Journal of Computing & Biomedical Informatics
ISSN: 2710 - 1606

Volume 05 Issue 01
2023

Research Article
https://doi.org/10.56979/501/2023

AI Credit: Machine Learning Based Credit Score Analysis

Suleman Qamar[1,*], Muhammad Hanif Durad[1], Faizan Ul Islam[2], Syeda Rozeena Saleha[3], Muhammad Hamza[2], Allay Hyder Urooj[4], Syed Muhammad Abrar Akber[5]

[1]Critical Infrastructure Protection and Malware Analysis (CIPMA) Lab, Department of Computer and Information Science, Pakistan Institute of Engineering and Applied Sciences, Islamabad 45650, Pakistan.
[2]Department of Computer and Information Science, Pakistan Institute of Engineering and Applied Sciences, Islamabad 45650, Pakistan.
[3]Kalam4Solutions, Islamabad, Pakistan.
[4]Directorate of Outreach, University of Agriculture, Faisalabad 38000, Pakistan.
[5]Department of Computer Graphics, Vision and Digital Systems, Faculty of Automatic Control, Electronics and Computer Science, Silesian University of Technology, Gliwice, 44-100, Poland.
[*]Corresponding Authors: Suleman Qamar. Email: m.sulemanqamar@gmail.com.

Received: March 25, 2023 **Accepted:** May 29, 2023 **Published:** June 05, 2023.

Abstract: Statistical credit rating indicates whether a person is able to pay his debts or not. In this study, the effectiveness of different machine learning (ML) classifiers for credit scoring and bankruptcy prediction is investigated and compared. The main advantage of this system is to reduce the risk of financial crisis and bankruptcy. This will help applicants who wish to find out if they are qualified for credit and how much they are capable of receiving as well as for creditors taking the risk factor into consideration. In the study, credit applicants were divided into two categories: those who paid their bills on time and those who didn't. Using this information, people were classified into those who were eligible for credit and those who weren't eligible for credit. The data set used for credit scores was "GIVE ME SOME CREDIT" from Kaggle. Flask was used to build and deploy the final model. Several machine learning classifiers were used, including bagging, logistic regression, gradient classifier, and random forest classifier. XGboost performed the best among all, achieving 94% accuracy, ROC of 86%, and 92% F1 score.

Keywords: Credit score; Bagging and Boosting; Machine learning; Bankruptcy

1. Introduction

Financial companies and lenders use credit scoring [1] to determine a person's creditworthiness. Lenders rely on credit scores to decide whether or not to extend credit to a person. The possible range of credit ratings is 300 to 850, with 850 being the highest number. The capacity of a person to access financial products like credit cards, mortgages, auto loans, and personal loans depends on their credit score. Nowadays, even the way of life has revolutionized. Everything has become so fast that no one wants to travel a long way to complete their tasks. Every person thinks that all their solution to their problem was right at their fingertips.

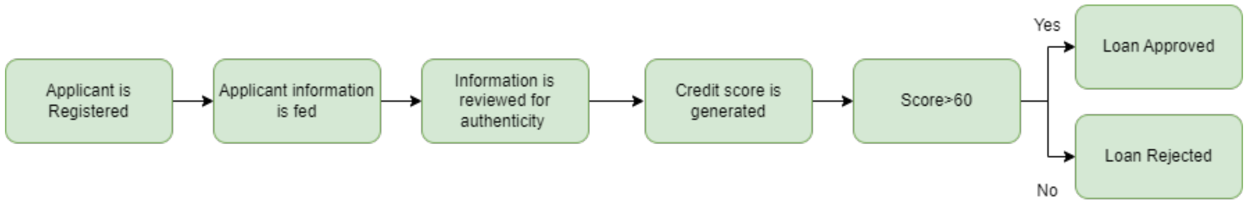

Figure 1. Credit scoring Workflow

In Pakistan, one has to fill out a form, have the attached documents notarized by reference persons, and meet other requirements, and then it is sent to the National Bank of Pakistan for processing. Even then, one has to wait for about one to two months to get an answer on eligibility. Pakistan has been facing a lot of problems related to the financial system and especially credit. One of the problems was that applying for a loan is a long and hectic process with limited transparency [2].

Moreover, the decision about creditworthiness depends on experts who can consider only a limited number of variables. Decisions about lending and pricing, new customer sales, customer retention, cash flow, and revenue forecasting all rely on credit scoring algorithms [3].Credit cards, loans, and other credit-based products and services are all impacted by your credit rating [4]. Understanding how credit scoring works can help you improve your credit practices to maintain a good credit score. Because they are similar concepts, credit ratings and credit scoring should not be confused. Corporations, sovereigns, sub-sovereigns and the securities they issue, as well as asset-backed securities, are all assigned credit ratings.

Credit scoring models build a picture of a person's credit history, and scores from the three major credit agencies may differ (although not significantly). The interest rate on a loan and whether or not a borrower's loan or note is accepted both depend on creditworthiness. Generally, credit scoring algorithms take into account how many times one has missed payments recently, how much they owed, and how long they were past due. Credit history provides information about how many of a person's credit accounts were past due based on all of their accounts. The credit score can be negatively impacted if a user has 10 credit accounts and five of them are delinquent. Credit scoring workflow is shown in Figure 1.

Payment history includes bankruptcies, foreclosures, income garnishments, and accounts reported to collection agencies, and can have a significant impact on some credit scores because credit scoring models take all of this information into account. For decades, consumer lending relied heavily on character assessments, and the credit industry has been slow to adapt to a new, standardized system of credit scoring. Only one company, American Investments, used Fair Isaac's statistical scorecard [5] when it was introduced in 1958. Credit card companies, auto lenders, and banks were among the first to adopt the system after its introduction in the late 1950s. They relied on the Fair Isaac system to quickly and accurately determine whether or not a borrower was creditworthy. By the late 1970s, the majority of lenders were using credit scoring. Today, it is estimated that 26 million Americans have no credit history with any of the major credit bureaus people who have few or no credit accounts on file and no recent financial activity are considered "credit unworthy". This work contributes in the following ways:

- Kaggle dataset[6] named “GIVE ME SOME CREDIT” is preprocessed extensively to make it more suitable for machine learning;
- Multiple machine learning approaches are applied and contrasted identifying and separating credit scorable people from others;
- Model is deployed using flask for ease of use;
- This major significance of this is preprocessing the dataset “GIVE ME SOME CREDIT” and applying different machine learning techniques on it in order to address the problem of credit scoring for banks to decide whether or not to grant credit to a particular person, and for general users to check if they are eligible for a certain amount of credit.

