Eligible Personal Loan Applicant Selection using Federated Machine Learning Algorithm

Mehrin Anannya, Most. Shahera Khatun, Md. Biplob Hosen, Sabbir Ahmed,
Md. Farhad Hossain, M. Shamim Kaiser
Institute of Information Technology, Jahangirnagar University, Bangladesh

***Abstract*—Loan sanctioning develops a paramount financial dependency amongst banks and customers. Banks assess bundles of documents from individuals or business entities seeking loans depending on different loan types since only reliable candidates are chosen for the loan. This reliability materializes after assessing the previous transaction history, financial stability, and other diverse kinds of criteria to justify the reliance of the bank on an applicant. To reduce the workload of this laborious assessment, in this research, a machine learning (ML) based web application has been initiated to predict eligible candidates considering multiple criteria that banks generally use in their calculation, in short which can be briefed as loan eligibility prediction. Data from prior customers, who are authorized for loans based on a set of criteria, are used in this research. As ML techniques, Random Forest, K-Nearest Neighbour, Adaboost, Extreme Gradient Boost Classifier, and Artificial Neural Network algorithms are utilized for training and testing the dataset. A federated learning approach is employed to ensure the privacy of loan applicants. Performance analysis reveals that Random Forest classifier has provided the best output with an accuracy of 91%. Based on the mentioned prediction, the web application can decide whether the customers' requested loan should be accepted or rejected. The application was developed using NodeJs, ReactJS, Rest API, HTML, and CSS. Furthermore, parameter tuning can improve the performance of the web application in the future along with a usable user interface ensuring global accessibility for various types of users.**

Keywords*—*Loan eligibility prediction; machine learning; random forest; K-Nearest Neighbour; Adaboost; extreme gradient boost; artificial neural network; federated learning

I. Introduction

People all over the world reckon on banks to gain various kinds of financial support depending on their needs. Besides, depositing individual money it provides loans to its customers assessing different conditions and criteria. In general, banks variably provide sixteen types of loan applications [1]. In recent years, the lend-leasing industry has created significant growth increasing number of individuals seeking personal loans for various purposes. This increase in demand has led to a need for more efficient and accurate methods of loan applicant selection. Loan approval criteria defer from bank to bank. Forbes refers to the top five banks in the world providing different personal loan applications and sanctioning criteria with some common attributes [2]. Assessing those top five [3][4][5][6][7] banks, it is seen that some attributes like - credit score, social security number, loan amount, loan type, mortgage information, employment, etc. are common. Depending on these criteria, traditional loan application processing carries forwards with manual reviews and human judgment which can be subjective and biased, leading to inefficient loan processing and higher default rates consuming a huge time in taking a decision which is a cumbersome task of the banking system. Due to human error, sometimes loans are sanctioned mistakenly to some people who cannot repay banks' money with interest in proper time. Moreover, banking sectors more or less face challenges with huge data management and security issues during data processing. But the use of FL in processing all the eligible loan applicants at a time is left behind. The primary motivation behind this research is to tackle the aforementioned challenges progressively, aiming to alleviate the burden on bankers in identifying loan defaulters and streamline the loan sanction process efficiently. By providing swift decisions, this research aims to support loan applicants in making informed choices that depend on the approval of their loans. Additionally, the research aims to expedite the loan sanctioning process, reducing the waiting time for loan applicants. The research introduces a web application developed using ML and DL algorithms for selecting eligible personal loan applicants in an FL approach to ensure security and a better data management process. Since, today's modern world increasingly depends on ML for any type of big data analysis and prediction because of having different statistical models, and banks need more accurate predictive systems, in this research ML models are used for personal loan prediction. In a study [8], loan prediction has been done with a random forest algorithm providing better performance than a decision tree. Thereupon, in this research, the best accuracy-giving algorithm is selected among four ML and one DL algorithms for achieving better performance of data in checking the eligible personal loan applicants among all the submitted applications. The app uses data-driven approaches for analyzing vast amounts of data and making predictions about the candidates who are likely to be selected for the loan sanction. This leads to a more objective assessment of loan applicants and a reduced risk of loan defaults. And another lesson that has been found from analyzing different research on the loan prediction arena is, very few concrete systems have been developed for predicting eligible personal loan applicants ensuring the privacy of loan applicants. The key contributions of the research are:

- To train and test a loan prediction dataset with four ML and one DL algorithm that has been found after the literature review.
- To choose the best-performing ML algorithm among those five for loan prediction.
- To ensure the privacy, security, and robustness of data processing, an FL approach will be utilized with different loan applicant selection datasets.

- Lastly, to develop a web application for checking eligible loan applicants when customers request a loan with an online application to the bank.

The rest of the paper includes a literature review in Section II, methodology in Section III, result analysis and discussion in Section IV, and a conclusion in Section V wraps up the overall research.

II. Background Study

In this section, the research background has been categorized into three subsections: ML-based loan prediction, FL-based loan prediction, and web applications using ML for loan prediction.

