Enhancing Bank Loan Authentication with Machine Learning –Based Customer Credibility Predictions: A comparative Study

K.C.Bhanu[1], Dr.P.Uma Maheswari Devi[2]
[1]Research Scholar, Department of Commerce and Management Studies
Adikavi Nannayya University ,Rajahmundry
[2]Associate Professor , Department of Commerce and Management Studies
Adikavi Nannayya University ,Rajahmundry
bhanu1605@gmail.com,umadevi_4@yahoo.com

ABSTRACT

The banking industry contributes significantly to the global economic development of any country. The quantity of credit that has been granted to the public is one of the primary sources of a bank's income. Credit gains account for the majority of the bank's financial gains. Now, it's important for the banks to make a sound decision when determining who should be given credit. The personnel are unsure of whether the accepted consumer would return the credit even after the long manual vetting of credit applicants. We are attempting to lessen the vulnerability behind the authorised person in order to lessen the stress on the bank staff and to save their resources. To do this, a dataset from the Kaggle library that includes the historical data of credit applicants who were approved for credit may be used. Our main objective is to determine if a consumer is eligible for a loan or not by predicting their eligibility. We have tested a broad range of machine learning techniques. To minimise staff effort, which results in a computation error while trying to locate a candidate for credit authentication. Here, we look at the different information about the customers, with features like income, past credit history, educational status, and their asset information from previous records of credit applicants regarding their loan approval, and the best elements are chosen which have a clear impact on the outcome for our credit authentication system.

KEYWORDS: Credit authentication, Machine learning, Classification models, Testing and training set, Applicant details, Decision making.

INTRODUCTION

Loan disbursements are the majority of banks' main source of income. The vast majority of a bank's income comes from loans issued to applicants. The banks in this area charge interest on loans issued to applicants. The bank's principal objective is to spend its resources in reliable customers. If a certain consumer is eligible for and capable of repaying a loan, it is determined using a model called Loan Prediction System. The user's marital status, income, spending patterns, and other criteria are all examined by this procedure. Credit authentication is quite beneficial for both clients and lenders. *Random classifier to assess the constructive power of the two-way categorization systems in terms of actual customer credibility rank and the* *banking credit proposals[2]*. This study's objective is to choose verified applicants, making it a simple, fast, and efficient method for the credit approval procedure.

The following are the primary objectives.

- Based on a customer's key characteristics, determining if they can repay the loan.
- Selecting the ideal model for approving the customer's credit request.
- Assessing the creditworthiness of the applicants using the customer segmentation.

