International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

An Approach for Product Recommendation using Light GBM

Dr. I. S. Siva Rao[1], Parasa Rajya Lakshmi[2], Dr. Dasari N. V. Syma Kumar[3], Akkala Yugandhara Reddy[4], Dr. Jayavarapu Karthik[5], Badugu Bhavana[6]

Submitted: 22/12/2023 **Revised**: 28/01/2024 **Accepted**: 05/02/2024

Abstract: Attracting clients is the main task of online e-commerce websites. Systems for providing recommendations are essential for engaging clients. Customer reviews play a crucial role in analyzing the product. Product insights can be provided by sentiment analysis of customer reviews. Websites routinely recommend products despite bad user reviews, which dissatisfy customers. Hence there is a need for a more accurate model recommending the products. In this work, a machine learning model is proposed that suggests a product with a greater user sentiment for positivity. Models are developed to analyze the sentiment of product reviews using the algorithms ADABoost, Light GBM, Gradient Boosting, Extreme Gradient Booting, and Extreme Gradient Boosting coupled with Random Forest. Based on the performance of the models, the Light GBM model is considered for building the product recommendation system. The proposed model gave better results when compared to existing models.

Keywords: *ADABoost, Extreme Gradient Boosting, Gradient Boosting, Light GBM, Random Forest, Recommendations.*

1. Introduction

Sentiment analysis [1, 2, 4, 5, 7, 8, 18, 19, 21, 22, 23, 24, 25, 26] is the process of analysing the text and determining the viewpoint of the text. The sentiment analysis [1, 2, 4, 5, 7, 8, 18, 19, 21, 22, 23, 24, 25, 26] is used in many areas of research. The tone of the text can helps us to know the preferences of the customers. These days the purchase decisions are generally made by reading reviews given by the fellow customers. The user review may have positive sentiment, neutral sentiment or negative sentiment. There are three main approaches for sentiment analysis that include lexicon approach, hybrid and machine learning approaches. This work includes working with machine learning algorithms.

The XGBoost algorithm [4] is a boosting mechanism. It's a well-known supervised machine-learning (ML) algorithm. The gradient-boosted trees approach is efficiently implemented in XGBoost. On large datasets, it can be utilized for both regression and classification. XGBoost provides reliable results while avoiding overfitting by using sequentially created decision trees.

Random Forest [2, 20] is an established ML algorithm that employs supervised learning [27]. The random forest algorithm can be used to tackle both classification and regression problems [27]. It is a set of decision trees. Instead of depending on a single decision tree, the random Forest classifier accepts predictions from all trees and concludes the final output based on the predictions of the majority of trees.Adaptive Boosting in short called AdaBoost is a statistical meta-algorithm that is frequently used to handle classification problems. The AdaBoost algorithm initially builds a model and gives every data point the same amount of weight. Later it assigns larger weights to data points that are wrongly classified. Now, the more heavily weighted data points are given priority in the consequent model, and this mechanism is continued till the error is negligible.

Light GBM (Gradient Boosting is a decision tree-based machine-learning algorithm. Light GBM is an implementation of decision trees where the tree grows leaf-wise, which means only a single leaf gets split every time. Light GBM performs extremely well with larger datasets. Light GBM is sensitive and overfits small datasets, but this problem can be avoided by limiting the depth of the tree. The Light GBM algorithm mainly focuses on the accuracy.

Gradient Boosting is a well-known ML boosting algorithm. It is more often used in classification and

[1] *Associate Professor, Department of CSE, GITAM University, Rushikonda, Visakhapatnam.*
[2] *Assistant professor, Department of Information Technology, Prasad V. Potluri Siddhartha Institute Of Technology, Vijayawada, AP.*
[3] *Associate Professor, Department of CSE, Bapatla Engineering College, Bapatla, AP, India.*
[4] *Assistant professor, Department of Computer Science and Engineering, Chirala Engineering College, Chirala, A.P, India.*
[5] *Associate Professor, Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, A.P, India.*
[6] *Department of CSE, Velagapudi Ramakrishna Siddhartha Engineering College, Vijayawada, India.*
** Corresponding Author Email: badugubhavana99@gmail.com*

regression problems. It is based on the assumption that the overall prediction error is decreased when earlier models are combined with the best upcoming model. It's essential to specify the anticipated outcomes for this future model in order to minimize errors.

