniversity.ac.in

2023 5th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N) | 979-8-3503-3086-1/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICAC3N60023.2023.10541732

***Abstract*— Sentiment analysis has garnered considerable attention in recent years, primarily on account of the exponential expansion of user-generated content on social media platforms and the escalating demand for businesses to extract valuable insights from this data. This manuscript presents a comprehensive comparative examination of different machine learning algorithms employed in the task of sentiment analysis on product evaluations obtained from social media data. To execute the comparative analysis, a diverse dataset of product evaluations from various social media platforms is assembled and processed. Subsequently, this dataset utilizes a range of varied machine learning algorithms. The effectiveness of each algorithm is assessed through the use of fundamental metrics including F1-score, precision, accuracy and recall. The findings of the investigation uncover the merits and deficiencies of each computational procedure when employed in sentiment analysis tasks involving data derived from social media. This scientific inquiry delivers invaluable perspectives into the selection of machine learning procedures for sentiment analysis on social media-based evaluations of products. It functions as a manual for practitioners and scholars in the process of choosing the most appropriate procedure based on the unique attributes of their data and the goals of their analysis. Furthermore, it contributes to the broader comprehension of sentiment analysis and its significance in utilizing social media data for making business decisions.**

Keywords—Sentiment analysis, NLP, Positive Reviews, Negative Reviews, SVM, Classification, Accuracy

I. INTRODUCTION

Sentiment analysis, a form of text classification, is a crucial and fundamental aspect of Natural Language Processing (NLP) that involves the categorization of text based on the underlying emotional tone conveyed through opinions. Its significance lies in its ability to ascertain the emotional context of a given text, discerning whether it leans towards the positive or negative spectrum. This analytical process is colloquially referred to as opinion mining, as it aims to extract and interpret the viewpoint and attitude of customers. The foundation of sentiment analysis rests upon the collection and aggregation of opinions from users, a practice that proves invaluable in facilitating further enhancements. Social networks, in particular, serve as prominent platforms where users frequently express their opinions, and these user-generated posts are subsequently utilized for the purpose of classification [1].

Significant research endeavors are presently in progress in the field of sentiment analysis owing to its essentiality in competitive marketing and the constantly evolving consumer preferences. In order to conduct this particular research, data is obtained from the renowned online marketplace Amazon, specifically comprising reviews pertaining to tablets. After the acquired data has been subjected to the requisite preprocessing, a range of machine learning algorithms are employed in order to categorize the reviews into distinct negative or positive classifications. The proposed model for sentiment analysis facilitates customers in making prompt decisions regarding any product or service, as well as enabling business organizations to enhance the quality of their products by gaining a comprehensive understanding of customer perspectives.

II. BACKGROUND AND RELATED WORK

Khan et al. used BoW to extract features from movie reviews and created a vector. Then, they used Naive Bayes to categorize reviews into positive and negative classes. They created a weighted network based on semantic similarities between review sentences. They computed the score of each sentence using WGRA. The summary was generated through the process of carefully choosing the sentences that ranked the highest. The suggested approach outperformed other approaches [2].

Iqbal F et al. conducted an examination of a comprehensive framework that establishes a connection between machine learning approaches and lexicon-based approaches. The scholars developed a method for reducing features based on Genetic Algorithm (GA) in order to tackle the issue of scalability. The authors successfully diminished the feature set while upholding accuracy through the employment of the suggested hybrid approach. A comparison was made by the authors between their technique and feature reduction techniques based on Latent Semantic Analysis (LSA) and Principal Component Analysis (PCA). PCA encountered an augmentation of 15.4 percent in accuracy, whereas LSA experienced a 40.2 percent increase in accuracy. [3].

Munuswamy S et al. proposed a novel sentiment-based rating prediction method for constructing a recommendation system using user reviews on social media. The approach predicts preferred details based on user ratings. Sentiments are evaluated using a sentiment dictionary, and item reputations are determined using three sentiments, resulting in accurate recommendations. The n-gram methodology and SVM are integrated to enhance accuracy and classify reviews on social media platforms [4].

Nandal et al. introduced a distinctive approach that revolves around identifying sentiments at the aspect level, with a particular emphasis on the features of the item. The study employed Amazon customer reviews as a source of data, which were obtained through web crawling. Initially, the aspect phrases of each review were determined. The dataset underwent preprocessing procedures such as stemming, tokenization, casing, and stop word removal, in order to extract meaningful information. Subsequently, a supervised machine learning classification method was employed to assign a positive or negative class to each aspect [5].