The management of money and financial resources is a critical component of business and economics, and it includes financial analysis, investment management, risk management, and financial planning. It applies economics, accounting, statistics, and mathematics ideas to help people and organisations make sound financial choices. The machine learning loan approval system is developing an algorithm that can estimate the likelihood of loan default based on a variety of characteristics such as the customer's credit rate, salary, prior employment records, debt-to-income ratio, and other relevant information. The algorithm will be trained using past loan data to identify patterns and connections between loan repayment and a variety of factors. The programme will then be tested against a new set of data to determine its accuracy and recommend areas for improvement. The algorithm employed must find a compromise between prediction accuracy and erroneous positives and false negatives. Machine learning is increasingly being used in many areas, including banking, healthcare, transportation, and marketing, to automate processes, make predictions, and draw insights from data. The three primary subcategories of machine learning are supervised learning, unsupervised learning, and reinforcement learning. Unsupervised learning is a technique for discovering patterns and structure in unlabeled data, as opposed to supervised learning, which involves training a model using tagged data. Reinforcement learning is used to educate a decision-making agent to consider input from the environment. It is quite important to determine if an application will return the credit or not.To concentrate on these particular clients, the task has been set to identify the optimal customer groups that are qualified for the appropriate loan type.The major goal of the challenge is to predict which customers will be able to return their credits and which will not be able to return the credits. We will use various ML algorithms such as logistic regression, decision trees, random forests, and gradient boosting classifiers, as well as customer segmentation to determine which type of credit the applicants are eligible for. classify them into clusters so that we can determine which type of credit the applicants are eligible for. It is necessary to develop an accurate model with a lower error percentage. The thesis's main purpose is to prove that granting credit to a certain consumer is trustworthy. The most recent advances in data mining and machine learning approaches have prompted interest in implementing these technologies in a variety of fields. The banking sector is no exception, and the rising necessity for financial institutions to have effective threat management has prompted a strong interest in developing new risk estimating approaches.Banks consider a number of risk variables when calculating the possible loss they may incur in the future. The anticipated loss that the bank could experience if an application declines is also derived from their measurements. Logistic Regression, which is said to be a superior approach for assessing the chance of application default. A range of machine learning methodologies will be investigated and researched in this article to determine whether those models can supplant the methods now in use. This article examines the previous records of accepted and disapproved applicants for the bank loan approval procedure, which were obtained from the Kaggle public repository. Using these data, the project conducts several machine learning approaches analysis to determine which supervised ML algorithm displays greater accuracy for forecasting loan eligibility after training the previous records and storing the model for our bank loan authentication system. In addition, customer segmentation was performed, which classifies applicants using an unsupervised machine learning algorithm based on the most dominant characteristics retrieved using the feature engineering approach and saving the cluster model to the bank loan authentication system. The models have been optimised for the bank loan authentication system in order to lower the risk of default and increase customer satisfaction by making quick and accurate loan decisions. Loan approval and complexity may be forecasted using a variety of data analytics approaches. To predict the kind of loan, prior data of approved and disapproved applicants must be trained using multiple algorithms before being compared to test data. Identifying commonalities in a training dataset of frequently approved loans and then developing a model based on these discovered patterns. The training dataset is sent on to the ML models, and the best classifier is generated using this training dataset. The specifics of each prospective application form serve as test data. To develop the model, several ML methods such as logistic regression, random forest, decision tree, support vector classifier, gradient boosting, GaussianNB classifier, KNN classifier, and lastly linear discriminant were employed. To get reliable prediction results, the better models with more accuracy have been subjected to hyperparameter tweaking. All classification models are built and trained using classification methods. Following a study of the prediction results, the random forest classifier is finally chosen as the best model for classification since it predicts with more accuracy than other machine learning models.In addition, customer segmentation was performed, which classifies applicants using an unsupervised machine learning algorithm based on the most dominant characteristics retrieved using the feature engineering approach and saving the cluster model to the bank loan authentication system.

DATASET

For our loan authentication procedure, the dataset was obtained from the Kaggle open access data source. There are 614 observations in the sample data, containing 13 characteristics. The labels for the input and output are part of the data that was gathered.The characteristics (input labels) include the following: Loan number, sexual orientation, marital status, number of dependents, employment status, educational level, credit amount and duration, credit history, and asset information. The decision of the lenders' loan approval procedure is shown by the output labels.The majority of the time, these labels are binary, with 1 signifying an approved loan and 0 signifying a rejected loan.

The dataset link is given below:
https://www.kaggle.com/datasets/vikasukani/loan-eligible-dataset

The dataset was gathered from a bank loan application and comprises of sample observations with 13 attributes. The dataset was pre-processed to handle missing values, encode category categories, and scale numerical variables.

Variable Name	Description	Data Type
Loan_ID	Loan reference number (Unique I.D.)	Numeric
Gender	Applicant gender	Categorical
Married	Applicant marital status	Categorical
Dependents	Number of family members	Numeric
Education	Applicant educational qualification (graduate or not graduate)	Categorical
Self_Employed	Applicant employment status (yes for self-employed, no for employed/others)	Categorical
Applicant_Income	Applicant's monthly salary/income	Numeric
Coapplicant_Income	Additional applicant's monthly salary/income	Numeric
Loan_Amount	Loan amount	Numeric
Loan_Amount_Term	The loan's repayment period (in days)	Numeric
Credit_History	Records of applicant's credit history (0: bad credit history, 1: good credit history)	Numeric
Property_Area	The location of the applicant's home (Rural/Semi-urban/Urban)	Categorical
Loan_Status	Status of loan (Y: accepted, N: not accepted)	Categorical

METHODOLOGY

MACHINE LEARNING

The "machine learning" sector of artificial intelligence (AI) entails training computer systems to detect patterns in data, draw inferences from it, and make predictions or judgements without being specifically directed to do so. Big datasets are used in machine learning projects to train machine learning models, which are subsequently applied to fresh data to provide predictions or choices. The ML algorithm aids in the discovery of patterns and correlations in data. It also entails developing a sophisticated model that can be trained autonomously from data and can predict future events and provide results without the need for human intervention. Machine learning is the process of employing algorithms and statistical models to analyse data, identify patterns, and make predictions or judgements based on that data. It is a powerful tool that may be used to a variety of jobs to improve accuracy, efficacy, and decision-making abilities.