A. ML-based Loan Prediction

Authors in [9] used the ML approach to predict eligible candidates to receive loan amounts by collecting previous banks' data who are accredited before. For predicting loans, a simple comparative study was made in [10] based on six machine-learning classification models in R to find out whether allocating a loan to a certain person is risky or not without recommending any specific algorithm. In 2021, a comparison of seven different classifiers was performed in [11], which also showed a method for combining results from multiple classifiers. In [12], authors suggested a better method for performing the identical function in banking procedures. In terms of accuracy, it is shown in a study [13] that the Decision Tree ML algorithm outperformed rather than Logistic Regression and Random Forest ML techniques, according to the results of the trial. In [14], authors made a comparative analysis comprising Random Forest and Decision Trees, declaring the latter to have the highest accuracy when evaluated on the same dataset. To forecast an outcome on loan prediction, a Decision Tree ML algorithm was employed in [15]. In [16], Big Data mining was utilized to collect approved clients' previous data and for training and testing the ML models. Among the four ML models, the Decision tree algorithm gave the best accuracy result. In [17], three ML models are utilized to train the past data to decide whether the loan request will be accepted or not, and among them, the Decision tree algorithm outperformed than Random Forest and Logistic Regression ML approaches. An understandable artificial intelligence (AI) decision-support system was researched to automate the loan underwriting process with a belief-rule-base (BRB) and was capable of learning from and incorporating human knowledge through supervised learning, and historical data [18]. In recent times, authors of [19] made a comparative study in predicting eligible customer loan receivers using five ML algorithms recommending a Decision tree with AdaBoost ML to have the highest accuracy rate where the data cleansing mechanism played an important role. In [20], a logistic regression model was utilized for predicting the problem of forecasting loan defaulters fetching the Kaggle dataset, depending on sensitivity and specificity as the two parameters to compare the performance of the ML model. Authors of [21] used the Logistic regression model to estimate various performance metrics providing a wide range of outcomes disregarding two important variables, such as gender and marital status. A technique was utilized in [22] for developing a model using the information and outcomes of loan applicants who had already submitted applications which discovered that the logistic regression model performs better than other models. Under the assumption that loan quality has a direct impact on a bank's profitability, in [23], a combined logistic regression method and artificial neural network (ANN) was utilized to improve the predictive performance based on real data from a rural commercial bank. In [24], a research project was made intending to create a cutting-edge algorithm to predict events for different financial institutions to protect them from fraudsters while also streamlining the pre-approval procedure for loan applications and the associated verification process. For performing data categorization with good accuracy, K-nearest neighbor (K-NN), decision tree, support vector machine, and logistic regression models are taken into account to measure their performance. A loan default dataset was used in [25], which is taken from the lending club. To address the dataset's class imbalance issue, the ADASYN (Adaptive Synthetic Sampling Approach) method was used in increasing the prediction accuracy. Following an experimental comparison, it was discovered that the fusion model proposed in this paper outperformed using three other models—Logistic Regression, Random Forest, and CatBoost—in terms of its ability to predict the likelihood of customer loan default which was trained with the dataset lowering the external risk posed by customer loan default for the online loan platform. To classify a Kaggle dataset with the best degree of feasible accuracy, it is found that the random forest classification approach provided better performance in loan candidate classification [26]. The authors of the paper [27], researched that the loan grants were given to people in previous years after mining them in their recommended model using random forest ML to predict the loan grants to develop a better risk prediction system for the network loan platform reducing its risks. In [8]also showed that Random Forest Classification outperformed better than the Decision Tree algorithm with a mean accuracy of 89.94% in finding eligible loan applicants after their loan application in a bank. Data Mining Techniques are used in [28], to assess the manual way of loan sanctions made by banks, and following that deep learning models are used to perform the task for prediction. In [29], a proprietary dataset from an agency was utilized to compare the efficacy of a variety of regression models and ML algorithms for forecasting the probability of paying the loan discovering rule-based algorithms to outperform other approaches. A model is created by Debnath et al. in [30], to forecast whether to approve credit for or deny credit utilization for clients using loan application data from consumers. The proposed model took into account the factors that affect a person's loan status and produces precise results for approving or rejecting the customer's request for credit after carefully assessing all available possibilities. To entrench the convolutional neural network (CNN) and the integration model of stacking, a loan risk prediction model called Stacking+CNN was proposed by Li et al. [31]. The prediction model created in this work was superior to the single model and other integrated models in terms of forecasting accuracy and recall rate, according to empirical results. A mechanism for foretelling loan failure was developed by Muslim et al. [32]. For the prediction analysis procedure, an enhanced light gradient boosting machine via features selection using swarm methods such as ant colony optimization and bee colony optimization was applied having a 95% success rate. Authors in [33], utilized an ML method to anticipate loan defaults recommending the Naive Bayes model to perform better than other models. Arutjothi et al. in [34] build a credit rating model using loan status. Credit rating models are used to distinguish defaulters and legitimate consumers. This research used credit data to develop a rating model and presented an ML-based data analysis methodology with K-NN and Min-Max normalization. The proposed approach was 75.8% accurate.

B. FL-based Loan Prediction

In a recent study by [35], Yang et al., an overall description of FL with its use in different sectors like health and communication had been made. Then it found its drawback in security issues and finally discussed its future and its use in the application layer. In [36], an FL approach was utilized by Gu et al. in processing the trained model data and updating the parameters on the centralized server ensuring accuracy, privacy, and model fairness. FL approach had been remarked on by Kawa et al. in [37] for assessing credit risks by learning shared prediction models from different banks collaboratively to update their data in the central repository. Authors in [38] proposed an FL model to predict the loan requester's financial situation using the clients' banking information concluding that the F1-score metric gave identical results in both the centralized and decentralized environment. In [39], federated learning (FL) is used in finding the loan applications that have less possibility to repay the loans in due time, and a Synthetic Minority oversampling Technique (SMOTE) is used in solving the imbalanced data.

C. Web Application using ML for Loan Prediction

Sujatha et al. in [40] referred to the deployment of a web application project that utilizes an ML algorithm named logistic regression for loan prediction with a high accuracy rate. In another study by Thomas et al. [41], a similar type of suggestion has been given to achieve eligible loan applicants. But, comparisons have been made among XGBoost, K-NN, and support vector machine, recommending XGBoost to have the highest accuracy rate of about 91.6%. In a study by Shukla et al. [42], research on loan prediction-based web applications using logistic regression, random forest classification, and XGB ML algorithms has been made using Stream the lit library. The application shows either "Loan denied" or "Loan approved" status to the loan applicant customer after prediction using ML algorithms. The app can be modified to increase its accuracy in the future.

Along with the above three categories, the paper of Divate et al. [43], also predicted the outcome by mining the data of previously accepted clients. The system was developed using an AI model that delivered the most accurate result in this research. Authors in [44], employed LightGBM in predicting categorization outcomes using observational datasets as the most successful algorithm after multi-observation and multi-dimensional data cleaning. In [45], Blaszczy´enski et al. used an upgraded dataset for pre-programmed loan applications to test a tool for financial fraud prediction named DRSA-BRE and found that it performed better than existing methods. Robisco et al. Authors of [46] presented a new framework to compare ML approaches and model risk adjustments. To solve this issue, they first identified up to 13 risk variables using internal ratings-based methods, then grouped them into three primary categories: statistics, technology, and market conduct. Using natural language processing and risk terminology based on expert knowledge, they calculated the weight of each type based on the frequency of its mentions.

The above discussion on background works assisted that myriad works prevail in the selection of eligible loan applicants using ML algorithms with good prediction providing impressive accuracy rate. But still, most of the paper indicates to increase in this accuracy rate. Moreover, web applications based on loan applicants' prediction process couldn't reach huge popularity in research sectors ensuring data security. Therefore, to develop a comprehensive web application that utilizes ML algorithms in predicting eligible loan applicants in an FL environment, further research is needed to address these challenges and ensure the fairness and transparency of the system.

III. Methodology

A. Overview

In this article, an end-to-end solution for loan prediction using ML algorithms with a series of features related to scalability, and security with a distributed federated transfer learning model has been proposed. To ensure client-side rendering with data protection, the aim is to provide a microsystem structure with exchangeable FL capabilities and client-side rendering. Elaborately, the research is working combining three parts namely - ML prediction using loan data, FL for client-side rendering, and Web application development for sanctioning loans.