The remainder of the paper is organized as follows. The next section presents related work in the field of credit scoring. Section 3 explains the dataset and machine learning models, while Section 4 discusses the materials and methodology. Section 5 presents the results and discussions and draws a comparison with other studies. Section 6 presents and discusses the deployment of the proposed XGBoost credit scoring model. Finally, Section 7 concludes this study.

2. Literature review

Credit bureaus emerged in the 1800s. These tools were used by businesses to determine who was creditworthy. Urbanization in the United States made it more difficult to get credit. There was no standard way to ensure that a borrower could repay their loans if they moved and needed a bank loan. Because of the increased demand, businesses cared less and less about the needs of their customers. Businesses needed help figuring out who they could trust when it came to cash flow. The first consumer credit agency in the United States was created by businesses. Credit bureau representatives collected personal and credit information on tenants and employees (e.g., debt, moral character, marital status, etc.). Courthouses were also visited, where newspaper clippings and other official documents were obtained for research purposes.

In most cases, the information contained in these files was of high quality. Therefore, in the 1930s, department stores began to use quantitative methods and assign credit scores to customers based on this information. However, this approach to calculating credit scores was not based on research. People were given points based on their race, income, where they lived, and where they worked to assess their moral character. A statistical method is better at predicting the likelihood that a borrower will repay his or her debt than one based on the borrower's personal characteristics. An unbiased credit scoring system was one of the goals of the company, which was founded in 1956 by engineers Bill Fair and Earl Isaac. Standardized credit evaluation and lending methods would, in theory, eliminate the biases that had existed for decades. FICO [7] is the new name for Fair, Isaac and Company.

Numerous statistical and machine learning techniques were used to implement the credit scoring algorithm. In [8], seven well-known financial datasets from different countries were used, namely Australia, Germany, Japan, Taiwan, Poland, and India, in addition to "Give Me Some Credit" from Kaggle [9]. While the other datasets were used to study credit scoring, the Polish and Indian datasets were used to predict bankruptcies. Several traditional machine learning classifiers were used to achieve high accuracy. The statistical analysis demonstrates that the Bayes Net [10], Random Forest [11], AdaBoost [12], and LogitBoost [13] machine learning classifiers offer effective credit scoring models. In the end, they came to the conclusion that Bayes Net and boosting classifiers are efficient general-purpose models for credit scoring. According to Sunil Bhatia et al's study, the accuracy of the models investigated for credit scoring and bankruptcy prediction is shown by the numerical results, which are equivalent to those of earlier studies. "Credit Scoring Using Machine Learning Methods" [14], ensemble learning has a greater level of categorization and prediction accuracy. Therefore, it is widely used for the evaluation of personal loans. The machine learning methods used in this study are given in the following:

- Linear Discriminant Analysis (LDA) [15];
- Random Forest;
- Logistic Regression [16];
- XGBoost [17].

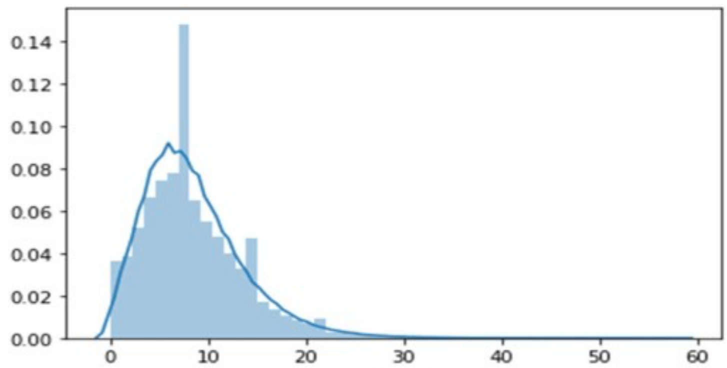

Figure 2. Number of credit lines and loans

While examining the performance metrics, it was discovered that XGBoost and Random Forest produced the greatest outcomes. The study by Fan et al. [18] titled "Improved ML-Based Method for Credit Card Scoring in Online Financial Risk Control" uses the Logistic Regression (LR) and XGBoost models for credit scoring. Due to its ability to translate, LR is most frequently employed for credit rating. The system was assessed using techniques like accuracy rate, true positive rate, false positive rate, accuracy rate, F1 score, and confusion matrix. They decided to use a large peer-to-peer (P2P) online lending platform in China as a research case study. Borrowers are classified into numerous groups based on their different performance levels. Borrowers were categorized as D0 and D1 ("good borrowers") and 6045 borrowers were categorized as D2 ("poor borrowers").

Six different factors, including education level, income, age, gender, and credit card purchasing history, are included in the sample data. Indicators like "KS," "AUC," "GINI," and others are crucial in assessing credit scoring algorithms. The GINI coefficient is used to assess the model's capacity to discriminate risk, while AUC serves as the standard for weighing the advantages and disadvantages of classifiers. The biggest difference between the cumulative percentage of undesirable and suitable consumers is used to determine the GINI coefficient, which assesses the model's capacity to distinguish between different types of clients. The capacity of the version to differentiate between different types of risk is considered satisfactory if the GINI score is close to 0.5. If the model threshold is properly established and the AUC value is in the range of 0.7 to 0.8, the classifier is superior to random estimation and there is a safe predictive value. The GINI value shows that the model's capacity for risk discrimination is appropriate when the AUC value is between 0.7 and 0.8. Data cleaning and feature selection, processing the unbalanced data set, setting the ML algorithm, and analyzing the output of the suggested ML model were the four techniques used.

3. Datasets and machine learning classifiers

For this study, we used Kaggle's dataset "GIVE ME SOME CREDIT", the largest credit scoring dataset available on the Internet. The number of credit lines and loans is shown in Figure 2. In this study, six different datasets related to credit scoring were considered, namely:

1. Australian;
2. German;
3. Taiwan;
4. Japanese;
5. Give me some credit.

As machine learning algorithms generally perform better when there more samples available of the data and in this case "GIVE ME SOME CREDIT" had the most samples so it chosen as the main dataset so that better credit scoring models could be generated. Figure 3 shows age of the customer and Figure 4 represents different features' correlation with monthly income.