SUPERVISED MACHINE LEARNING

What is meant by supervised learning is that it requires marked/classified datapoints to create an algorithm that properly categorises the datasets or forecasts the outcomes. Both the input and output datasets are marked.A model is trained using a labelled dataset with

ISSN PRINT 2319 1775 Online 2320 7876

Research paper © 2012 IJFANS. All Rights Reserved, UGC CARE Listed (Group -I) Journal Volume 11, Iss 9, Sep 2022

known inputs (features) and known intended outputs (labels) in this form of machine learning.

UNSUPERVISED MACHINE LEARNING

A model is trained on a dataset with no labels but known inputs and unknown outcomes using this form of machine learning. In a loan approval project, unsupervised learning might be utilised to discover patterns in the data, such as grouping similar credit application forms based on their features.

REINFORCEMENT MACHINE LEARNING

The algorithms are mainly concerned with agents and the activities that agents should do in each environment. As the programme progresses through its inquiry, information is offered in the form of incentives and punishments. This algorithm is programmed into the computer to make choices.

MODEL BUILDING

A bank loan approval system's goal is to predict whether a credit application will be approved or rejected based on a variety of variables, such as income, credit score, job status, and loan amount. This binary classification issue has two classifications for the target variable: authorised (1) and refused (0).

MODEL 1

Logistic regression: This linear model is straightforward, fast to train, and easy to grasp. It works well, particularly when the qualities and the target variable have a roughly linear connection. In LR, a linear equation is used to represent the relationship between the attributes and the target variable's log-odds. The logistic function is then used to determine the likelihood that the loan application will be approved given the characteristics using the log-odds.

$$p = 1 / (1 + e^{-(b0 + b1*x1 + b2*x2 + ... + bn*x n)})$$

where:
p is the predicted probability of the event occurring
e is the mathematical constant (approximately 2.71828)
b0 is the intercept term.
b1, b2, ..., bn are the coefficients for the predictor variables x1, x2, ..., xn

Decision Tree: As a non-parametric model, it can handle both category and numerical information. It can be understood and is able to capture complex relationships between traits. A tree-like structure is created when employing the decision tree technique, with each node signifying a feature and each edge signifying a possible value of the feature. The tree (Approved or rejected) is produced by recursively splitting the data into the features that most precisely divide the target variable. The tree is then pruned to avoid

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 9, Sep 2022

overfitting. *decision-tree predictive model to make it easier to determine whether bank customers are eligible based on their attributes[5]*

$$\boldsymbol{Decision}(\boldsymbol{x}) = \boldsymbol{argmax}\ \boldsymbol{I}\ [\boldsymbol{P}(\boldsymbol{Y} = \boldsymbol{I}\ |\boldsymbol{X} = \boldsymbol{x}\ )]$$

Where:
Decision (x) is the predicted class label for input x,
P(Y=I | X=x) is the Probability of class I given input x,
And argmax is the function that returns the argument that maximizes the expression inside the brackets

Random Forest: To increase prediction accuracy, this ensemble model combines several decision trees. It works well with big datasets and can tolerate noisy data. The random forest algorithm is an effective and flexible technique that may be used to address classification problems in loan approval systems. The random forest method uses a set of randomly chosen data and feature subsets to construct a collection of decision trees. Each decision tree is built using a random subset of the attributes and an arbitrary sampling of the training data. The optimal outcomes are obtained by combining the forecasts of all these different decision trees.

$$\mathbf{y = mode(T1(x), T2(x), ..., Tn(x))}$$

where:
y is the predicted output for the input x
T1, T2, ..., Tn are the individual decision trees in the forest.
mode is a function that returns the most frequently occurring output among the predictions of the individual trees.

Support Vector Machine: On small to medium-sized datasets, this non-parametric model works well. It can manage interactions between linear and non-linear features.SVM works (accepted or rejected) by finding a hyperplane that best splits the data into discrete classes. By choosing the hyperplane, the margin—the distance between the hyperplane and the closest data points from each class—is maximised. The kernel approach, used by SVM, moves the datapoints into a higher-dimensional area where a hyperplane may split them when the data cannot be divided linearly.

$$\mathbf{y = sign\ (w\ \hat{}T\ x + b)}$$

where:
y is the predicted output for the input x
w is the weight vector
b is the bias term
x is the input vector

IJFANS INTERNATIONAL JOURNAL OF FOOD AND NUTRITIONAL SCIENCES

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 9, Sep 2022

Gradient Boosting Classifier: To improve prediction accuracy, this ensemble model also brings together a number of weak learners. It works well with big datasets and can tolerate noisy data.When using gradient boosting, the algorithm is trained repeatedly on the training set, focusing on the data points that the previous iteration mistakenly categorised. The method creates a new decision tree to try to correct the errors made by the previous trees in the ensemble.

f(x) = sum_i=1^M gamma_i h_i(x)

where:

F(x) is the predicted output for the input x.