B. Working Procedure

1) Web application development for sanctioning loan: This is the main software system with whom the bankers (Admin or Bank Employees) will interact. It will work with all online loan applications from customers. The workflow of the proposed system is shown in Fig. 1. Here, a web app is developed commencing with individual access to the system. There are three types of users, namely- Admin, Customer, and Employee. The user Authentication Section will give the required roles according to the logged-in user. If the user type is Admin, then it will be redirected to the "Controls and Operates the whole system". If the user type is Customer, then the individual customers can request a loan from the bank. the system provides the loan sanctioning form to the customer. Customers fill up the form and submit it to the system. Then, the customer has to wait for its approval or rejection. If the user type is Employee, then it can view all the loan requests of the customers. When the bank employee hits the Submit button, the ML prediction analysis starts working with all the loan requests to sort the eligible loan applicants using the best ML algorithm. The process to find the best ML algorithm is shown in Fig. 2. Based on this ML prediction result, the Employee can view the customers who are accepted and rejected for the loan request. The following tools and techniques are used for its development:

- NodeJs is used for backend coding and calling the REST API using a GitHub link.
- Tensorflow javascript library is used to load and run those data.

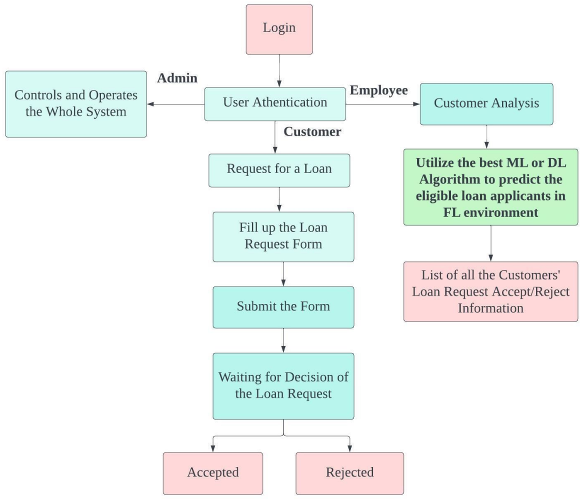


Fig. 1. Workflow diagram of the proposed system.

- After that, the generated result is uploaded to GitHub.
- Using javascript, reactjs, HTML, and CSS, the front end is developed.

2) ML Prediction using loan data: In the literature review section, the use of different ML algorithms has been observed, among which the ML algorithms, which are least popular, performed badly, and worked with similar datasets are chosen. Comprising all the ML and DL algorithms found to be used in similar research, four ML algorithms, namely - Random Forest Classifier, K-NN Classifier, AdaBoost Classifier, XGB Classifier, and one DL algorithm, namely - ANN have been used in this research. These five algorithms are then used to find the best one for developing the web application to predict the customers to whom the loan can be sanctioned or declined. The working procedure to find the best ML algorithm for the system is shown in Fig. 2.

At first, the data is collected from a popular dataset available on Kaggle [47]. It contains genuine 10,001 records of a bank. Then the data pre-processing has been done maintaining the following steps: *i) Null value elimination:* There are some cells in the dataset which has no values. These cells can result in improper results when tested. These null values are filled by the statistical estimation method. *ii) Label Encoding:* Some values are string-type in nature which are converted into numeric values. *iii) Correlation:* Since some attributes (LoanID, CustomerID, and Tax Liens) are not relevant to the model, this process automatically chooses useful features while removing redundant or unnecessary characteristics. Discarding a feature results in an O coefficient value. The data has 19 attributes of customers. Among these 15 attributes are used as independent attributes and 1 attribute as a dependent attribute. The attributes are given in Table I. The whole Dataset is then split into two parts: The Training Dataset and Test Dataset. All five ML models are trained with 8000 data and then tested with the rest. Then an analysis among the models has been done to select the best-performing one with the highest accuracy level. Noticeably, since ANN is a DL algorithm it is trained in the FL environment. A comparative analysis is made among ML and

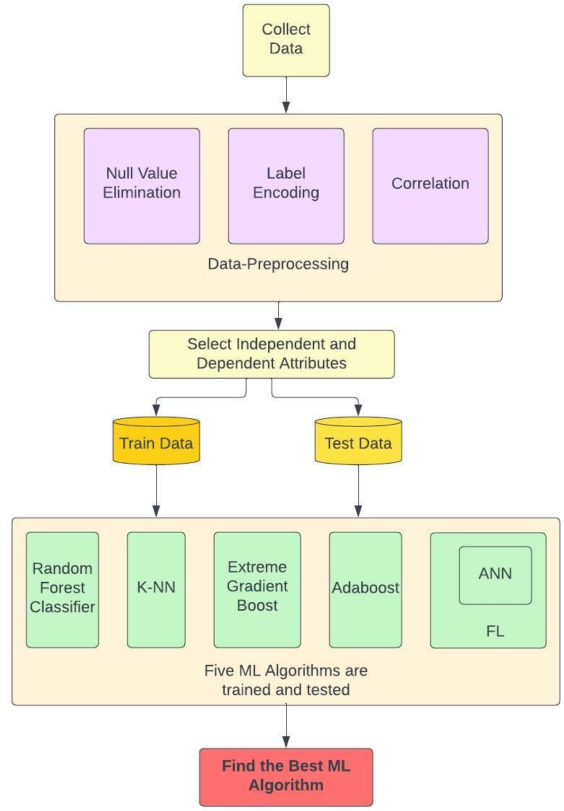


Fig. 2. Workflow diagram to find the best ML algorithm.

TABLE I. ATTRIBUTES

Dependent Attributes	Independent Attributes	
Loan Status	Current Loan Amount	Home Ownership
	Number of Open Accounts	Term
	Number of Credit Problems	Purpose
	Current Credit Balance	Credit Score
	Years of Credit History	Monthly Debt
	Maximum Open Credit	Annual Income
	Years in current job	Bankruptcies
	Months since last delinquent	

DL algorithms to choose the best one. Using the best ML or DL model, eligible customers for loan sanction are predicted and utilized in the proposed system's Utilize the best ML or DL Algorithm to predict the eligible loan applicants in the FL environment as mentioned in Fig. 1.