2. Literature Survey

A deep learning network to forecast the opinion or feelings of a product was presented in 2020 by the authors of the work [1]. The Deep Learning Modified Neutral Network (DLMNN) is used to implement the proposed model, and IANFIS is then used to enhance it. In this paper, the authors conducted a review analysis using the Food Review Dataset. Grade-based, content-based, and collaboration-based categories were used to separate the data, which was then subjected to review analysis utilizing a variety of natural language processing approaches. Precision, accuracy, recall, and f-score are just a few metrics that are used to evaluate the model.

The authors of the article [2] aimed to solve the sentiment polarity categorization issue in 2015. The proposed approach is assessed using data gathered from Amazon.com product reviews. They conducted tests for classification at the sentence and review levels and offered insightful information. Three machine learning algorithms—Naive Bayesian, SVM, and Random Forest—are employed in the presented work.

In 2022, the authors of the journal [3] conducted a systematic literature survey on existing studies in big data analytics. In this work, the authors presented a framework of a systematic literature review of multi-disciplinary outlooks. One is an organizational outlook that studies the theoretical foundations and different research models that gives insights about the performance using big data analytics and the other one is a technical outlook that explains the big data techniques, and different algorithms developed for supply chain functions. They also provided directions for future research in big data analytics.

A sympathetic approach to data analysis was offered by the writers of the journal [4] in 2020. It combines the neutrality detector model used to preprocess the data with extreme Gradient Boosting and a genetic algorithm. Through an Arabic social network, the model successfully predicted and analyzed the sentiment of e-payment service users. The dataset combines information from Twitter and Facebook for the years 2017 and 2019, respectively. The analysis of client satisfaction with Jordan's e-services was aided by this study. Accuracy, precision, and recall are performance factors taken into account when assessing the proposed system.

In order to determine the future extent of consumer sentiment analysis on online customer evaluations in the tourist and hospitality industries, the authors of the article [5] presented a study in 2021 that explored, analyzed, and compared previous attempts at filling research gaps. This work provided comprehensive information on the customer sentiment analysis research conducted in the past and planned for the future. The information used comes from a variety of sources, including airline reviews, hotel reviews, travel reviews, etc.

The writers of the article [6] looked at how different factors affect economic outcomes and how reviews affect product sales in 2011. In order to identify key components, they also looked into a variety of factors, including subjectivity, readability, and spelling problems in the review text. The study found that having a lot of subjective information hurts a product's sales. In order to forecast how the review content will affect product sales, they used random forest classifiers. The importance of three feature categories—including reviewer-related elements, review subjectivity, and review readability—was examined. They discovered that employing just one of the three can produce a performance that is comparable to using all three. The data is gathered from amazon.com.

A methodology for analyzing product reviews was given in 2018 by the authors of the paper [7]. In order for the designers to make the best decisions and try to determine what the needs of the clients are. Large amounts of qualitative content were to be converted into quantitative perceptions using the proposed framework. Amazon product reviews were selected to verify the framework's effectiveness. A bridge between qualitative and quantitative results was created by the framework.

The author of the research [8] concentrated on creating aspect term extraction of topic models by taking into account product specifics in 2018. They put out the SA-ASM and SA-PSM models, two models that were extended from the Aspect Sentiment Unification Model. Reviews of laptops and mobile devices are used to assess the suggested model. According to the findings, the seller-aided product-based Sentiment model works better in highlevel aspects like category recognition, while the seller-aided Aspect-based Sentiment model performs better in smaller aspects like sentiment categorization.

The study's authors [9] looked at how Internet product reviews affected product sales in 2019. In this study, topics in the online review content are extracted using a joint sentiment-topic model. The study provided a good understanding of how electronic word-of-mouth influences product sales as well as how numerical rating and review texts contribute to product sales.

Based on methods of sentiment analysis and intuitionistic fuzzy set theory and taking into account the online reviews of the products, the authors of the work [10] ranked the products in 2016. To enhance the evaluations, preference ranking organization techniques are also employed. To demonstrate the effectiveness of the proposed approach, a case study is also taken into consideration.

The writers of the article [11] stated the effect of product reviews on sales in 2019. A recommendation system utilizing the ensemble approach was proposed in the work. From several official sources, the authors gathered product reviews. Data pre-processing is first carried out, then classifiers like the Naive Bayes and SVM algorithms are trained. The authors discovered that the aforementioned models' performance is ineffective in the interim. As a result, the authors recommended a model employing Support Vector Machine and Naive Bayes as an ensemble technique. In comparison to earlier models, the suggested approach yielded better outcomes.