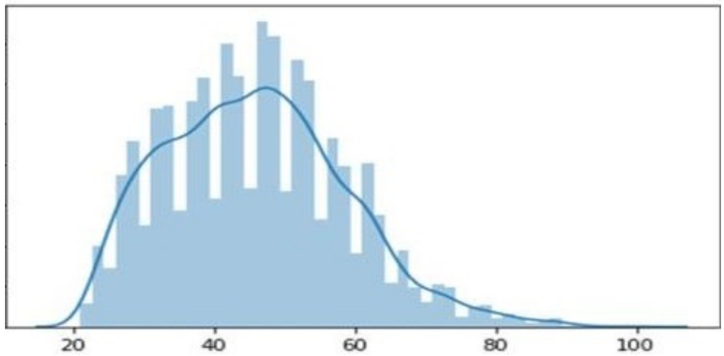

Figure 3. Age of the customer

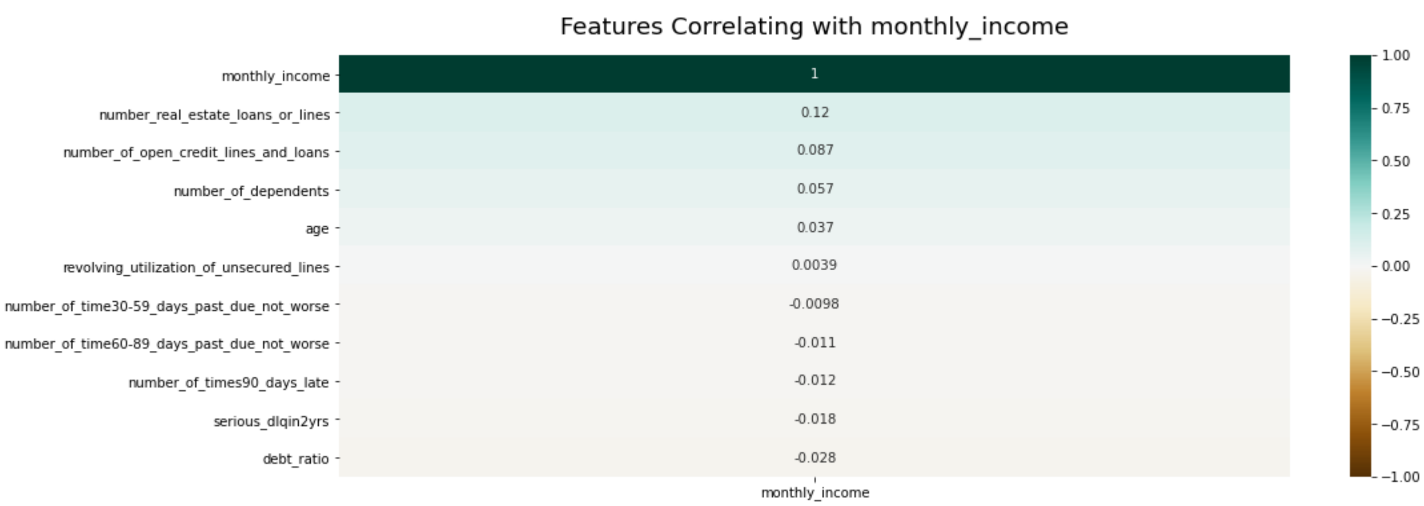

Figure 4. All features correlation with monthly income

Furthermore, other datasets were either highly biased due to regional samples or had too few values which are not ideal for machine learning algorithms. Also, these datasets have a lot of Null values which render some of the important learning features useless.

Different machine learning classifiers are employed to achieve state of the art performance. An ensemble contains multiple base learners and combines those weak learners into a strong classifier. Following classifiers were employed for learning:

1. Gradient boosting classifier;
2. Logistic regression classifier;
3. Bagging;
4. Random forest classifier;
5. Boosting.

3.1. Gradient Boosting Classifier

Gradient-boosting classifiers combine a set of learners that are less qualified to create a single learner that is better qualified. The decision trees themselves are not very effective learners. Each tree is sequentially related to its predecessor, and each tree strives to reduce the error rate of its predecessor. On the other hand, boosting algorithms are often slow to train due to their sequential linkage, but have high accuracy. It has been found that models that are slower to learn perform better in statistical learning.

3.2. Logistic Regression Classifier

To forecast the likelihood of an outcome variable, a supervised learning classification technique like logistic regression is used. There are only two possible categories to take into consideration because the goal or dependent variable is a dichotomous variable. Or, to put it another way, the data for the target variable can be coded as either binary 1 (which stands for successful/yes) or binary 0 (which stands for unsuccessful/no).

3.3. Bagging

Machine learning algorithms can be made more accurate and efficient by using a technique known as bagging, sometimes known as bootstrap aggregation. It manages bias-variance tradeoffs and reduces the variance of a prediction model. To avoid overfitting the data, bagging, particularly decision tree approaches, is utilized for both regression and classification models [19].

3.4. Random Forest Classifier

Using machine learning, a random forest classifier can be used to solve classification and regression issues. It employs ensemble learning, a method for resolving complex issues by merging various classifiers. Many decision trees make up a random forest algorithm. To train the "forest" created by the random forest approach, utilize bagging or bootstrap aggregation. By grouping algorithms together, the meta-algorithm of bagging improves algorithm accuracy. The shortcomings of a decision tree classifier are overcome by a random forest approach. Accuracy is increased and overfitting of the data set is decreased.

3.5. Boosting

The term "boosting" describes a group of algorithms that transform weak learners into strong learners. Boosting is an acronym for "Boosting Family of Algorithms" Boosting [19] is an ensemble strategy that can be used to improve the predictive performance of any learning algorithm. Boosting is a method in which weak learners are trained incrementally, with each stage attempting to improve the previous one.

4. Materials and methodology

The first step in the proposed technique is cleaning of data to make it more viable for learning better patterns and features. To do so, an exploratory analysis was carried out with the following results:

1. Around 6% of total samples were defaulted. In the dataset, when column “NumberOfTimes90DaysLate” has values above 17, furthermore there were 267 instances in the dataset where three columns share same values, namely “NumberOfTimes90DaysLate”, “NumberOfTimes60-89DaysPastDueNotWorse”, “NumberOfTime30-59DaysPastDueNotWorse.
2. Many values of monthly income and dependents were empty.
3. The dataset was imbalanced as number of good loans was too much and number of bad loans was too low.