3) FL for client-side rendering: A federated learning approach is adopted to train the loan property detection model. This approach involves multiple clients, each possessing its local dataset. During each training round, the clients independently perform local training using their respective datasets. This process allows the clients to learn from their data, capturing the specific characteristics and patterns of their datasets. After the local training phase, the clients generate model updates based on their trained models. These updates typically consist of either the updated model parameters or gradients, which represent the direction and magnitude of the parameter updates. The clients then transmit their model updates to the aggregator, a central entity responsible for coordinating the federated learning process. The federated transfer learning approach of ANN consists of several clients, where n number of client nodes N_1, N_2, N_3 .. $\in N_n$ can participate in processing, assuming each device has at least P computational power. The local batch size B and iteration C are adaptive, depending on the user end of the generated data. Each N_i can train on private data PD and a central server-based classification model is shared between all of the nodes with a synchronized upload and download time T_u and T_d respectively. The objective function is to minimize binary cross entropy loss (BCE) over all models referring to equation 1, where y is the ground truth and ε is the prediction with optimization of all of the weight and biases denoted by w_i and b_i respectively as mentioned by Zhang et al. in [48]. This BCE loss is calculated individually for each ANN model on the client side.

$$BCE = -(y \log(\varepsilon) + (1 - y) \log(1 - \varepsilon)) \tag{1}$$

Upon receiving the model updates (BCE) from all participating clients, the aggregator computes their mean, where ∂ $BCE(X_t)$ of equation 2 denotes the gradient descent and η_t is the learning rate of each node, N_i which also denotes the local update of i-th nodes with a learning convergence assumption. Again, X_t refers to the current instance of input in the i-th nodes, which can be calculated from previous instances. A large number of local models are aggregated (e.g., averaged) on the client side to create the global model. As local models are developed utilizing client-specific training data on devices, local and global models often differ. This aggregation step consolidates the model updates into a single global update, representing the collective knowledge from all the clients. By computing the mean of the model updates, the aggregator ensures a fair combination of local knowledge while preventing the dominance of any particular client. By averaging the model updates, the aggregation process balances the contributions of individual clients and facilitates the convergence towards an accurate and generalized eligible loan applicant's prediction model.

$$X_t = X_{t-1} - \sum_{i=1}^{N} \eta_t \, \partial BCE_i(X_t) \tag{2}$$

On deployment, the ML model is trained on the user side and only the prediction and updated model are sent over the network. Thus, any practical or private information needed for the ML model to operate will be separated from the central cloud storage, resulting in a more secure and reliable application system. It also solves critical issues like data security, privacy, and authorized access. This is also a more decentralized approach where edge devices actively participate in computation, reducing the computation complexity on a central server. Therefore, this process also enables reducing the unnecessary model parametric complexity.

IV. Result Analysis and Discussion

This research can determine the eligibility of a customer to get a loan. After getting all the related information about customers, the system checks that using the best ML algorithm, the system can approve or reject the loan applicants.

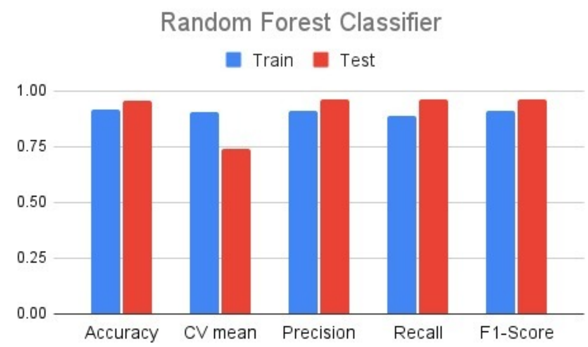


Fig. 3. Performance comparison of random forest classifier.

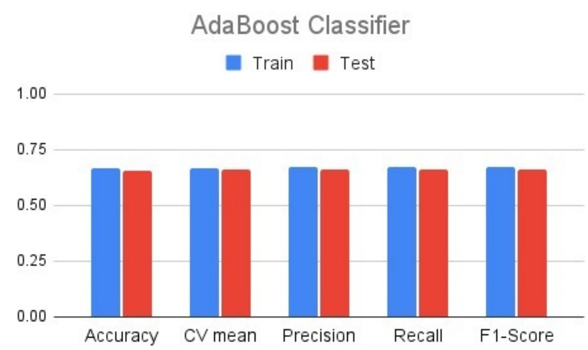


Fig. 4. Performance comparison of Adaboost classifier.

A. ML Prediction Result

To determine the best-performing ML algorithm for the dataset [47], accuracy, CV Mean, Precision, Recall, and F1-score metrics of the confusion matrix are used since most of the research papers referred to in the literature review section used them. The graphs for each of the five algorithms using the above-mentioned metrics are analyzed here:

1) Random forest classifier for loan prediction: Fig. 3 describes that the test data performed better than the training data generating a value near 1 for each of the attributes except for the CV mean. CV means calculated a significant degradation in value compared to all the train and test data. All the metrics of the train data are generating a value near 0.9 except for recall which is slightly below compared to the other metrics.

2) Adaboost classifier for loan prediction: From Fig. 4, it is observed that the training data performed better than the test data for all metrics and are near 0.65 which is poor than the Random Forest Classifier in Fig. 3.

3) K-NN classifier for loan prediction: Fig. 5, Fig. 6, Fig. 7, and Fig. 8 describe graphs for four different values of k. Here, K=7, 11, 13 and, 17 were used to of K-NN identify any significant change in its pattern. When the value of K in Fig. 5 was 7, it was observed that for all the metrics of confusion matrix, the probability was above 0.75 except for CV-mean. But, when the value of K in Fig. 6 was bit increased to 11, it was observed that for all the metrics of confusion matrix,

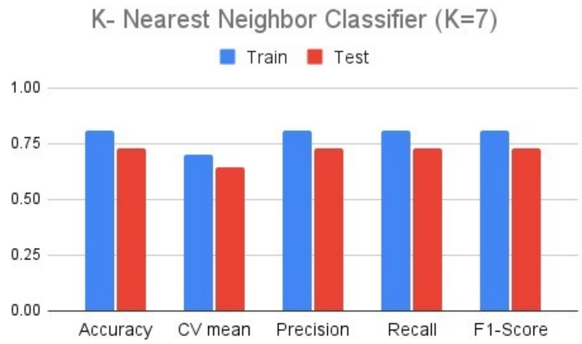


Fig. 5. Performance comparison of K-NN classifier (n=7).

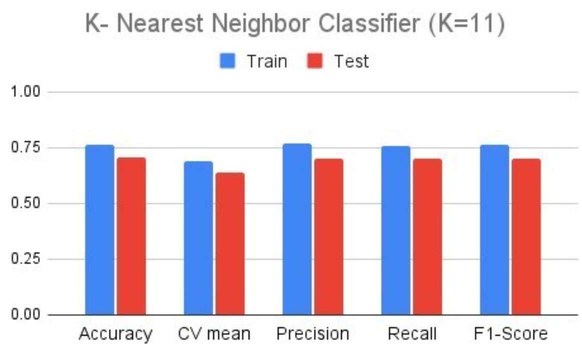


Fig. 6. Performance comparison of K-NN classifier (n=11).