Ayyub K et al. conducted an investigation into a variety of classifiers and collections of features in order to perform sentiment quantification in this particular study. Through the utilization of the feature set, an empirical assessment was performed on conventional machine learning (ML) approaches, ensemble-based approaches, and deep learning (DL) methods. The results obtained indicate that different collections of features have an impact on the performance of the classifier in sentiment quantification. Furthermore, the findings demonstrate that DL methods outperform the traditional machine learning algorithms [6].

III. A FRAMEWORK FOR ANALYZING SENTIMENTS: THE MODEL PERSPECTIVE

A systematic methodology is involved in the development and implementation of sentiment analysis models within the domain of natural language processing (NLP).The subsequent Fig. 1 illustrates the sequential methodologies for conducting sentiment analysis. The process begins with the collection of data, obtaining a dataset of text samples that have been labeled with sentiment categories. Subsequently, the data undergoes preprocessing to prepare it for analysis, which involves tokenization and other transformations. Feature extraction is then employed to convert the text into numerical representations that can be understood by machines. The next step involves selecting an appropriate machine learning or deep learning model, followed by training and evaluation phases to ensure the model's effectiveness.

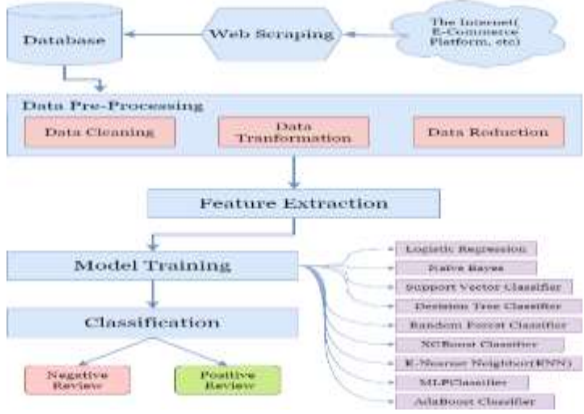

Fig.1.Framework for Sentiment Analysis Model

A. Data Collection

For the purpose of this study, the collection of data was carried out from the Amazon E-commerce Website by means of the process known as web scraping. Web scraping is a methodology employed to extract data by retrieving information from various websites. This process entails the automated retrieval of web content, often in the form of HTML or structured data, from one or more web pages.This Dataset encompasses a comprehensive compilation of more than thirty-four thousand consumer appraisals concerning Amazon products, specifically the Fire HD 8 Tablet equipped with Alexa technology, featuring an eight-inch high-definition display and a storage capacity of sixteen gigabytes.

B. Data Preprocessing

Data preprocessing is an integral part of the examination of data and the formulation of machine learning models. It involves the cleaning, transformation, and organization of raw data to make it suitable for analysis or model training. This step is important because it addresses issues such as missing values, outliers, and inconsistencies, which can affect the quality of the data. Additionally, data may need to be transformed through methods such as scaling, encoding categorical variables, or reducing dimensionality. In classification tasks, handling imbalanced data is crucial, while in natural language processing tasks, preprocessing the text is essential [7]. The proper preprocessing of data is vital because it directly influences the quality and reliability of subsequent analyses and machine learning models. Ultimately, this leads to more accurate and interpretable results and helps to mitigate problems associated with noisy or inconsistent data.

The procedure of Data Preprocessing encompasses several sequential steps, namely Data Sampling, Allocation of Sentiment Score to the Dataset, Purification of Text Process, Tokenization, Elimination of Stop words, Lemmatization, and ultimately Storing the resultant data into a fresh Dataset for subsequent processing.

C. Feature Extraction

Feature extraction within the sentiment analysis procedure involves the conversion of raw textual data into numerical characteristics that can be utilized by machine learning models. These characteristics encompass vital information derived from the text, such as the frequency of words or phrases, their significance within the document, or their contextual associations. Feature extraction plays a crucial role as it converts unstructured textual data into a structured format, rendering it suitable for modeling [8]. Through the representation of textual data as numerical features, sentiment analysis models are able to discern patterns and relationships within the data, thereby enabling the identification of sentiment and the generation of accurate predictions pertaining to positive or negative sentiments [14].

The vector representation has been generated by employing the TF - IDF (Term Frequency - Inverse Document Frequency) methodology for the purpose of feature extraction, as utilized for the specific objectives of this research [9][10].