According to the analysis, the available data was not in correct form to be used to train machine learning models. Thus, some preprocessing operations on the dataset were required to convert it into a more useful form. In order to do so, following techniques were performed to handle missing values and eliminate the features which were not important.

Feature importance is calculated using MDI (Mean Decrease in Impurity) also called GINI Importance represented by Figure 5 while Figure 6 shows correlation Heat-map with 1 representing very highly correlated.

4.1. Handling NULL Values

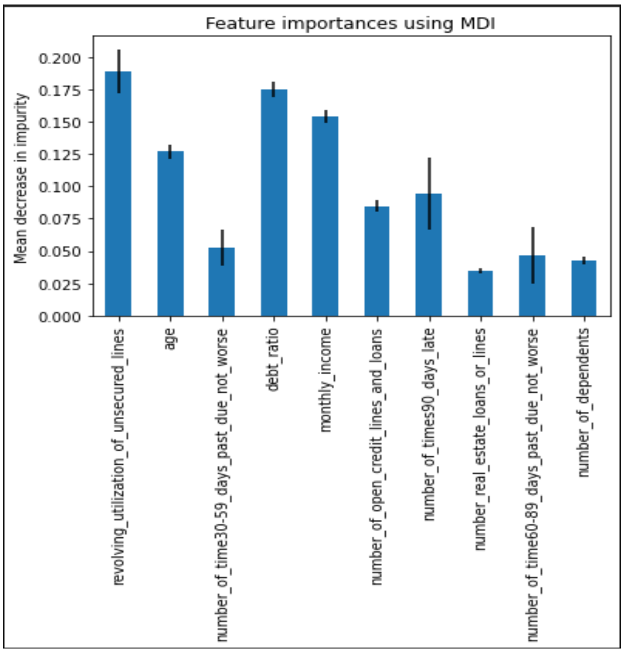

Figure 5. Important features found using MDI (Mean Decrease in Impurity)

At first, the null values had to be removed. Only "Monthly Income" and "number of dependents" had null values. Among those "number of dependents" was not a key value/attribute so it can be made null and it would not affect the prediction capability the model but the monthly income could not be left null. As it has a 20% part consisting of null values so mean or mode can't be employed to fill these null values as it will result in having a more skewed distribution instead of a normal distribution which may result in less accurate models. In order to deal with this, the null values had to be replaced with intelligent values. To do so, an imputer was created to impute intelligent values to fill null values. In the next step, Dataset was split into 2 groups: training and test. Then the training set was used to calibrate the model. After calibration, test set was employed to evaluate the effectiveness of the model. Train data was split in two columns, one containing data along with null values and the other not containing any null values. Then the correlation of monthly income with other features was computed. 'Number_real_estate_loans_or_lines', 'number_of_open_credit_lines_and_loans' were taken as features while monthly income as target variable to train the model. The reason for taking these two columns was of them having greater effect on monthly income.

These two columns were taken because they have a greater effect on monthly income. K Neighbor Regressor was employed to train the model with a monthly income that does not have null values and predict the instances that contain null values. In this way, intelligent values for monthly income were incorporated.

4.2. Balancing Dataset

Unbalanced data is a major issue during training machine learning algorithms [20], in this case, dataset had over 90 percent of the loan samples being positive and fewer than 10 percent samples being negative. Therefore, dataset must be balanced in order to create a reliable model. For this purpose, synthetic minority oversampling technique (SMOTE) [21] was employed. It is a statistical method for proportionally increasing the number of cases in a dataset. The component generates new instances from the minority class that is provided as input. This SMOTE implementation has no effect on the number of majority cases. The new occurrences are not just repeats of existing minority situations. Instead, the method gathers feature space samples for each target class and its close neighbors. After that, the computer creates new instances with traits from both the target case and its neighbors. This tactic broadens the capabilities that are available to each class and makes the examples more applicable. SMOTE only raises the proportion of minority cases while accepting the entire dataset as input. As a result, 100,000 samples of "bad loans" were added.

4.3. Feature Reduction

Feature reduction, also known as dimensionality reduction, is a technique for reducing features in calculations using a lot of resources without sacrificing crucial information [22]. Because there are fewer qualities, there are fewer variables as well, which makes the computer simpler and faster. The amount of resources needed to complete a calculation or job decreases as functionality is reduced. The computer can complete more tasks in a shorter amount of time and with less storage space. Feature reduction improves machine learning models by eliminating multi co-linearity during machine learning. Another advantage of dimensionality reduction is that it helps facilitate data visualization for humans, especially when data is reduced to 2d or 3d that may be displayed graphically.

The curse of dimensionality [23] is an interesting problem that may be addressed by reduction. Reducing the number of dimensions in data can make it easier for machine learning algorithms to work with. This can make the data more manageable and statistically significant. Firstly, we checked the correlation of variables with each other. It can be seen that some of the independent variables have a very high correlation with each other. There is a strong association between "Number of times 30-59 days past due not worse," "Number of times 60-89 days past due not worse," and "Number of times 90 days late." Hence, eliminating the phrases "Number of times, 30-59 days past due not worse" and "Number of times, 60-89 days past due not worse" will not have an impact on the suggested model.