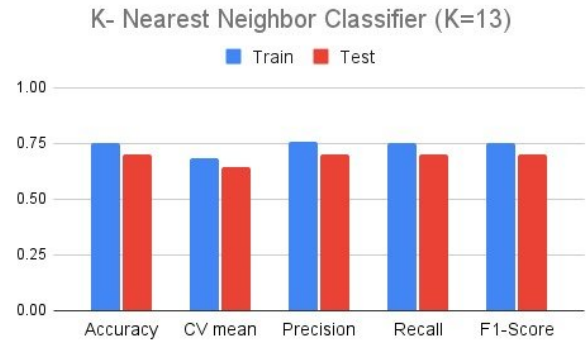


Fig. 7. Performance comparison of K-NN classifier (n=13).

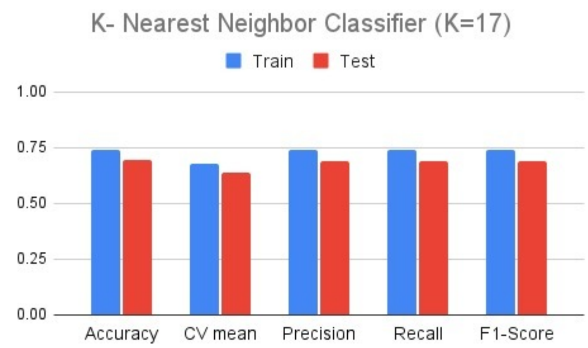


Fig. 8. Performance comparison of K-NN classifier (n=17).

the probability was 0.75 except for CV-mean. Similarly, when the value of K in Fig. 7 was a bit increased from 11 to 13, it was observed that for all the metrics of confusion matrix, the probability was 0.75 except for CV-mean. However, when the value of K in Fig. 8 was searched covering wide range to 17, it was observed that for all the metrics of confusion matrix, the probability was near to 0.75 except for CV-mean. Furthermore, no significant differences were observed in it. For all the four values of k, the training data performed better than test data similar to Adaboost Classifier in Fig. 4 calculating a value around 0.75 for each of the metrics with a significant decrease in the value of CV Mean metrics. But the values are less than the Random Forest Classifier in Fig. 3.

4) Extreme gradient boosting classifier: Fig. 9 describes that the training data performed slightly better than the test data calculating a value near 0.60 for each of the metrics similar to the Adaboost Classifier in 4 and K-NN in Fig. 5 to 8. But couldn't outrage the Random Forest Classifier in Fig. 3.

5) ANN: This section generates Fig. 10 and Fig. 11 using the equation 2 and 1 respectively. Since it uses a neural network to perform the calculation, with the increase in the number of epochs [49] i.e. the learning rate, observing the Fig. 10 and Fig. 11, it is seen that accuracy of FL-based ANN is also increasing for both the training and the testing data with a corresponding decrease in loss value. But since the measurement is made on a scale of 1, the peak value of it is around 0.8 which is less than the Random Forest Classifier in Fig. 3. Considering all the values of each confusion matrix for all the ML and DL algorithms, a comparative graph is created in Fig. 12. In this graph, for the overall analysis, only the accuracy and F1-score metrics are selected for both training and test data since they gave excellent results for all the ML algorithms. However, the FL-based ANN couldn't beat the ML algorithm even after having multiple iterations. Hence, it is concluded that Random Forest Classifier's performance is the best in comparing all the other ML and DL algorithms. This Random Forest Classifier is then used in the FL environment for data analysis of the web application.

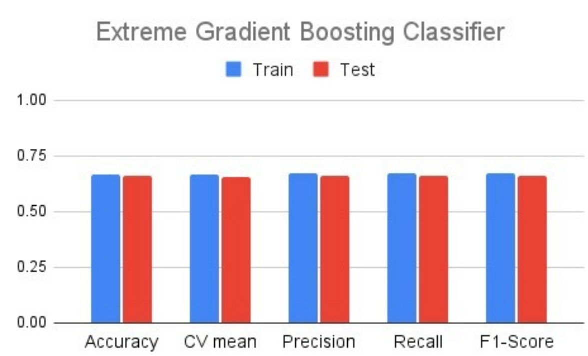


Fig. 9. Performance comparison of XGB classifier model

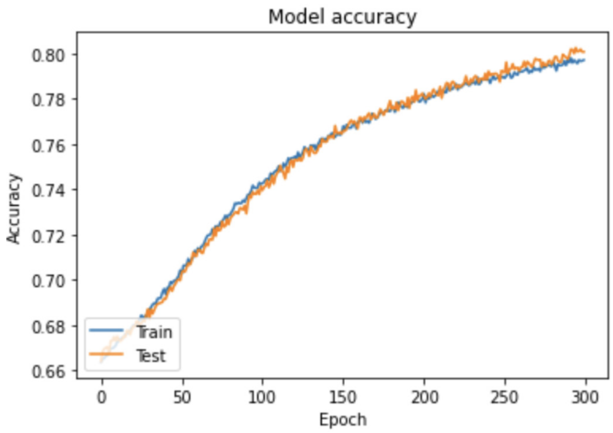


Fig. 10. ANN classifier model accuracy.

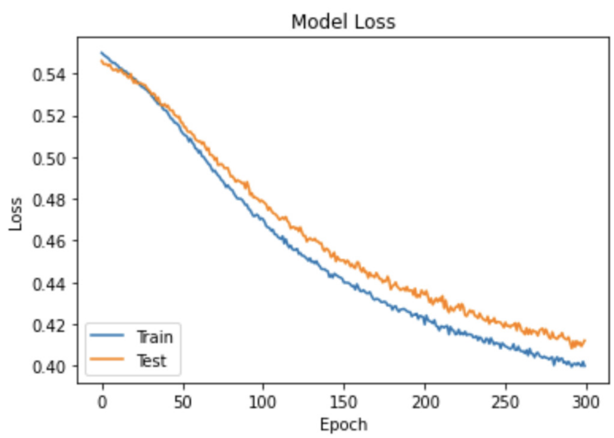


Fig. 11. ANN classifier model loss.

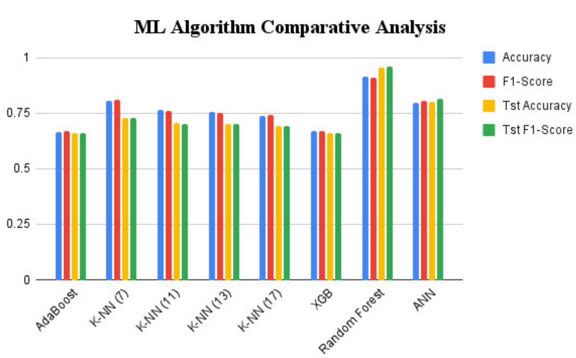


Fig. 12. Comparative analysis of ML algorithms.