Gamma I is the learning rate for the I-model h I(x)

h I(x) is the I-week model that is trained to predicted the residual error of the previous model.

Gaussian Naive Bayes: This approach uses a random process to decide how probable it is for a given sample to belong to each class based on the values of each characteristic. In Gaussian Naive Bayes, the characteristics within each class are supposed to be independent and consistently distributed. The algorithm determines the means and variances of each feature for each class using the training data. The Bayes theorem is then used to predict the probability that a given sample belongs to each class using these parameters. When the independence assumption is broken if the feature distributions are not Gaussian, it could not work as well.

P(y | x1, x2, ..., xn) = P(y) * P(x1 | y) * P(x2 | y) * ... * P(xn | y) / P(x1, x2, ..., xn)

where:
P(y | x1, x2, ..., xn) is the posterior probability of the target class y given the input features x1, x2, ..., xn
P(y) is the prior probability of the target class y
P(x1 | y), P(x2 | y), ..., P(xn | y) are the likelihood probabilities of the features given the target class y
P(x1, x2, ..., xn) is the evidence probability of the input features

K-Nearest Neighbour :The basic principle underlying KNN is to classify a new sample using the class labels of its k-nearest neighbours in the training set, where the distance between the new sample and the neighbours is determined using a distance metric, such as Euclidean distance.The user must choose the hyperparameter k, which is used by KNN. Higher values of k may lead to smoother choice borders, whilst lower values of k can lead to more complicated decision borders. The curse of dimensionality will, however, cause it to perform poorly when the dataset is unbalanced or if there are several attributes.

$$y = mode(y_1, y_2, ..., y_k)$$

where:
y is the predicted output for the input data point
$y_1, y_2, ..., y_k$ are the output values of the k-nearest neighbors of the input data point
mode is the function that returns the most common output value among the k-nearest neighbors

Linear Discriminant:The main objective of LDA is to identify an ordered collection of characteristics that best differentiates between various types of loan approval results. (Rejected or accepted).The mean and covariance matrices are initially generated using the LDA approach for each class of loan approval outcomes. The next step is to compute the within-class scatter matrices and the between-class scatter matrices in order to identify the linear combination of characteristics that maximises the ratio of between-class variance to within-class variance. The discriminant function, a linear combination, may be used to group new loan applicants into different groups based on their characteristics. The curse of dimensionality will, however, cause it to perform poorly when the dataset is unbalanced or if there are several attributes.

$$y = \text{argmax } c \ (P\ (C=c|\ X=x))$$
$$P\ (C=c|\ X=x) = P\ (X=x|\ C=c) * P(C=c) / P(X=x)$$
$$P(X=x) = \text{sum } c\ (P\ (X=x|\ C=c) * P(C=c))$$

Where:

Y is the predicted class label for the input x

C is the set of possible class labels

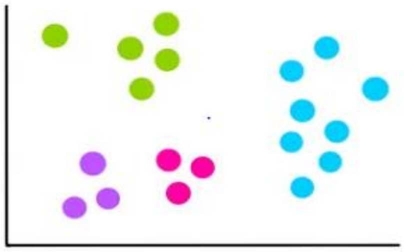
P(C=c | X=x) is the probability of class c given the input x ,which is estimated using Bayes theorem .

It is crucial to evaluate the effectiveness of each algorithm using the appropriate standards and choose the one that best addresses the specific issue at hand.

MODEL 2

Customer segmentation is a crucial step in developing a bank loan authentication system since it helps lenders to identify and classify borrowers based on their characteristics, behaviour, and trustworthiness. Additionally, it helps candidates to understand the grounds for approval and rejection. Lenders may do this by modifying their lending practises and risk assessment procedures in order to better match the needs of each industry and decrease their susceptibility to default risk.

K-Means Clustering: The K-means cluster technique divides a dataset into k clusters, and then assigns each data point to the cluster with the closest centre. The approach seeks to limit variance of data points inside each cluster and increase variance between clusters.



$$J(V) = \sum_{i=1}^{c} \sum_{j=1}^{ci} (||x_i - v_i||)^2$$

$||x_i-v_i||$ is the Euclidean distance between x_i and v_i
'c_i' is the number of data point in i^{th} cluster
'c' is the number of cluster centers

RESULTS

Following the successful implementation of our suggested technique in the bank loan approval system, we arrive at the following categorization and customer segmentation model findings.