D. Model Training

In the domain of sentiment analysis, the process referred to as 'model training' encompasses the core task of instructing a machine learning or deep learning model to identify and categorize sentiment within textual data. Throughout the process of model training, the specified algorithm or neural network acquires knowledge from a labeled dataset that consists of text samples and their corresponding sentiment labels, while adjusting its internal parameters through optimization techniques like gradient descent. The objective is to enable the model to identify underlying language patterns and subtleties that correspond to various sentiments, such as positive or negative. Once the model has been trained, it can then generalize its comprehension to formulate sentiment predictions on new, unfamiliar textual inputs, thereby laying the groundwork for accurate applications of sentiment analysis. The effectiveness of sentiment analysis heavily depends on the efficacy of this phase of model training in establishing a robust and precise predictive model.

The present study employs nine different machine learning algorithms to train the model. To achieve this objective, the dataset is divided into two separate subsets, specifically referred to as the training dataset and the testing dataset, with a proportion of 80:20.This section provides a concise overview of the nine Machine Learning Algorithms that have been employed to attain the intended results.

1) *Naive Baye:* Naive Bayes is a potent algorithm in classification, particularly in natural language processing and text analysis. It relies on Bayes' theorem and assumes independent features, resulting in practical success. Naive Bayes is effective for tasks like spam email detection, sentiment analysis, and document classification. It calculates the likelihood of an instance being assigned to a specific category based on feature probabilities. Despite its simplicity, Naive Bayes is accurate, efficient, and capable of handling high-dimensional data [11].

2) *Support Vector Machine:* The Support Vector Machine (SVM) is a robust algorithm in the field of machine learning that is highly effective for performing classification and regression tasks. SVM finds a clear separation boundary between classes in the data. It maximizes the margin between classes. SVM uses different kernel functions for linear and nonlinear classification tasks. Support Vector Machines (SVM) exhibit efficacy in handling datasets of both small and large sizes, making them prevalent in the domains of image recognition, text classification, and bioinformatics [11].

3) *Random Forest:* Random Forest is a widely respected machine learning algorithm known for its accuracy and flexibility in performing classification and regression tasks. It utilizes multiple decision trees in the training phase and combines their predictions using a voting or averaging method. This approach adopted by Random Forest is effective in mitigating over fitting and enhancing the model's capacity to generalize. Furthermore, Random Forest excels in managing data with high dimensions, effectively capturing intricate relationships, and displaying resistance to outliers. Additionally, it offers valuable insights into the importance of features, aiding in the selection of relevant features and facilitating data comprehension. Given these strengths, Random Forest finds wide application in diverse fields such as image classification, sentiment analysis, and finance, where reliable and interpretable predictions are of utmost importance [8].

4) Decision Tree Classifier: A Decision Tree is a crucial algorithm in machine learning for classification and regression tasks. It uses a tree-like structure to represent decisions, with internal nodes representing features and branches representing decision rules. Decision Trees are easily understood and useful for extracting insights from data. They can handle both numerical and categorical data [17].

5) XGBoost Classifier: The XGBoost Classifier is highly respected and widely used in machine learning. It is known for its accuracy and efficiency in classification problems. It is part of the ensemble learning family and can handle different data types and complex relationships in datasets. XGBoost builds decision trees and optimizes their combination to minimize prediction errors. It also uses regularization techniques to avoid over fitting and supports various objective functions, making it suitable for different tasks like classification, regression, and ranking [7].

6) *Logistic Regression:* Logistic Regression is a crucial algorithm in machine learning, primarily for binary classification tasks. The tasks aim to predict one of two potential outcomes. The logistic function is employed by this algorithm in order to model the probability of an input instance being part of the positive class. Logistic Regression is capable of effectively segregating classes by creating a linear decision boundary in feature space. Moreover, the model's interpretability is a standout advantage, providing a thorough understanding of how each feature influences classification decisions. Regularization techniques are commonly employed to prevent over fitting [16].

7) *K-Nearest Neighbor Classifier (KNN):* The K-Nearest Neighbor Classifier (KNN) is a simple but effective algorithm used in machine learning for classification and regression tasks. It determines the label of a data point by considering the majority class of its nearest neighbors in the feature space, typically measured using Euclidean distance.KNN is non-parametric, meaning that it does not impose strong assumptions regarding the underlying data distribution. This characteristic renders it adaptable to various types and structures of data. The selection of hyper parameter K, which determines the number of neighbors considered, can have a significant impact on its performance [13].