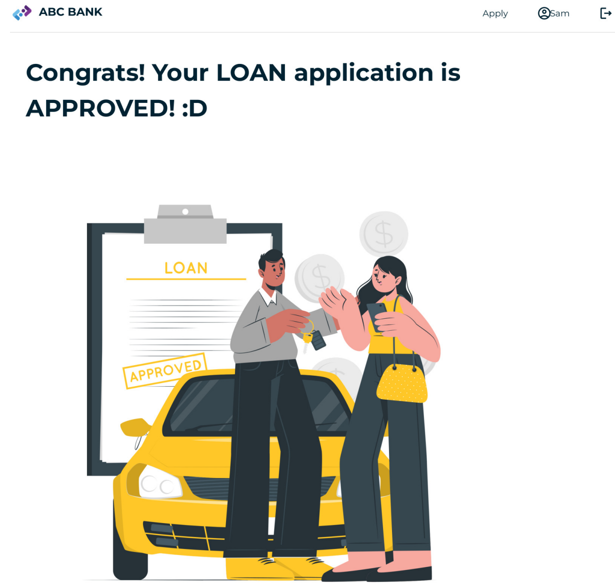


Fig. 13. Loan approval UI of a customer.

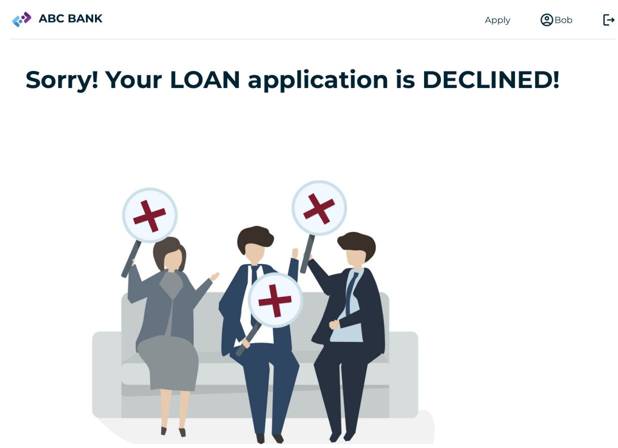


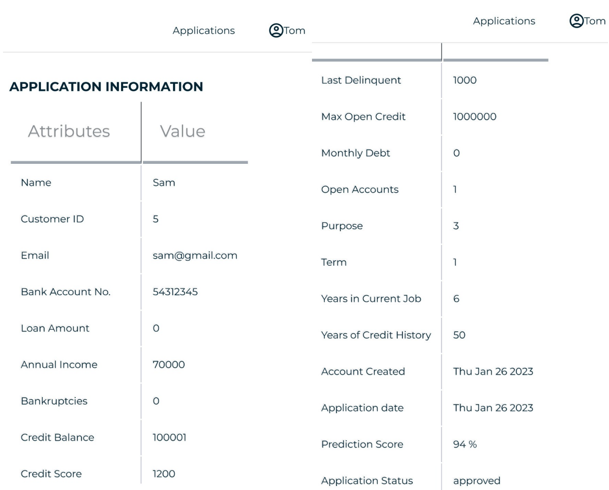
Fig. 14. Loan decline UI of a customer.

B. User Interface (UI) of Web Application

In this section, some salient figures of the developed web application have been highlighted using which the customer and bank employee will interact for loan processing. Here, ABC Bank is considered an exemplary name of a bank.

- *User-Customer:* Fig. 13 shows a customer named Sam has been sanctioned with his requested loan and Fig. 14 shows a loan decline UI for a customer named Bob. However, the customer's application form's UI is skipped from inclusion.
- *User-Employee:* Fig. 15 shows the employee dashboard UI which comes after processing the customers' loan application using ML prediction techniques. In Fig. 15, it is seen that the customer with LoanID: 1 is declined from getting the loan, LoanID: 2 has been approved for loan sanction, and LoanID: 3 and 4's loan requests are still on review status. Employees can review loan requests using ML algorithms. If the

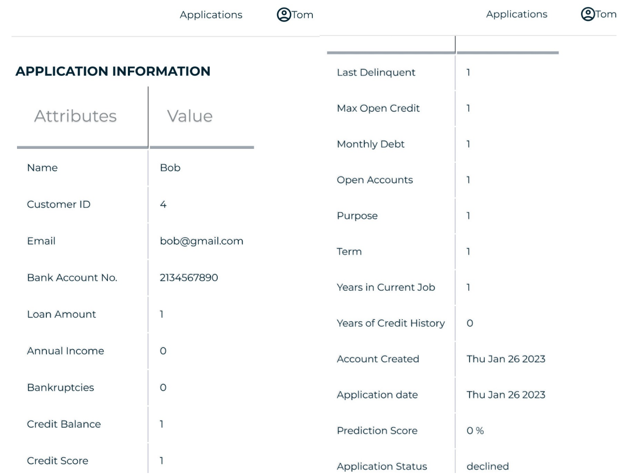
Fig. 15. Employee dashboard.



Attributes	Value
Name	Sam
Customer ID	5
Email	sam@gmail.com
Bank Account No.	54312345
Loan Amount	0
Annual Income	70000
Bankruptcies	0
Credit Balance	100001
Credit Score	1200
Last Delinquent	1000
Max Open Credit	1000000
Monthly Debt	0
Open Accounts	1
Purpose	3
Term	1
Years in Current Job	6
Years of Credit History	50
Account Created	Thu Jan 26 2023
Application date	Thu Jan 26 2023
Prediction Score	94 %
Application Status	approved

Fig. 16. Employee view of the loan approved customer.

employee wants to view the detailed information of the loan applicants then the UI regarding that is shown in Fig. 16 gives an approved loan applicant's details and Fig. 17 gives a declined loan applicant's details. Elaborately, both the figures show all the information of the customer to whom the loan has been sanctioned and whose loan application is rejected respectively using ML algorithms.



Attributes	Value
Name	Bob
Customer ID	4
Email	bob@gmail.com
Bank Account No.	2134567890
Loan Amount	1
Annual Income	0
Bankruptcies	0
Credit Balance	1
Credit Score	1
Last Delinquent	1
Max Open Credit	1
Monthly Debt	1
Open Accounts	1
Purpose	1
Term	1
Years in Current Job	1
Years of Credit History	0
Account Created	Thu Jan 26 2023
Application date	Thu Jan 26 2023
Prediction Score	0 %
Application Status	declined

Fig. 17. Employee view of the loan declined customer.