CLASSIFICATION MODEL 1

We come up with the results after evaluating with several metrics, such as cross validation score, accuracy, precision, recall, f1 score, and roc curve, after training the eight supervised machine learning algorithms, which include logistic regression, support vector classifier, random forest, decision tree, gradient boosting, Gaussian NB classifier, KNN classifier, and finally linear discriminant analysis for our binary classification problem.

Model Name	**Cross Validation Score**	**Accuracy**	**Precision**	**Recall**	**F1 Score**	**ROC Area**
Logistic Regression	0.80478	0.80180	0.77777	1.0	0.87500	0.61
Support Vector Machine	0.79389	0.79279	0.77	1.0	0.87005	0.65
Decision Tree	0.72337	0.74774	0.82666	0.80519	0.81578	0.71

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 9, Sep 2022

Random Forest	0.79027	0.75675	0.77173	0.92207	0.84023	0.77
Gradient Boosting	0.77400	0.79279	0.78125	0.97402	0.86705	0.69
Gaussian Naïve Bayes	0.78668	0.82882	0.80208	1.0	0.89017	0.70
K-Nearest Neighbour	0.73415	0.71171	0.73684	0.90909	0.81395	0.61
Linear Discriminant Analysis	0.80478	0.80180	0.77777	1.0	0.87500	0.61

Since we trained the model using cross fold combinations, we see the cross-validation metric as the most crucial metric for our model evaluation. Therefore, a summary of the overall performance of the different models using the cross validation measure is shown below.

Logistic Regression	0.80478
Support Vector Machine	0.79389
Decision Tree	0.72337
Random Forest	0.79027
Gradient Boosting	0.77400
Gaussian Naïve Bayes	0.78668
K-Nearest Neighbour	0.73415
Linear Discriminant Analysis	0.80478

We typically choose the top three models with the highest cross validation score from this report. Later, using randomised search to induce hyperparametric tuning on the three models (Logistic Regression with a score of 80.48, Support Vector with a score of 79.39, and Random Forest with a score of 79.03), the best model for our proposed system is discovered.

Below are the model results after hyperparametric tuning:

Logistic Regression score before Hyperparametric tuning	80.48
Logistic Regression score after Hyperparametric tuning	80.48

SVC score before Hyperparametric tuning SVC score after Hyperparametric tuning	79.38 80.66
Random forest Classfier score before Hyperparametric tuning Random forest Classfier score after Hyperparametric tuning	79.03 80.66

After hyperparametric tweaking, Random Forest and Support Vector tend to provide results that are similar, therefore Random Forest Classifier is selected as the best model for our bank loan approval system based on the cross validation score.

Accuracy	0.81
Precision	0.78
Recall	1.0
F1-Score	0.88

The Random Forest Classifier's score increased from 79.03% to 80.66% after the hyperparameter adjustment, which is a substantial improvement. This suggests that the model's performance was able to be optimised by the hyperparameters, making it more suitable for the loan approval system.It generates predictions on whether to approve or refuse credit applications based on the attributes included in the dataset by choosing and storing the model to our Bank Loan Authentication System. These attributes may include details on the applicant's credit standing, income, employment status, loan amount, etc.

CUSTOMER SEGMENTATION MODEL 2

The dataset has been split into two groups, authorised and disapproved applications, for segmenting the applicants based on the most crucial attributes. With three clusters for the accepted applicants, KMeans clustering is utilised to discover trends in applicant income, credit history, and loan amount since they were selected as the most crucial variables for our model from the feature selection. The candidates that were turned down are grouped into the same three KMeans clusters. The results of the clustering are shown using scatter plots. The findings of the clustering show that the data has clear patterns that may be utilised as a guide when deciding whether to accept a loan application.

The Silhouette score metric, which measures how effectively the clustering method worked on the training dataset, is used to assess the performance of the Kmeans clustering model. It assesses how closely a given item resembles its own cluster in relation to the other clusters.The following table shows the Silhouette rating of the kmeans clustering model for the two groups of accepted and rejected applicants:

Silhouette score for approved Customers	0.7381

Silhouette score for unapproved Customers	0.7472

Since the silhouette score for each category ranges from -1 to 1, it may be concluded that the clusters are well defined and spaced out enough from one another. Additionally, the score is higher than 0.5, which shows that the model performs well for our clustering model. In order to anticipate the applicant's cluster in the future based on the most predominate characteristics we chose, the K-means model is kept. Both lenders and applicants may learn about loan acceptance and refusal thanks to this.