8) *MLP Classifier (Multi Layer Perceptron):* The MLP Classifier is a type of artificial neural network used in machine learning for supervised classification tasks. It consists of interconnected layers of nodes, called neurons, organized into input, hidden, and output layers. Each node performs a mathematical transformation on its inputs and passes the result to the next layer. MLPs are known for their ability to identify complex patterns and nonlinear relationships in data, making them suitable for various applications. They are trained using the back-propagation technique, whereby the network adjusts its internal weights and biases in order to minimize the prediction error during the training process. MLPs can be refined through the utilization of various activation functions, layer architectures, and regularization techniques, thus enabling them to model intricate decision boundaries and achieve a high level of predictive accuracy.

9) *AdaBoost Classifier:* AdaBoost, which stands for Adaptive Boosting, is an ensemble classifier designed for binary classification tasks. Its methodology involves combining various weak classifiers, such as shallow decision trees, by assigning higher importance to erroneously classified data points during each iteration. This iterative procedure culminates in the creation of a robust classifier that excels in managing intricate datasets, adapts effectively to noisy data, and exhibits resistance to over fitting. The versatility and widespread utilization of AdaBoost can be attributed to its remarkable ability to significantly enhance classification performance [18].

E. Classification

The process of categorizing sentiments in sentiment analysis is the pivotal stage wherein an educated model assigns predetermined sentiment labels to input textual data, commonly classifying sentiments as either positive or negative. The accuracy of sentiment classification relies on various factors, including the quality of the training data, the selection of the prototype, and the preference for attributes, thus encompassing applications that involve the analysis of customer reviews, the tracking of sentiment on social media, and the assessment of market sentiment to derive insights into public opinions and attitudes [15].

IV. EVALUATION METRICS

In the field of machine learning, assessing the quality and efficacy of a trained model requires an evaluation matrix or performance metric. This matrix consists of quantitative measures that provide valuable insights into the model's performance on a specific task. Common evaluation metrics in machine learning include F1-score, recall, accuracy and precision. These metrics assist data scientists and researchers in evaluating the model, understanding its strengths and weaknesses, and making informed decisions regarding model selection, hyper parameter tuning, and improvement.

In order to evaluate the efficacy of the model, we have employed the metrics of F1-score, Recall, Precision and Accuracy. The results of all the models are presented, which offer distinct guidance in the selection of a suitable algorithm for conducting sentiment analysis on a substantial amount of data. The following Fig. 2 elaborate on the equations used to compute the measures of Precision, Recall, F1-score, and Accuracy [12].

Fig.2.Evaluation Metrics Equations

V. FINDINGS AND PERFORMANCE OF THE MODEL

A confusion matrix functions as an invaluable tool for assessing the efficacy of classification algorithms, specifically in binary classification tasks, within the fields of machine learning and statistics. It provides a concise representation of the correlation between expected and actual class labels across a dataset. The confusion matrix is commonly employed to derive crucial potential metrics like F1-score, recall, accuracy and precision, facilitating a comprehensive evaluation of model performance [16].

The ensuing illustration in Fig. 3 portrays a range of confusion matrices while assessing these Machine Learning Algorithms through the utilization of the Term Frequency-Inverse Document Frequency (TF-IDF) feature extraction procedure.