V. Conclusion

Sanctioning a loan is a challenging task for bankers since there occur some phenomena when the borrowers cannot return their debts in due time. Sometimes, debt cannot be collected too due to some misjudgment. Various types of loans are provided by the banks. In this research, an ML-based web application has been used to check the eligibility of personal loan applicants. To conduct the task, data is used for prediction using four ML and one DL algorithm. The prediction has been performed depending on some attributes in which the most crucial factors that are considered in taking decisions are - loan amount, loan length, loan term, and age. Among those five ML algorithms, Random Forest Classifier has been suggested to be used by the banks since it has given the best result for all the metrics of the confusion matrix. Moreover, another remarkable component of the research is the implementation of a decentralization technique in local PC for data processing using the FL approach to ensure its data security and robustness.

However, the research lacks working with more real and relevant data that can effect the accuracy augmentation of the ml algorithms. It could have worked with more latest ml algorithms which have not been used in this type of research. The back-end architecture of the web application have been developed with modern programming tools.

In future, the research could have work with more real data integrating more empirical attributes that the banks follow and use during their assessment so that the accuracy of prediction can be enhanced. Furthermore, the user interface of the web application can also be enhanced in the future using modern tools and techniques.

References

[1] Forbes, *16 Types of Loans to Help You Make Necessary Purchases*, https://www.forbes.com/advisor/loans/types-of-loans/, Accessed: 18 January 2023.

[2] Forbes, *Top 5 International Banks for Worldwide Banking*, https://statrys.com/blog/best-international-banks/, Accessed: 18 January 2023.

[3] Citi, *APersonal Lines & Loans - See all lines & loans from Citibank® - Citibank*, https://online.citi.com/US/ag/personal-loan?intc=1 7 50 5 LOBA PersonalLoans Pos4/, Accessed: 18 January 2023.

[4] HSBC, *Personal Loan* — HSBC Bank Bangladesh, https://www.hsbc.com.bd/1/2/retail-banking/loans/personal-loan/, Accessed: 18 January 2023.

[5] Capital One, *Let's get personal: Understanding how to get a personal loan*, https://www.capitalone.com/bank/money-management/banking-basics/how-to-get-a-personal-loan/, Accessed: 18 January 2023.

[6] JPMorgan Chase & Co., *AChoosing the Right Mortgage Loan — Chase*, https://www.chase.com/personal/mortgage/mortgage-purchase/choose-a-loan/, Accessed: 18 January 2023.

[7] Thomas (TJ) Porter, *How to Get Bank of America Personal Loans 2022*, https://www.mybanktracker.com/personal-loans/faq/how-to-get-bank-of-america-personal-loans-268385/, Accessed: 18 January 2023.

[8] Kumar, Narra Rahul, and L. Rama Parvathy, *Higher Accuracy on Loan Eligibility Prediction using Random Forest Algorithm over Decision Tree Algorithm*, Baltic Journal of Law & Politics 15, no. 4 (2022): 241-251.

[9] Gupta, Anshika, Vinay Pant, Sudhanshu Kumar, and Pravesh Kumar Bansal, *Bank Loan Prediction System using Machine Learning*, In 2020 9th International Conference System Modeling and Advancement in Research Trends (SMART), pp. 423-426. IEEE, 2020.

[10] Arun, K., Ishan, G. and Sanmeet, K., *Loan approval prediction based on machine learning approach*, 2016. IOSR J. Comput. Eng, 18(3), pp.18-21.

[11] Łuczak, Aleksandra, Maria Ganzha, and Marcin Paprzycki, *Probability of Loan Default—Applying Data Analytics to Financial Credit Risk Prediction*, In Intelligent Systems, Technologies and Applications: Proceedings of Sixth ISTA 2020, India, pp. 1-16. Springer Singapore, 2021.

[12] Ghildiyal, Bhawana, Shubham Garg, and Vivek Raturi, *Analyze of different algorithms of machine learning for loan approval*, In Smart Trends in Computing and Communications: Proceedings of SmartCom 2021, pp. 719-727. Springer Singapore, 2022.

[13] Tejaswini, J., T. Mohana Kavya, R. Devi Naga Ramya, P. Sai Triveni, and Venkata Rao Maddumala, *ACCURATE LOAN APPROVAL PREDICTION BASED ON MACHINE LEARNING APPROACH*, Journal of Engineering Science 11, no. 4 (2020): 523-532.

[14] Madaan, Mehul, Aniket Kumar, Chirag Keshri, Rachna Jain, and Preeti Nagrath, *Loan default prediction using decision trees and random forest: A comparative study*, In IOP Conference Series: Materials Science and Engineering, vol. 1022, no. 1, p. 012042. IOP Publishing, 2021.

[15] Gomathy, C. K., Ms Charulatha, Mr AAkash, and Ms Sowjanya, *The Loan Prediction Using Machine Learning.*, International Research Journal of Engineering and Technology 8, no. 10 (2021).

[16] Supriya, Pidikiti, Myneedi Pavani, Nagarapu Saisushma, Namburi Vimala Kumari, and K. Vikas, *Loan prediction by using machine learning models*, International Journal of Engineering and Techniques 5, no. 2 (2019): 144-147.

[17] Singh, Vishal, Ayushman Yadav, Rajat Awasthi, and Guide N. Partheeban, *Prediction of modernized loan approval system based on machine learning approach*, In 2021 International Conference on Intelligent Technologies (CONIT), pp. 1-4. IEEE, 2021.

[18] Sachan, Swati, Jian-Bo Yang, Dong-Ling Xu, David Eraso Benavides, and Yang Li, *An explainable AI decision-support-system to automate loan underwriting*, Expert Systems with Applications 144 (2020): 113100.

[19] Kumar, Ch Naveen, D. Keerthana, M. Kavitha, and M. Kalyani, *Customer Loan Eligibility Prediction using Machine Learning Algorithms in Banking Sector*, In 2022 7th International Conference on Communication and Electronics Systems (ICCES), pp. 1007-1012. IEEE, 2022.

[20] Sheikh, Mohammad Ahmad, Amit Kumar Goel, and Tapas Kumar, *An Approach for Prediction of Loan Approval using Machine Learning Algorithm*, In 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), pp. 490-494. IEEE, 2020. Kathe2021

[21] Pramod, Ms Kathe Rutika, Ms Dapse Punam Laxman, Ms Panhale Sakshi Dattatray, and Ms Avhad Pooja Prakash, *Prediction of Loan Approval using Machine Learning Algorithm: A Review Paper*, (2021).

[22] Rath, Golak Bihari, Debasish Das, and BiswaRanjan Acharya, *Modern approach for loan sanctioning in banks using machine learning*, In Advances in Machine Learning and Computational Intelligence: Proceedings of ICMLCI 2019, pp. 179-188. Springer Singapore, 2021.