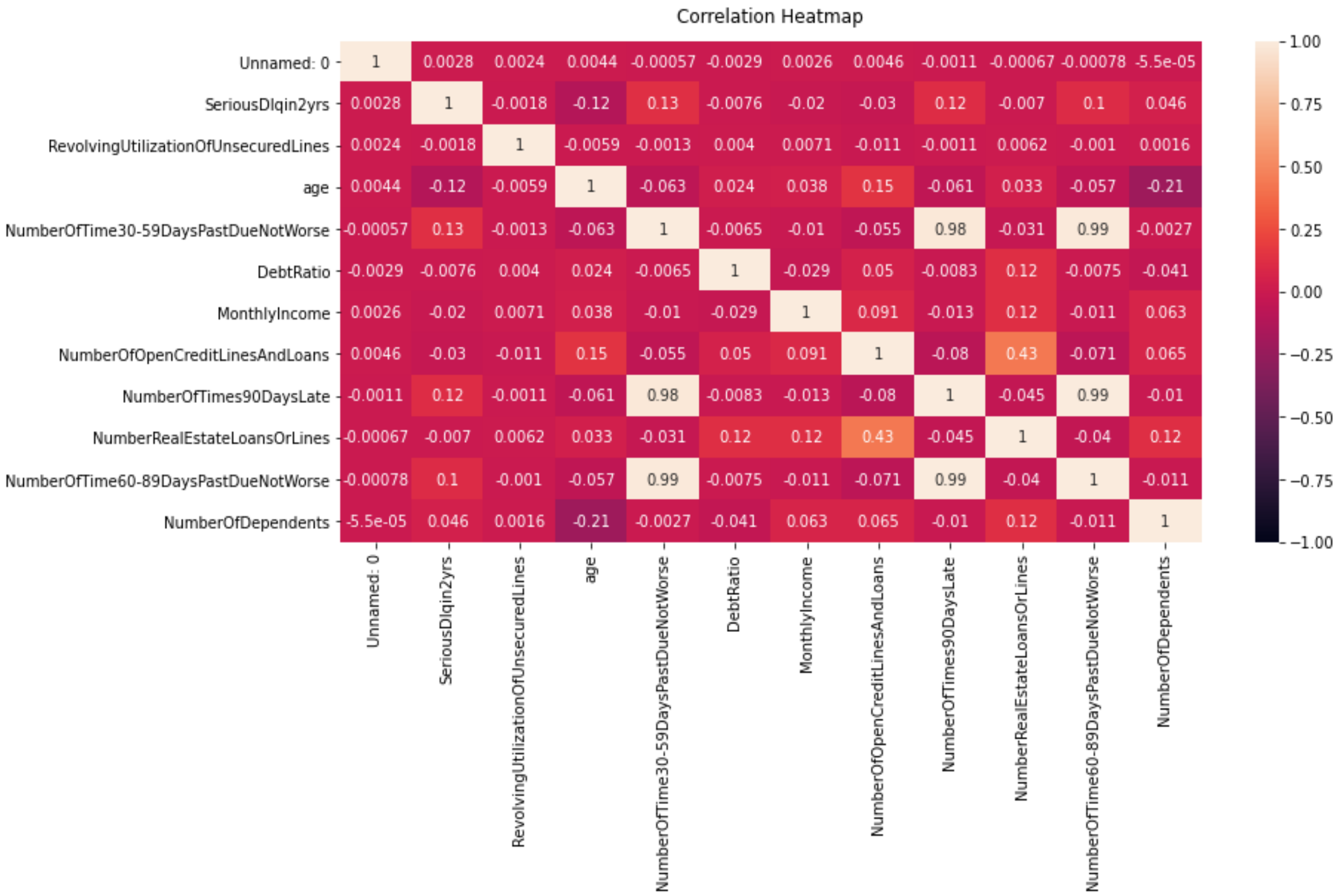

Figure 6. Correlation of features with each other

5. Results and discussions

To evaluate the performance of the proposed technique, following performance metrics [24] were employed.

1. Accuracy;
2. Precision;
3. Recall;

4. F1-score;
5. ROC (%).

After preprocessing the dataset by using above mentioned methods, different algorithms were applied on them. Resultant ROC curves are given in Figure 7 to 11 for all the trained models. XGBoost performed the best in all metrics having F1 score of 92 and accuracy of 94%, as shown table 1.