The significance of data mining techniques in customer review analysis was examined by the research's authors [12] in 2021. The authors gathered customer feedback from internet shopping portals. From 2015 through 2020, they examined and analyzed customer reviews posted online. They found that more study on the analysis of online customer evaluations has been done in recent years. The issues and research trends are highlighted in this study. The authors also offered information on upcoming studies in the study of online consumer reviews.

The sentiment analysis of the corpus's textual data was examined in 2019 by the journal's research team [13]. The authors trained the data using a variety of machine-learning techniques, including KNN, NB, SVM, LR, RF, and DT. They discovered that their system outperformed the basic system by about 9%. To clean textual data, they employed a variety of natural language processing techniques, including tokenization, lemmatization, and n-gram techniques. F-score and accuracy measures are used to assess the system's performance.

The contributors of the article [14] researched the value of sentiment analysis for addressing real-time problems in 2017. They claimed that because the amount of data coming from different sources is growing daily, it is necessary to analyze the data in order to gather meaningful information. Recent research on sentiment analysis utilizing deep learning models, such as convolutional neural networks, recursive neural networks, and deep learning neural networks, was emphasized by the authors in this article. Recent studies have focused on a range of real-time issues, such as sentiment classification, textual, visual, and multilingual issues, as well as product review analysis.

To determine the opinion on the textual data, the authors of the work [15] brought up the Enhanced XGBoost model and the Tailored Random Forest model in 2020. These algorithms are trained and tested using customer tweets. The authors discovered that the Enhanced XGBoost model [27] outperforms Tailored Random Forest and Naive Bayes in classifying the sentiment of consumer tweets. The effectiveness of the above models is evaluated using measures like accuracy.

The user is unable to predict the products based on customer reviews because of inaccurate decision-making processes. Using different decision tree classifiers, the authors of the work [16] suggested a solution in 2019. The three models of decision tree classifiers that are employed are Random tree, Hoeffding tree, and Adaptive boosting + Random tree. Researchers discovered that the suggested models perform better than more conventional methods. A DOM parser is used to extract the data from Amazon product reviews. There are several feature extraction methods employed, including FD and AFINN. Different sentiment analysis methods, such as NPRS, are applied. Utilizing evaluation measures like kappa statistics, F-measure, and error rate, the presented models are evaluated.

3. Methodology

The proposed methodology is divided into two modules. They are; Data pre-processing [3, 5] and

Data Visualization module and Model Building using machine learning algorithms.

3.1. Data pre-processing and Data visualization

The data is collected from various sources. The dataset contains reviews of different products [2] like household chemicals, movies, music, books, hair care, medicines, lubricants, and so on. The dataset contains around 30000 records of different product reviews [7, 19, 22]. Since the data is collected from different sources it contains unwanted data. Hence the data is pre-processed using many natural language processing techniques to extract useful information.

Some columns in the dataset have missing rows, which may be due lack of information at the time of entry. These missing rows may create noise and can affect the performance of the model. Hence all the missing rows are calculated and removed. The dataset contains supervised data. The class labels given are positive and negative. Hence the class labels are mapped to binary values. The positive sentiment to the class label 1 and the negative sentiment to the class label 0.

The dataset contains attributes like the manufacturer, brand, review text, review ratings, user sentiment, and so on. On observation, it was found that some reviews contain negative user sentiment but, the review rating is greater than or equal to 4. Similarly, the reviews with positive user sentiment are having review ratings of less than 4. These inconsistencies can make the model error-prone. Hence, the customer reviews [7, 19, 22, 23, 24, 25, 26, 28] with a user sentiment of 1 and a review rating of less than 4 are mapped to class label 0. Similarly, the customer reviews that have a user sentiment of 0 and a review rating greater than or equal to 4 are mapped to class label 1.

Unfortunately, the dataset is suffering from a class imbalance [27]. The percentage of positive user sentiment reviews is larger when compared to negative user sentiment reviews. Hence the model will be trained rigorously on positive reviews and less on negative reviews which may affect the performance of the model. Hence Synthetic Minority Oversampling Technique (SMOTE) Algorithm is used to over-sample the minority class. It resulted in creating artificial samples of class label 0 so that the model can be trained rigorously even on negative reviews.

There are more than 250 products in the dataset, and each product contains may contain positive and negative reviews. The sentiment percentage of each product is calculated to analyze the viewpoint of the product. The formula that is used to calculate the sentiment percentage is shown below.

*Sentiment% = Positive revies count*100/Total count of the product* (1)

As mentioned earlier the dataset contains various features like brand, manufacturer, review text, review title, review rating, and so on. The review title and the review text are combined for better analysis of data. The review texts are cleaned by removing unwanted punctuations.