[23] Li, Yiheng, and Weidong Chen, *Entropy method of constructing a combined model for improving loan default prediction: A case study in China*, Journal of the Operational Research Society 72, no. 5 (2021): 1099-1109.

[24] Hemachandran, Kannan, Raul V. Rodriguez, Rajat Toshniwal, Mohammed Junaid, and Laxmi Shaw, *Performance analysıs of different classıfıcatıon algorıthms for bank loan sectors*, In Intelligent Sustainable Systems: Proceedings of ICISS 2021, pp. 191-202. Singapore: Springer Singapore, 2021.

[25] Li, Xingyun, Daji Ergu, Di Zhang, Dafeng Qiu, Ying Cai, and Bo Ma, *Prediction of loan default based on multi-model fusion*, Procedia Computer Science 199 (2022): 757-764.

[26] Bhanu, L., and Dr S. Narayana, *Customer Loan Prediction Using Supervised Learning Technique*, International Journal of Scientific and Research Publications 11, no. 6 (2021): 78.

[27] Pandimurugan, V., D. Usha, M. Nageswara Guptha, and M. S. Hema, *Random forest tree classification algorithm for predicating loan*, Materials Today: Proceedings 57 (2022): 2216-2222.

[28] Pradeep Kumar Singh , Zdzislaw Polkowski Sudeep Tanwar, Sunil Kumar Pandey Gheorghe Matei, Daniela Pirvu, *Innovations in Information and Communication Technologies*, Proceedings of International Conference on ICRIHE - 2020, Delhi, India: IICT-2020.

[29] Bellotti, Anthony, Damiano Brigo, Paolo Gambetti, and Frédéric Vrins, *Forecasting recovery rates on non-performing loans with machine learning*, International Journal of Forecasting 37, no. 1 (2021): 428-444.

[30] Debnath Bhattacharyya,N. Thirupathi Rao, *Machine Intelligence and Soft Computing*, Proceedings of ICMISC 2020.

[31] Li, Meixuan, Chun Yan, and Wei Liu, *The network loan risk prediction model based on Convolutional neural network and Stacking fusion model*, Applied Soft Computing 113 (2021): 107961.

[32] Muslim, Much Aziz, Yosza Dasril, Muhammad Sam'an, and Yahya Nur Ifriza, *An improved light gradient boosting machine algorithm based on swarm algorithms for predicting loan default of peer-to-peer lending*, Indonesian Journal of Electrical Engineering and Computer Science 28, no. 2 (2022): 1002-1011.

[33] Kadam, Ashwini S., Shraddha R. Nikam, Ankita A. Aher, Gayatri V. Shelke, and Amar S. Chandgude, *Prediction for loan approval using machine learning algorithm*, International Research Journal of Engineering and Technology (IRJET) 8, no. 04 (2021).

[34] Arutjothi, G., and C. Senthamarai, *Prediction of loan status in commercial bank using machine learning classifier*, In 2017 International Conference on Intelligent Sustainable Systems (ICISS), pp. 416-419. IEEE, 2017.

[35] Yang, Aimin, Zezhong Ma, Chunying Zhang, Yang Han, Zhibin Hu, Wei Zhang, Xiangdong Huang, and Yafeng Wu., *Review on application progress of federated learning model and security hazard protection*, Digital Communications and Networks 9, no. 1 (2023): 146-158.

[36] Gu, Xiuting, Zhu Tianqing, Jie Li, Tao Zhang, Wei Ren, and Kim-Kwang Raymond Choo., *Privacy, accuracy, and model fairness trade-offs in federated learning*, Computers & Security 122 (2022): 102907.

[37] Kawa, Deep, Sunaina Punyani, Priya Nayak, Arpita Karkera, and Varshapriya Jyotinagar., *Credit risk assessment from combined bank records using federated learning*, International Research Journal of Engineering and Technology (IRJET) 6, no. 4 (2019): 1355-1358.

[38] Imteaj, Ahmed, and M. Hadi Amini, *Leveraging asynchronous federated learning to predict customers financial distress*, Intelligent Systems with Applications 14 (2022): 200064.

[39] Shingi, Geet, *A federated learning based approach for loan defaults prediction*, In 2020 International Conference on Data Mining Workshops (ICDMW), pp. 362-368. IEEE, 2020.

[40] Sujatha, C. N., Abhishek Gudipalli, Bh Pushyami, N. Karthik, and B. N. Sanjana., *Loan Prediction Using Machine Learning and Its Deployement On Web Application*, In 2021 Innovations in Power and Advanced Computing Technologies (i-PACT), pp. 1-7. IEEE, 2021.

[41] Thomas, T. C., J. P. Sridhar, M. J. Chandrashekar, Makarand Upadhyaya, and Sagaya Aurelia., *Developing a website for a bank's Machine Learning-Based Loan Prediction System*.

[42] Shukla, Saurabh, Arushi Maheshwari, and Prashant Johri., *Comparative analysis of ml algorithms & stream lit web application*, In 2021 3rd International Conference on Advances in Computing, Communication Control and Networking (ICAC3N), pp. 175-180. IEEE, 2021.

[43] Diwate, Yash, Prashant Rana, and Pratik Chavan, *Loan Approval Prediction Using Machine Learning*, International Research Journal of Engineering and Technology (IRJET) 8, no. 05 (2021).

[44] Ma, Xiaojun, Jinglan Sha, Dehua Wang, Yuanbo Yu, Qian Yang, and Xueqi Niu, *Study on a prediction of P2P network loan default based on the machine learning LightGBM and XGboost algorithms according to different high dimensional data cleaning*, Electronic Commerce Research and Applications 31 (2018): 24-39.

[45] Błaszczyński, Jerzy, Adiel T. de Almeida Filho, Anna Matuszyk, Marcin Szelag, and Roman Słowiński, *Auto loan fraud detection using dominance-based rough set approach versus machine learning methods*, Expert Systems with Applications 163 (2021): 113740.

[46] Alonso Robisco, Andrés, and Jose Manuel Carbo Martinez, *Measuring the model risk-adjusted performance of machine learning algorithms in credit default prediction*, Financial Innovation 8, no. 1 (2022): 70.

[47] Kaggle, *Classificaçao - DataSet*, https://kaggle.com/code/jcaminha/classifica-ao-dataset/, Accessed: 18 January 2023.

[48] Zhang, Zhilu, and Mert Sabuncu., *Generalized cross entropy loss for training deep neural networks with noisy labels*, Advances in neural information processing systems 31 (2018).

[49] Deepchecks, *Epoch in Machine Learning*, https://deepchecks.com/glossary/epoch-in-machine-learning/, Accessed: 18 January 2023.