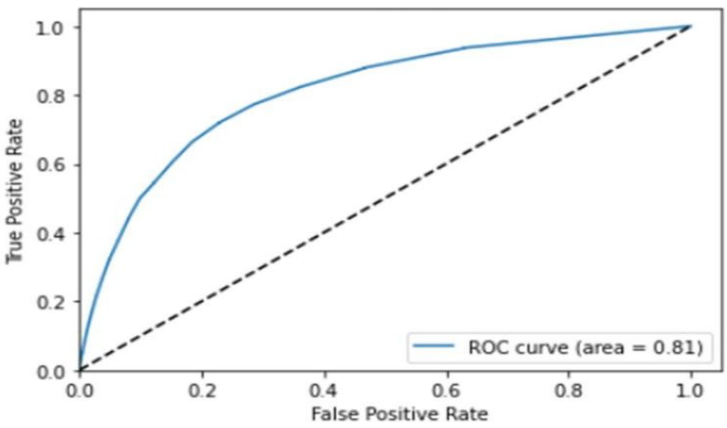

Figure 7. Bagging

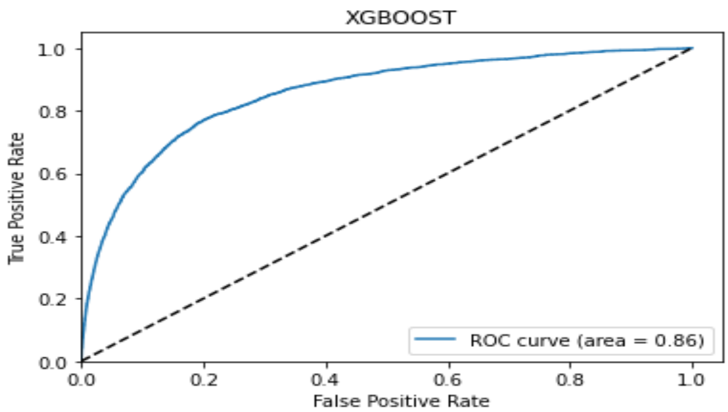

Figure 8. Boosting achieving 86% Roc and F1 Score of 92%

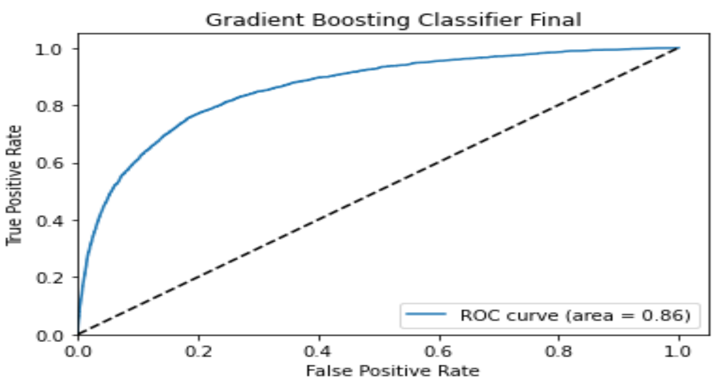

Figure 9. Gradient Boosting

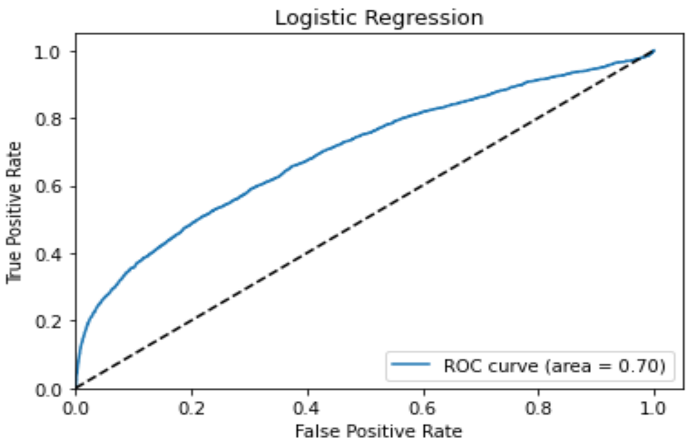

Figure 10. Logistic Regression

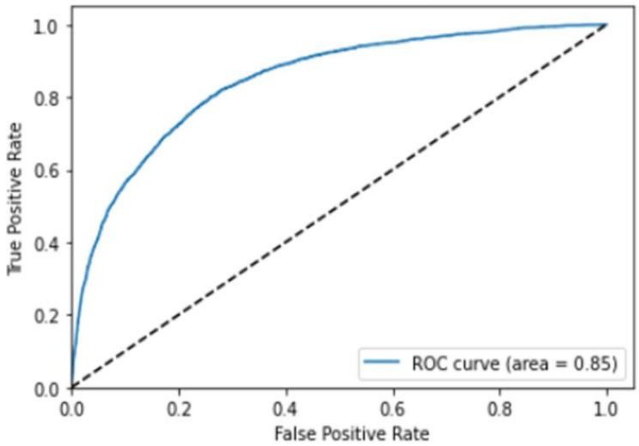

Figure 11. Random Forrest

Table 1. Comparative analysis of different classifiers

Classifier(s)	Precision	Recall	Accuracy	F1 score	ROC (%)
Bagging	91	93	93	91	81
Gradient Boosting	91	93	93	92	86
Logistic Regression	91	93	89	90	70
Random Forest	92	94	91	91	85
XGBoost	**92**	**94**	**94**	**92**	**86**

Our approach of preprocessing dataset and then applying XGBoost, performed better than the existing techniques achieving an accuracy of 94% and ROC value of 86% as shown in table 2. These can be further improved using meta-classifiers.

Table 2. Comparison with existing research

Technique(s)	ROC (%)
MLP[8]	81.4
Bayes Net[8]	85.5
Proposed approach (XGBoost)	86

6. Deployment of the proposed XGBoost credit scoring model

Machine learning model is deployed in flask [25] shown in figures 12 and 13, which was connected to a MySQL database as seen in figures 14 and 15, to give it a better appearance and allow users to use it in real time. It stored the values entered by users and printed the result and those values in real time.