Stop-word removal [1, 4] is a technique of natural language processing. The textual data contains a large number of stop-words which may not much contribute to the analysis. Hence all these stop-words are removed using the NLTK library. As we know the NLTK library defines all the possible stop-words, and it becomes easier to remove the stop-words from textual data. Every word of the review text is mapped to its root word by considering the position tags of each word. This process is known as lemmatization. Initially, the position tags [2] are mapped for each word and then passed to the lemmatizer function to map that word into its root form.

A Word cloud [7] is a cluster of words that gives us insights into the frequency of words in textual data. A word cloud of 40 words is generated among all the articles after the processing of text. The size and thickness of words in the word cloud depend upon the frequency of words in the corpus.

N-gram frequency [1] is a method of NLP that tells how frequently n words are occurring together in the corpus. The bigram frequency and trigram frequency are calculated for reviews text to identify the words that are frequently occurring in the corpus. Features [21] of the product also play a key role changing the opinion of the customers. Some people mention features of the product in reviews. The important features from all the product reviews are extracted using TF-IDF vectorizer. TF-IDF refers to Term Frequency - Inverse Document Frequency [4]. From 30000 product reviews around 650 features are extracted with max_df of 0.9, min_df of 7, and ngram range of 2.

TF(t, d) = The frequency of a word in a documbnet/ Total number of words in document (2)

$$IDF = \log \frac{N}{1 + df} \quad (3)$$

Where *t* is the word, *d* is the document, *N* is the total number of documents present in the corpus, and *df* is the documents that contain the word.

3.2. Model Building

The entire data is divided into training data and testing data. The 75% of the entire data is considered for training and 25% of data is used for testing. A model is developed and four boosting algorithms along with one ensemble method is considered for training and Evaluation of the model. The algorithms that are used are Adaptive Boosting (AdaBoost), Gradient Boosting, Extreme Gradient Boosting (XGBoost), XGBoost-RF (An ensemble technique of Extreme Gradient Boosting and Random Forest), and Light GBM. The mechanisms for each algorithm and the parameters are discussed below.

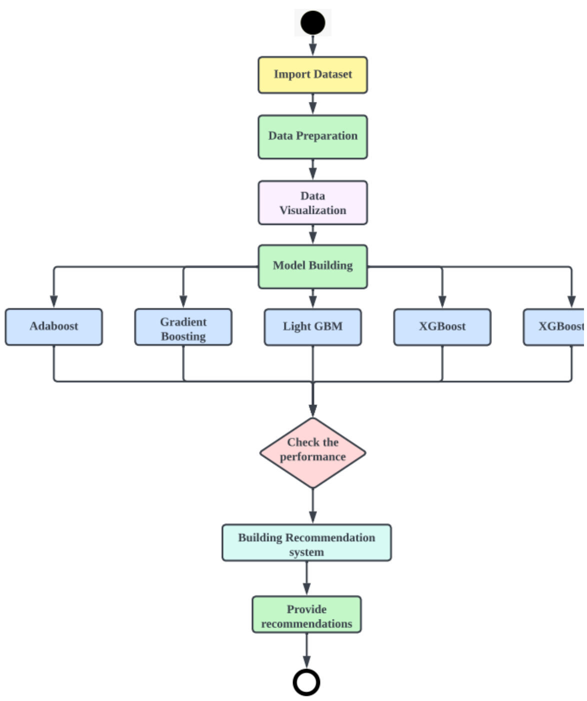

Fig 1. The process flow diagram

The AdaBoost method first creates a model and gives each data point an equal weight. Later, it gives incorrectly categorized data items heavier weights. The resulting model now prioritizes the data points with larger weights, and this method is continued until the error is minimal. The hyper parameters that are considered are learning rate of 0.15, estimators of 510, and random state of 44. A confusion matrix [4] to check the effectiveness of the adaboost model is generated.