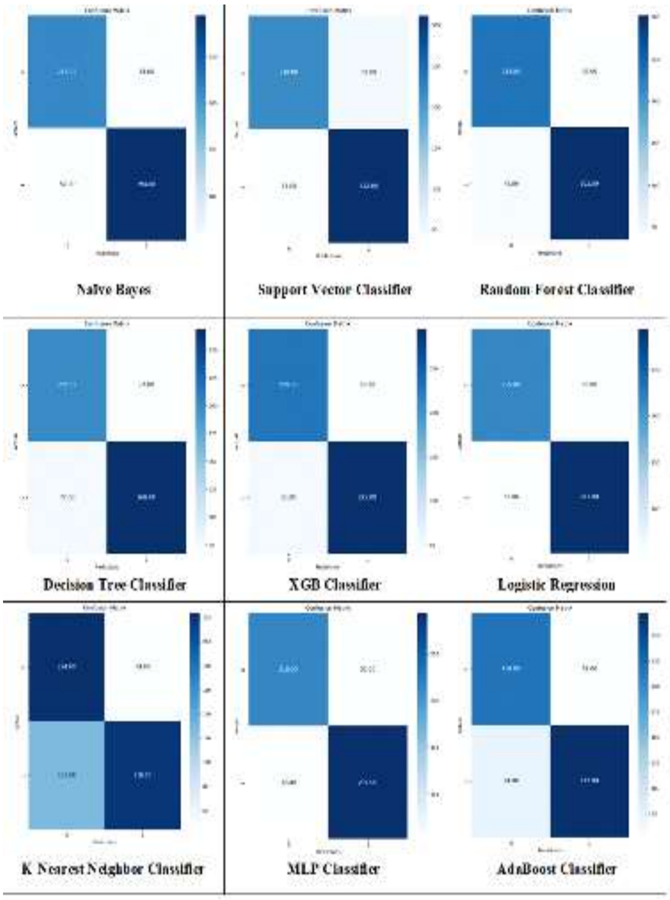

Fig.3.Confusion matrices derived by applying different models using TF-IDF

Table 1 presents the performance metrics of each distinct classifier employed in this particular investigation for the purpose of carrying out a comparative examination on the Dataset of Product Reviews.

TABLE I. COMPARISON OF PERFORMANCE METRICS FOR ALL THE CLASSIFIERS

Model	Precision	Recall	F1-Measure	Accuracy
Naive Bayes	0.84	0.85	0.85	82.41%
Support Vector Classifier	0.86	0.90	0.88	85.76%
Random Forest Classifier	0.90	0.87	0.88	87.11%
Decision Tree Classifier	0.80	0.78	0.79	76.37%
XGBoost Classifier	0.88	0.85	0.87	85.17%
Logistic Regression	0.85	0.85	0.85	82.45%
K-Nearest Neighbor (KNN) Classifier	0.83	0.63	0.72	73.44%
MLP Classifier	0.85	0.85	0.85	82.78%
AdaBoost Classifier	0.84	0.79	0.81	79.60%

The evaluation results of the classifiers are also presented in the subsequent Fig. 4, taking into consideration the parameter of Accuracy.

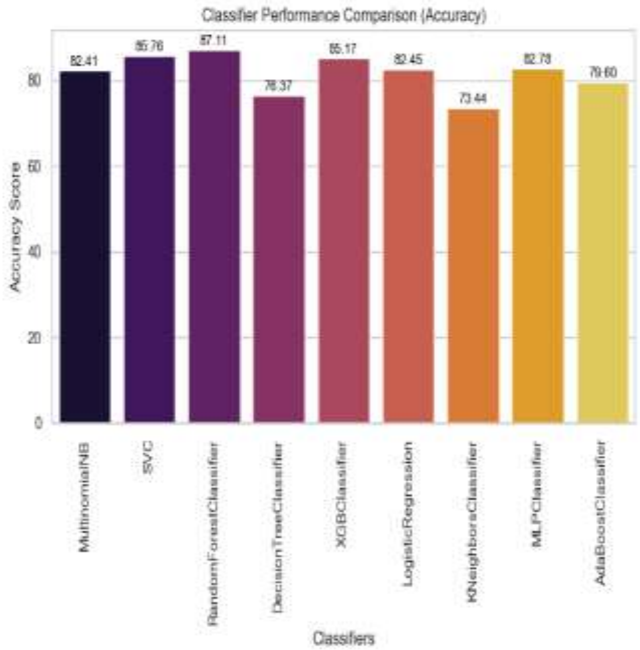

Fig..4.Assessment measures for all the classifiers

The analysis of the Performance indicates that the Random Forest Classifier exhibits the highest level of accuracy, standing at 87.11%, among all the classifiers considered. The Support Vector Classifier, on the other hand, offers the second highest level of accuracy, reaching 85.76%.

VI. CONCLUSION AND FUTURE DIRECTIONS

In this research, a comparative analysis was conducted on various machine learning methodologies in order to carry out sentiment classification. The purpose of utilizing the Amazon dataset was to achieve a higher level of accuracy in sentiment classification. The study employed machine learning techniques such as Naive Bayes, Support Vector Machine (SVM), Random Forest, Decision Tree, XGBoost, Logistic Regression, K-Nearest Neighbors (KNN), Multilayer Perceptron (MLP), and AdaBoost Classifiers. The findings of our analysis indicate that the Random Forest Classifier showcases remarkable expertise in sentiment classification, reaching the highest level of accuracy. Furthermore, the precision of sentiment analysis can be further enhanced by incorporating Deep Learning Techniques and applying Fuzzy Logic during the classification process. Future research will focus on exploring these avenues, leading to an increased accuracy of the classifier.