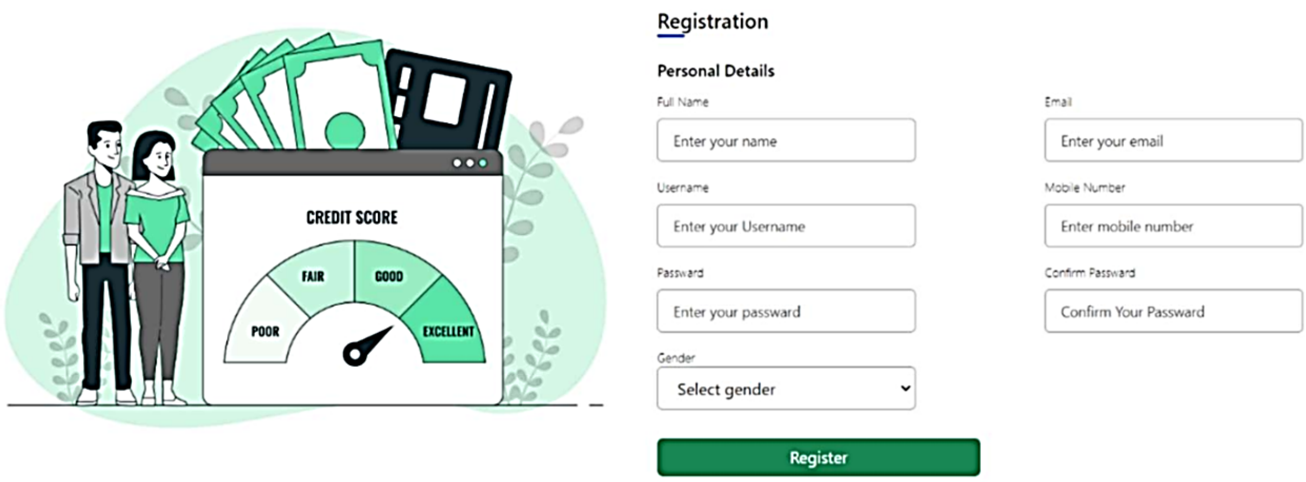

Figure 12. Sign Up Interface

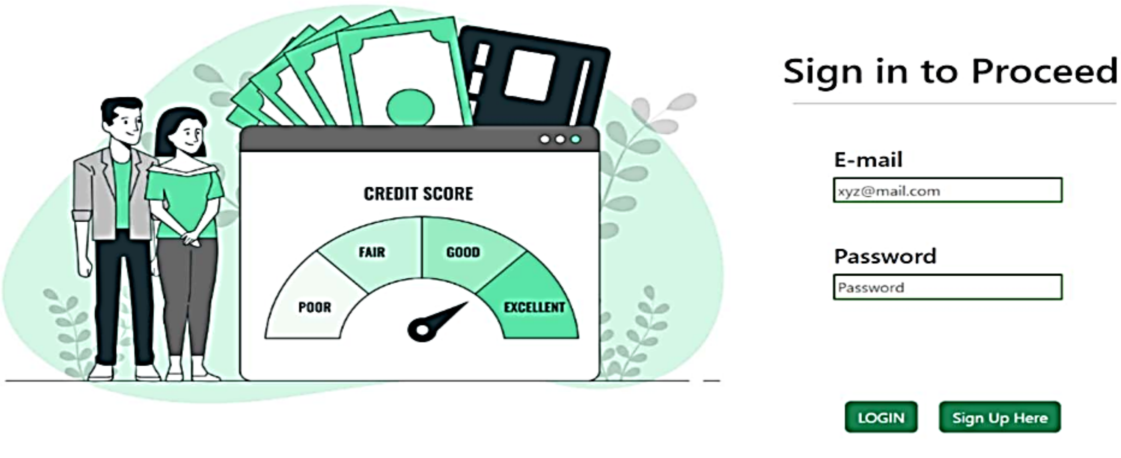

Figure 13. Login Interface

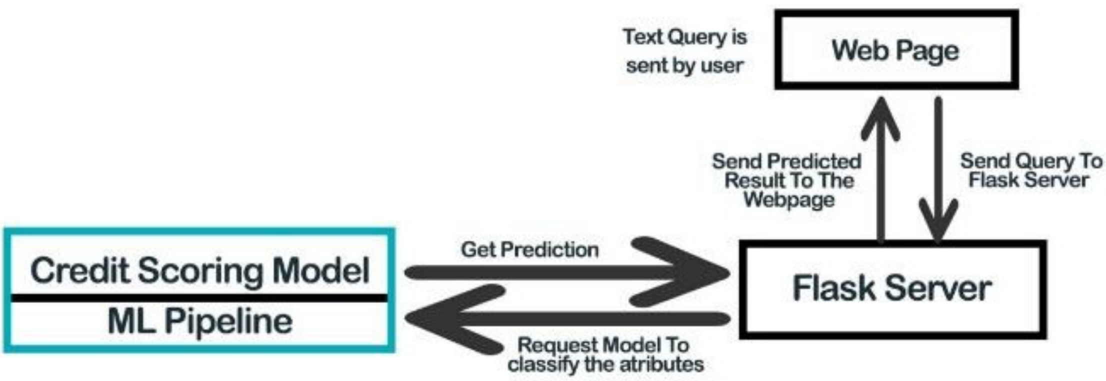

Figure 14. All components of the whole system and their relations

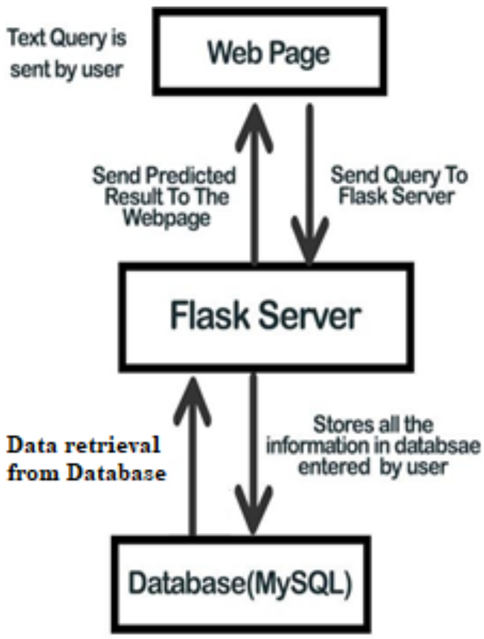

Figure 15. Interaction with Flask Server and database

After logging in, the user is prompted to enter information into the credit score checker, shown in Figure 16. This is used to determine the user's credit score. The user can find out whether they meet the requirements for credit or not. The authenticity of the information needs to be verified by a bank officer when the bank uses it to make a decision about granting credit. Figure 17 shows the final prediction screen.