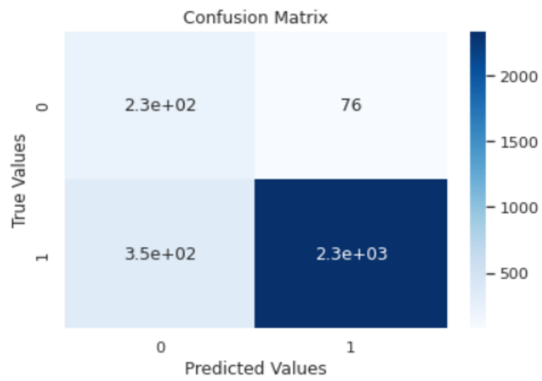

Fig 2. The confusion matrix obtained from adaboost model

Gradient boosting is based on the hypothesis that the overall prediction error is reduced when older models are combined with the best forthcoming models are combined with the best upcoming model. It's crucial to specify the anticipated outcomes for the subsequent model in order to minimize errors. The parameters that are considered for training of the model are learning rate of 0.15, estimators of 510, and random state of 44. A confusion matrix [4] that is generated for gradient Boosting model is shown below.

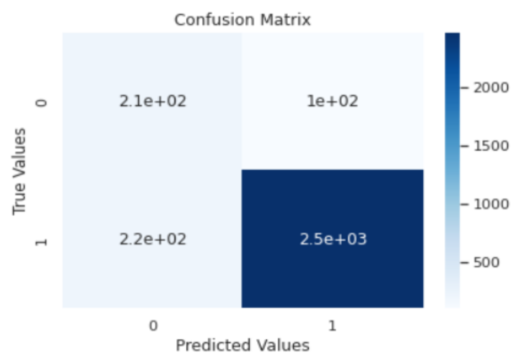

Fig 3. The confusion matrix obtained from Gradient Boosting model

In XGBoost [27], the gradient-boosted trees method is practiced. It can be used for classification and regression on huge datasets. XGBoost uses sequentially built decision trees to produce accurate results while preventing overfitting [27]. Using the xgboost model we trained and tested the model. The hyper parameters that are considered are learning rate of 0.15, maximum depth of 60, and a random state of 44. The generated confusion matrix [4] of xgboost model is shown below.

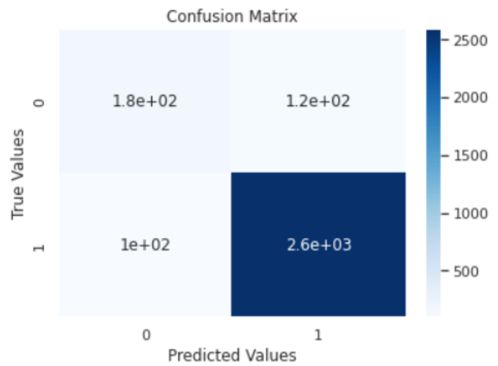

Fig 4. The confusion matrix obtained from XGBoost model

An ensemble method is a combination of two algorithms. The Extreme gradient Boosting algorithm[27] is combined with Random forest [20, 27] and tested against the data. The random forest is an tree based machine learning algorithm the collects the outputs from multiple decision trees and consolidate the output. The attributes that are used in the model are learning rate of 0.15, maximum depth of 135, and the random state of 44. The confusion matrix [4] is generated which gives the insights into the effectiveness of the model is shown below.

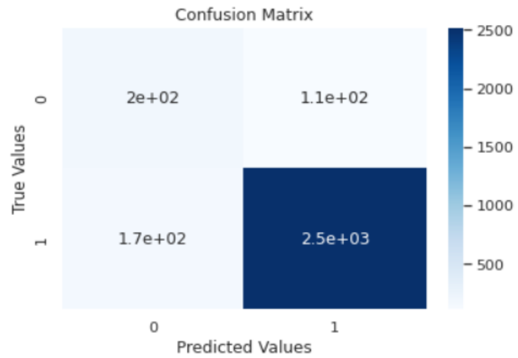

Fig 5. The confusion matrix obtained from XGBoost-RF model

Light GBM grows in a leaf-wise fashion. This implies that only one leaf is split each time. With larger datasets, Light GBM performs incredibly well. By restricting the depth of the tree, light GBM can prevent its sensitivity and overfitting of small samples. The accuracy of the results is the major focus of the Light GBM method. The hyper parameters that are considered are learning rate of 0.2, maximum depth of 13, and estimators of 900. The light GBM performed really well on the following dataset. The confusion matrix [4] generated is shown below.

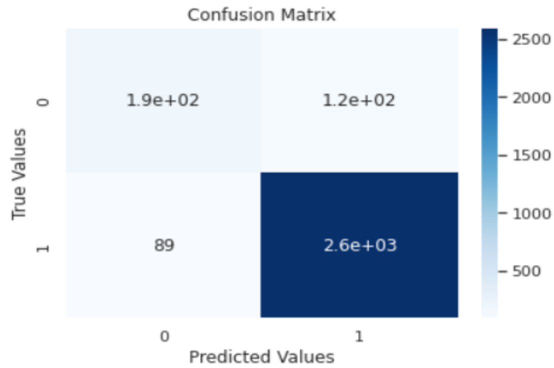

Fig 6. The confusion matrix obtained from LightGBM model

4. Results

The performance of each model is evaluated using the metrics like accuracy, precision, recall, F1-score, AUC score, computational time, and AUC score.