FUTURE WORK

As technology develops, automated loan approval systems based on machine learning have a bright future. Despite the fact that the Random Forest Classifier seems to be the most effective model based on these results, it is essential to thoroughly evaluate the model's effectiveness on a holdout dataset before using it in a practical setting. This ensures the model's reliable operation and its capacity for generalisation to new data. The moral implications of utilising machine learning algorithms to approve loans must also be considered, as must the impartiality and fairness of the process for all applicants.

CONCLUSION

A machine learning-based loan approval system might significantly improve the accuracy and efficiency of our automated bank loan authentication system. Using past loan data and other relevant information, machine learning algorithms may assess and identify patterns that point to creditworthiness or the potential for default. This may reduce the possibility of fraud and assist lenders in selecting which applications to approve or reject based on applicant segmentation. Additionally, the borrowers learn the reasons behind the acceptance or refusal of their loan. It's critical to realise that not all loan approval issues can be resolved by machine learning. Maintaining client privacy and security while ensuring impartiality and fairness in loan choices are still crucial. Additionally, machine learning models may have biases or flaws that result in bad judgements. They are not flawless.

REFERENCES

1. Zijiang Yang, Wenjie You, Guoli Ji,Using partial least squares and support vector ma chines for bankruptcy prediction,Expert Systems with Applications,Volume 38, Issue 7,2011,Pages 8336-8342,ISSN 0957-4174,https://doi.org/10.1016/j.eswa.2011.01.021.
2. A. Vaidya, "Predictive and probabilistic approach using logistic regression: Application to prediction of loan approval," 2017 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Delhi, India, 2017, pp. 1-6, doi: 10.1109/ICCCNT.2017.8203946.

ISSN PRINT 2319 1775 Online 2320 7876

Research paper © 2012 IJFANS. All Rights Reserved, UGC CARE Listed (Group -I) Journal Volume 11, Iss 9, Sep 2022

3. A. Vaidya, "Predictive and probabilistic approach using logistic regression: Application to prediction of loan approval," 2017 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Delhi, India, 2017, pp. 1-6, doi: 10.1109/ICCCNT.2017.8203946.
4. Xiaojun Ma, Jinglan Sha, Dehua Wang, Yuanbo Yu, Qian Yang, 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,Volume 31,2018,Pages 24-39,ISSN 1567-4223,
5. M. Alaradi and S. Hilal, "Tree-Based Methods for Loan Approval," 2020 International Conference on Data Analytics for Business and Industry: Way Towards a Sustainable Economy (ICDABI), Sakheer, Bahrain, 2020, pp. 1-6, doi: 10.1109/ICDABI51230.2020.9325614.
6. M. A. Sheikh, A. K. Goel and T. Kumar, "An Approach for Prediction of Loan Approval using Machine Learning Algorithm," 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), Coimbatore, India, 2020, pp. 490-494, doi: 10.1109/ICESC48915.2020.9155614.
7. P. Kirubanantham, A. Saranya and D. S. Kumar, "Credit Sanction Forecasting," 2021 4th International Conference on Computing and Communications Technologies (ICCCT), Chennai, India, 2021, pp. 155-159, doi: 10.1109/ICCCT53315.2021.9711790.
8. S. Barua, D. Gavandi, P. Sangle, L. Shinde and J. Ramteke, "Swindle: Predicting the Probability of Loan Defaults using CatBoost Algorithm," 2021 5th International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 2021, pp. 1710-1715, doi: 10.1109/ICCMC51019.2021.9418277.
9. H. Ramachandra, G. Balaraju, R. Divyashree and H. Patil, "Design and Simulation of Loan Approval Prediction Model using AWS Platform," 2021 International Conference on Emerging Smart Computing and Informatics (ESCI), Pune, India, 2021, pp. 53-56, doi: 10.1109/ESCI50559.2021.9397049.
10. Justin Munoz, Ahmad Asgharian Rezaei, Mahdi Jalili, Laleh Tafakori,Deep learning based bi-level approach for proactive loan prospecting,Expert Systems with Applications,Volume 185,2021,115607,ISSN 0957-4174,https://doi.org/10.1016/j.eswa.2021.115607.