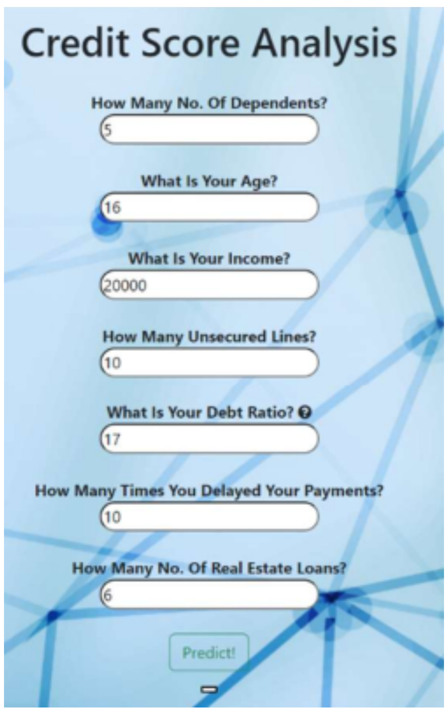

Figure 16. Graphical User Interface of the Credit Score Predictor

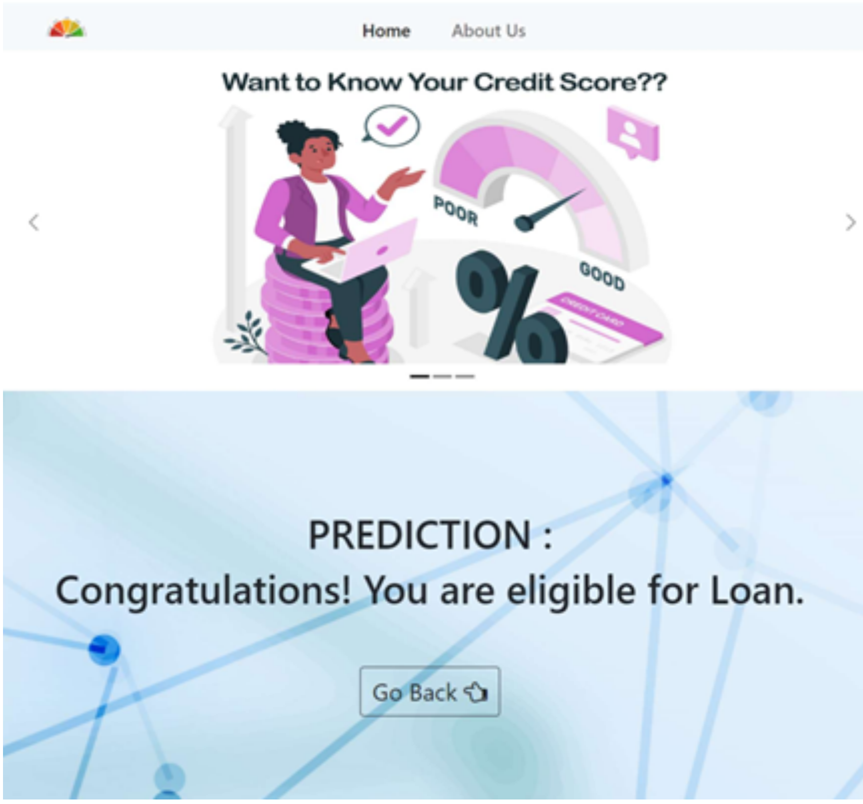

Figure 17. Prediction screen of the credit score predictor

7. Conclusion and future work

This research addresses the problem of credit scoring for banks to decide whether or not to grant credit to a particular person, and for general users to check if they are eligible for a certain amount of credit. Calculating whether or not to grant credit is not only a complex task, but also a tedious one. Therefore, by preprocessing the dataset and using machine learning, significant results could be achieved. Our proposed approach achieved 94% accuracy, 92% precision, 94% recall, 86% ROC score, and 92% F1 score. Flask was used to create and deploy the final model. Other credit scoring datasets such as the Pakistani dataset can be preprocessed using similar techniques, and then machine learning models can be used to create models for region-specific credit scoring. All datasets can be combined and refined to create a global model, and deep learning approaches can be applied in subsequent work to achieve even better performance.

References

1. Ziemba, P., Radomska-Zalas, A., & Becker, J. (2020). Client evaluation decision models in the credit scoring tasks. Procedia Computer Science, 176, 3301-3309.
2. Hussain, A., Khan, M., Rehman, S. U., & Khattak, A. (2019). Credit scoring model for retail banking sector in Pakistan. Journal of Managerial Sciences, 14(4), 153-161.
3. Lessmann, S., Baesens, B., Seow, H. V., & Thomas, L. C. (2015). Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. European Journal of Operational Research, 247(1), 124-136.
4. Ali, S., Yousaf, I., & Naveed, M. (2020). Role of credit rating in determining capital structure: Evidence from non-financial sector of Pakistan. Studies of Applied Economics, 38(3).
5. Arya, S., Eckel, C., & Wichman, C. (2013). Anatomy of the credit score. Journal of Economic Behavior & Organization, 95, 175-185.
6. Berry, J. A., Marotta, D. A., Savla, P., Tayag, E. C., Farr, S., Javaid, R., ... & Miulli, D. E. (2021). Predictive value of credit score on surgery resident and fellow academic and professional performance. Cureus, 13(6).
7. Liu, S., Wang, R., & Han, Y. (2021, June). Research on Personal Credit Evaluation Based on Machine Learning Algorithm. In 2021 6th International Symposium on Computer and Information Processing Technology (ISCIPT) (pp. 48-52). IEEE.
8. Boughaci, D., & Alkhawaldeh, A. A. (2020). Appropriate machine learning techniques for credit scoring and bankruptcy prediction in banking and finance: A comparative study. Risk and Decision Analysis, 8(1-2), 15-24.
9. Kaggle "Give me some credit" [Online] accessed 06/06/2022 https://www.kaggle.com/c/GiveMeSomeCredit
10. Ben, G. I. (2007). Bayesian networks. encyclopedia of statistics in quality and reliability.
11. Leo, B. (2001). Random forests. Machine learning, 45(1), 5-32.
12. Schapire, R. E. (2013). Explaining adaboost. Empirical Inference: Festschrift in Honor of Vladimir N. Vapnik.
13. Kamarudin, M. H., Maple, C., Watson, T., & Safa, N. S. (2017). A logitboost-based algorithm for detecting known and unknown web attacks. IEEE Access, 5, 26190-26200.
14. Bhatia, S., Sharma, P., Burman, R., Hazari, S., & Hande, R. (2017). Credit scoring using machine learning techniques. International Journal of Computer Applications, 161(11), 1-4.
15. Xanthopoulos, P., Pardalos, P. M., Trafalis, T. B., Xanthopoulos, P., Pardalos, P. M., & Trafalis, T. B. (2013). Linear discriminant analysis. Robust data mining, 27-33.
16. LaValley, M. P. (2008). Logistic regression. Circulation, 117(18), 2395-2399.
17. Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., ... & Zhou, T. (2015). Xgboost: extreme gradient boosting. R package version 0.4-2, 1(4), 1-4.
18. Fan, S., Shen, Y., & Peng, S. (2020). Improved ML-based technique for credit card scoring in internet financial risk control. Complexity, 2020, 1-14.
19. Aldrees, A., Awan, H. H., Javed, M. F., & Mohamed, A. M. (2022). Prediction of water quality indexes with ensemble learners: Bagging and Boosting. Process Safety and Environmental Protection, 168, 344-361.
20. Mooijman, P., Catal, C., Tekinerdogan, B., Lommen, A., & Blokland, M. (2023). The effects of data balancing approaches: A case study. Applied Soft Computing, 132, 109853.
21. Viloria, A., Lezama, O. B. P., & Mercado-Caruzo, N. (2020). Unbalanced data processing using oversampling: Machine Learning. Procedia Computer Science, 175, 108-113.
22. Nturambirwe, J. F. I., Hussein, E. A., Vaccari, M., Thron, C., Perold, W. J., & Opara, U. L. (2023). Feature Reduction for the Classification of Bruise Damage to Apple Fruit Using a Contactless FT-NIR Spectroscopy with Machine Learning. Foods, 12(1), 210.
23. Kabir, M. F., Chen, T., & Ludwig, S. A. (2023). A performance analysis of dimensionality reduction algorithms in machine learning models for cancer prediction. Healthcare Analytics, 3, 100125.
24. Wignall, T. J., Nakamura-Zimmerer, T., Scoggins, J. B., Snyder, S., Kumar, S., & Colbert, B. K. (2023). Comparisons of performance metrics and machine learning methods on an entry descent and landing database. In AIAA SCITECH 2023 Forum (p. 1183).
25. Mufid, M. R., Basofi, A., Al Rasyid, M. U. H., & Rochimansyah, I. F. (2019, September). Design an mvc model using python for flask framework development. In 2019 International Electronics Symposium (IES) (pp. 214-219). IEEE.