4.1. Accuracy

The accuracy [27] of all the models is calculated using the below formula. The values of TP, TN, FP, and FN are taken from the results of confusion matrix, where TP stands for True Positive values, TN refers to True Negative values, FP refers to False Positive values, and FN refers to False Negative values.

$$Accuracy = TP + TN/ TP + TN + FP + FN \qquad (4)$$

Table 1. Accuracy

Model	Accuracy
Adaboost	0.86
Gradient boosting	0.90
XGBoost	0.92
XGBoost-RF	0.90
Light GBM	0.93

4.2. Precision

The precision [27] of the model is calculated using the below formula. The values obtained from all the models are represented in the table below.

$$Precision = TP / TP + FN \qquad (5)$$

Table 2. Precision

Model	Precision
Adaboost	0.97
Gradient Boosting	0.96
XGBoost	0.95
XGBoost-RF	0.95
Light GBM	0.95

4.3. Recall

The recall [27] of the model is calculated using the below formula. The values obtained from all the models are represented in the table below.

$$Recall = TP/\ TP + FP \quad (6)$$

Table 3. Recall

Model	Recall
Adaboost	0.87
Gradient Boosting	0.93
XGBoost	0.96
XGBoost-RF	0.94
Light GBM	0.97

4.4. F1-Score

The F1-score of the model is calculated using the below formula. The values obtained from all the models are represented in the table below.

$$F1\ -\ score = 2TP/\ 2TP + FP + FN \quad (7)$$

Table 4. F1-Score

Model	F1-Score
Adaboost	0.92
Gradient Boosting	0.94
XGBoost	0.96
XGBoost-RF	0.95
Light GBM	0.96

4.5. Computational Time

The computational time is referred as the time taken for the system to run the program. The computational time of all the models is observed and represented in the table below.

Table 5. Computational Time

Model	Computational Time
Adaboost	59 sec
Gradient Boosting	130 sec
XGBoost	102 sec
XGBoost-RF	140 sec
Light GBM	52 sec

4.6. AUC Score

The Area under the curve of the model is calculated that gives the degree of separability. It gives the performance of classification models at different thresholds. The graphs obtained from all the models are represented in the table below.