REFERENCES

[1] Devika, M. D., Sunitha, C., & Ganesh, A. (2016). Sentiment analysis: a comparative study on different approaches. Procedia Computer Science, 87, 44-49.

[2] Atif K., Muhammad A. G., Mahdi Zz. ,2 R. R. Biswal, AsimZ.,Muhammad N., Yousaf S. and Naomie S., “Movie Review Summarization Using Supervised Learning and Graph-Based Ranking Algorithm,” Computational Intelligence and Neuroscience, vol. 2020,pp.1–14, June 2020.

[3] F. Iqbal et al., "A Hybrid Framework for Sentiment Analysis Using Genetic Algorithm Based Feature Reduction," in IEEE Access, vol. 7,pp. 14637-14652, Jan 2019.

[4] Munuswamy S., Saranya M. S., Ganapathy S., Muthurajkumar S.,Kannan A.,“Sentiment Analysis Techniques for Social Media-Based Recommendation Systems,” National Academy Science Letters, vol. 3, pp. 281–287, July 2020.

[5] Nandal N., Tanwar R., Pruthi J., “Machine learning based aspect level sentiment analysis for Amazon products,” Spatial Information Research, vol.5, pp. 601–607, Feb 2020.

[6] Ayyub K., Iqbal S., Munir E. U., Nisar M. W., Abbasi M., “Exploring Diverse Features for Sentiment Quantification Using Machine Learning Algorithms, “ IEEE Access, vol. 8, pp.142819–142831, July 2020.

[7] Agarwal, A., &Meena, S. (2022, November). A Comparative Study of Deep Learning and Machine Learning Algorithm for Sentiment Analysis. In 2022 3rd International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT) (pp. 1-7). IEEE.

[8] Kaur, G., & Sharma, A. (2022, April). Comparison of Different Machine Learning Algorithms for Sentiment Analysis. In 2022 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS) (pp. 141-147). IEEE.

[9] Maada, L., Al Fararni, K., Aghoutane, B., Fattah, M., &Farhaoui, Y. (2022, March). A comparative study of Sentiment Analysis Machine Learning Approaches. In 2022 2nd International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET) (pp. 1-5). IEEE.

[10] Kaushik, B., Sharma, A., Chadha, A., & Sharma, R. (2023, March). Machine Learning Model for Sentiment Analysis on Mental Health Issues. In 2023 15th International Conference on Computer and Automation Engineering (ICCAE) (pp. 21-25). IEEE.

[11] Dhola, K., &Saradva, M. (2021, January). A comparative evaluation of traditional machine learning and deep learning classification techniques for sentiment analysis. In 2021 11th international conference on cloud computing, data science & engineering (Confluence) (pp. 932-936). IEEE.

[12] Hossain, M. J., Joy, D. D., Das, S., & Mustafa, R. (2022, February). Sentiment analysis on reviews of e-commerce sites using machine learning algorithms. In 2022 International Conference on Innovations in Science, Engineering and Technology (ICISET) (pp. 522-527). IEEE.

[13] Bahel, V., Pillai, S., &Malhotra, M. (2020, June). A comparative study on various binary classification algorithms and their improved variant for optimal performance. In 2020 IEEE Region 10 Symposium (TENSYMP) (pp. 495-498). IEEE.

[14] Jain, A. P., &Dandannavar, P. (2016, July). Application of machine learning techniques to sentiment analysis. In 2016 2nd International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT) (pp. 628-632). IEEE.

[15] Jain, K., &Kaushal, S. (2018, August). A comparative study of machine learning and deep learning techniques for sentiment analysis. In 2018 7th International conference on reliability, infocom technologies and optimization (Trends and Future Directions)(ICRITO) (pp. 483-487). IEEE.

[16] Bari Antor, M., Jamil, A. H. M., Mamtaz, M., Monirujjaman Khan, M., Aljahdali, S., Kaur, M., ...&Masud, M. (2021). A comparative analysis of machine learning algorithms to predict alzheimer’s disease. Journal of Healthcare Engineering, 2021.

[17] Sheth, V., Tripathi, U., & Sharma, A. (2022). A Comparative Analysis of Machine Learning Algorithms for Classification Purpose. Procedia Computer Science, 215, 422-431.

[18] Ghosh, M., Raihan, M. M. S., Raihan, M., Akter, L., Bairagi, A. K., Alshamrani, S. S., &Masud, M. (2021). A Comparative Analysis of Machine Learning Algorithms to Predict Liver Disease. Intelligent Automation & Soft Computing, 30(3).