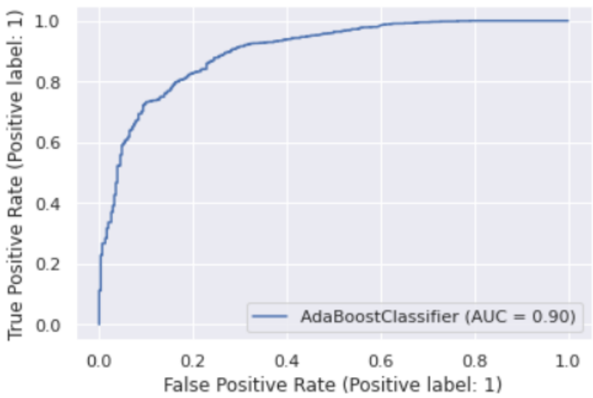

Fig 7. The AUC curve obtained for Adaboost classifier

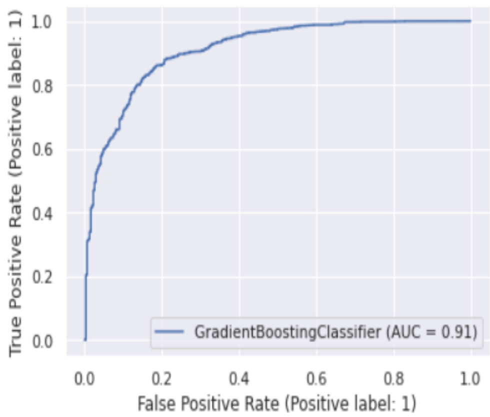

Fig 8. The AUC curve obtained for Gradient boosting

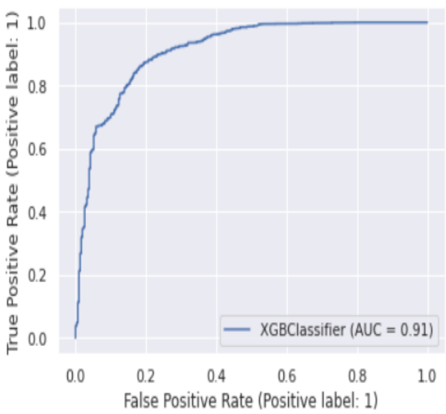

Fig 9. The AUC curve obtained for XGboot classifier

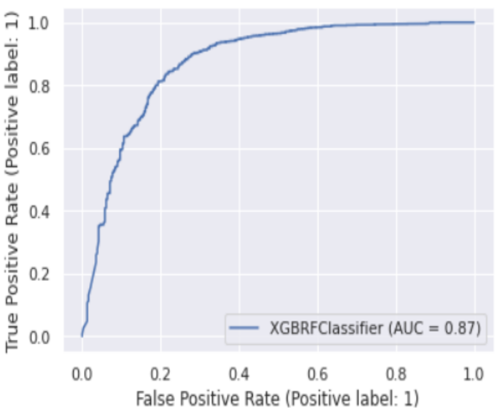

Fig 10. The AUC curve obtained for XGBoost-RF model

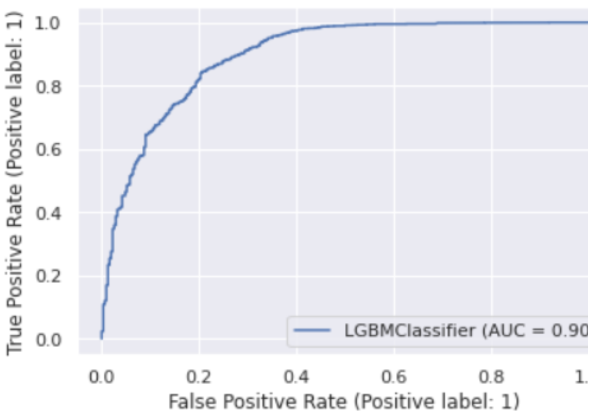

Fig 11. The AUC curve obtained for Light GBM model

5. Conclusion

To determine performance, all models, including Adaboost, Gradient Boosting, Light GBM, XGBoost, and XGBoost-RF, are tested against data [27]. The product recommendation system [17, 22] can be built using the Light GBM model since it performed well when compared to other models. The measures considered while evaluating the model's performance include precision, recall, accuracy, and F1 measure. The accuracy, precision, recall, and F1-score performance metrics of the model are 0.93, 0.95, 0.97, and 0.96 respectively. 52 seconds were spent on the Light GBM model's computations overall. When compared to current models, the suggested model is capable of making recommendations that are more accurate.

References

[1] P. Sasikala, L. Mary Immaculate Sheela, "Sentiment analysis of online product reviews using DLMNN and future prediction of online product using IANFIS", Journal of Big Data, pp. 1-20.

[2] Xing Fang, Justin Zhan, "Sentiment analysis using product review data", Journal of Big Data.

[3] In Lee, George Mangalaraj, "Big Data Analytics in Supply Chain Management: A Systematic Literature Review and Research Directions", Big data and cognitive computing.

[4] DANA A. AL-QUDAH, ALA M. AL-ZOUBI, PEDRO A. CASTILLO-VALDIVIESO, HOSSAM FARIS, "Sentiment Analysis for e-Payment Service Providers Using EvolutionaryeXtreme Gradient Boosting", IEEE Access volume 8, 2020.

[5] Praphula Kumar Jain, Rajendra Pamula, Gautam Srivastava, "A systematic literature review on machine learning applications for consumer sentiment analysis using online reviews", Computer Science Review 41 (2021) 100413.

[6] Anindya Ghose, Panagiotis G. Ipeirotis, "Estimating the Helpfulness and Economic Impact of Product Reviews: Mining Text and Reviewer Characteristics", IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, vol 23, no 10.

[7] Robert Ireland, Ang Liu, "Application of data analytics for product design: Sentiment analysis of online product reviews", CIRP Journal of Manufacturing Science and Technology 23 (2018) pp. 128-144.

[8] Reinald Kim Amplayo, Seanie Lee, Min Song, "Incorporating product description to sentiment topic models for improved aspect-based

sentiment analysis", Information Sciences 454-455 (2018) pp. 200-215.

[9] Xiaolin Lia, Chaojiang Wub, Feng Maic, "The effect of online reviews on product sales: A joint sentiment-topic analysis", Information & Management 56 (2019) pp. 172-184.

[10] Yang Liu, Jian-Wu Bi, Zhi-Ping Fan, "Ranking products through online reviews: A method based on sentiment analysis technique and intuitionistic fuzzy set theory", Information Fusion 36 (2017) pp. 149-161.

[11] Jayakumar Sadhasivam, Ramesh Babu Kalivaradhan, "Sentiment Analysis of Amazon Products Using Ensemble Machine Learning Algorithm", International Journal of Mathematical, Engineering and Management Sciences, Vol. 4, No. 2, 508–520, 2019.

[12] Anu Dahiya, Nidhi Gautam, Prashant Kumar Gautam, "Data Mining Methods and Techniques for Online Customer Review Analysis: A Literature Review", Journal of System and Management Sciences, Vol. 11 (2021) No. 3, pp. 1-26.

[13] Hassan Raza, M. Faizan, Ahsan Hamza, Ahmed Mushtaq, Naeem Akhtar, "Scientific Text Sentiment Analysis using Machine Learning Techniques", (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 10, No. 12, 2019.

[14] Qurat Tul Ain, Mubashir Ali, Amna Riazy, Amna Noureenz, Muhammad Kamranz, Babar Hayat, A. Rehman, "Sentiment Analysis Using Deep Learning Techniques: A Review", International Journal of Advanced Computer Science and Applications, Vol. 8, No. 6, 2017.

[15] Supriya B, C.B. Akki, "Sentiment Prediction using Enhanced XGBoost and Tailored Random Forest", International Journal of Computing and Digital Systems, ISSN (2210-142X).

[16] Maganti Syamala1, Nattanmai Jeganathan Nalini, "A Filter Based Improved Decision Tree Sentiment Classification Model for Real-Time Amazon Product Review Data", International Journal of Intelligent Engineering and Systems.

[17] Li Chen, Dongning Yan, Feng Wang "User perception of sentiment-integrated critiquing in recommender systems", International Journal of Human-Computer Studies.

[18] Efstratios Kontopoulos, Christos Berberidis, Theologos Dergiades, Nick Bassiliades "Ontology-based sentiment analysis of Twitter posts", Expert Systems with Applications.

[19] A.Nisha Jebaseeli, E.Kirubakaran, “A Survey on Sentiment Analysis of(product) Reviews”, International Journal of Computer Applications(0975-888), volume 47- no.11

[20] Gayatri Khanvilkar, Deepali Vora, “ Product Recommendation using Sentiment Analysis of reviews: A Random Forest Approach”, International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249-8958, Volume-8, Issue-2S2.

[21] Dr. D. Sivaganesan, Sridhar K, Dhruv Aggarwal, Arunkumar M, “ Feature based Sentiment Analysis for Product Reviews”, International Journal of Engineering Research & Technology, ISSN: 2278-0181, Volume 11, Issue 06.

[22] Adinarayana Salina, K Yogeswara Rao Yogi Rao, G.S.N Murthy, “Product Recommendation System from Users Reviews using Sentiment Analysis”, International Journal of Computer Applications, Volume-169.

[23] Najma Sultana, Pintu kumar, Sourabh Chandra, SK Safikul Alam, “Sentiment Analysis for Product Review”, International Journal of Soft Computing, pg:1913-1919.

[24] Dishi Jain, Bitra Harsha Vardhan, Sravanakumar Kandasamy, “Sentiment Analysis Of Product Review- A Survey”, International Journal of Scientific & Technology Research, Volume 8, Issue 12.

[25] Raktim Kumar Dey, Debabrata Sarddhar, Rajesh Bose, Sandip Roy, “ A Literature Survey on Sentiment Analysis Techniques involving Social Media and Online Platforms”, International Journal of Scientific & Technology Research, Vol 1 Issue 1.

[26] Rajkumar S.Jagdale, Vishl S. Shirsat, Sachin N.Deshmukh, “Sentiment Analysis on Product Reviews Using Machine Learning Techniques”, Cognitive Informatics and Soft Computing, Adavnces in Intelligent Systems and Computing 768, pg:639-647

[27] B. Bhavana, Jayavarapu Karthik, P. Lakshmi Kumari. “A Novel Approach for Product Recommendation using XGBOOST”, 2023

International Conference on Intelligent Data Communication Technologies and Internet of Things(IDCIoT), 2023.

[28] Yadav, Vijay, and Subarna Shakya. "Sentiment Analysis and Topic Modeling on News Headlines. "Journal of Ubiquitous Computing and Communication Technologies 4, no. 3 (2022): 204-218.