ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

CUSTOMER RELATIONSHIP MANAGEMENT: TWO DATASET COMPARISON IN PERSPECTIVE OF BANK LOAN APPROVAL USING MACHINE LEARNING TECHNIQUES

DURGESH KUMAR SINGH[1], NOOPUR GOEL[2]

[1]Research Scholar, Department of Computer Applications, VBS Purvanchal University, Jaunpur, India

[2]Professor, Department of Computer Applications, VBS Purvanchal University, Jaunpur, India

E-mail: [1]durgeshsingh111@gmail.com, [2]noopurt11@gmail.com

ABSTRACT

One of the main goals of any banking industry that wants to last for a long time is to become profitable. Understanding the customer is necessary for providing services and products to the customer in accordance with his preferences and requirements. Client division and profiling are essential in accomplishing two principal goals of CRM (Customer Relationship Management) i.e.; client maintenance and client advancement. Under CRM, loans are the most important product that banks and other financial institutions offer to meet customers' needs, but determining which customers are eligible for loans is a significant challenge. However, loans come with a risk of default, or the possibility that some borrowers will not be able to repay the loans they have been given. As a result, banks that have a lot of non-performing loans may go bankrupt or become unstable as a result. The progression of innovation like artificial intelligence (AI), getting helpful data from client information is of central significance in nowadays. By developing, comparing, and testing the accuracy of various models using datasets from two banks, we contribute to commercial banks' efforts to predict borrowers' behaviors. Throughout the manuscript, the first dataset is referred to as Dataset-1, and the second dataset is referred to as Dataset-2. In order to determine which machine learning method is most effective for predicting bank loan default, base learners, ensemble, and voting are used. The results demonstrate that Random forest (RF) outperformed all other classifiers. Precision, recall, the f1-score, and the ROC (AUC) curve all supported the classifiers' findings. Because it saves both time and money, I would suggest that financial institutions employ machine learning methods.
Keywords: *CRM, Bank loan, Artificial intelligence, Ensemble, Voting, Recall, ROC (AUC).*

1. INTRODUCTION

Due to financial constraints, banks' loans have become an important source of external financing for businesses and households. Lending to the economy is a very profitable business for commercial banks, and loans make up a significant portion of their assets [1]. However, there are a number of risks associated with the increase in loan lending, including credit risk and the risk of default, which refers to the borrower's inability to repay the loan on time. The creditor would profit from the borrowed funds if the debtor repays the loan [2]. However, the creditor loses both the invested money and its interest if the debtor defaults on the loan. As a result, creditors face the challenge of predicting the likelihood that a debtor will not be able to repay a loan. One of the main factors that contribute to financial instability is known to be credit risk [3]. Commercial banks try to minimize defaulting risks by assessing the borrower's ability to repay the loan and requesting collateral prior to loan supply because lending is seen by them as a high-risk activity in addition to a source of profit. Employing highly qualified professionals from commercial banks, this exercise verifies a candidate's worthiness for loan approval or rejection based on a variety of criteria and yields a numerical score [4].

With the development of technology, machine learning algorithms and neural networks were developed very recently to automatically predict an individual's credit score based on their historical data and separate credit defaulters from the crowd

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

before approving a loan [5]. Because it has such a significant impact on growth and profit, loan default prediction is one of the most significant and pressing issues that lending institutions like banks and other financial institutions face. Throughout the years, machine learning techniques were utilized to evaluate the historical data of a single individual to calculate and predict credit risk [6]. Financial institutions will benefit greatly from this study in assessing borrowers' creditworthiness and calculating risks by taking into account a variety of variables to predict borrowers' risk. However, a lot of people have trouble figuring out how much credit they can afford to repay. One of the most crucial aspects for banks and other financial institutions' profitability and continued operation in the highly competitive market is creditworthiness analysis [7]. They must establish precise lending criteria. In order to provide the necessary information about the structure of credit, borrowers, and payment methods, these criteria must be sufficient. To assess a person's creditworthiness and risk of default, numerical values known as credit scores have been assigned to them using machine learning [8]. Additionally, it has been utilized to foresee and assess the acknowledge risk related for a person by alluding to his/her verifiable information.

There are two goals for this study. Finding a classification solution that can accurately predict whether a borrower will be paid off or default is the first objective.

When evaluating loan applications, this will assist potential investors in determining whether borrowers are worthy of credit [9]. The second objective of this work is to discover and investigate relationships and associations among the attributes that contribute to the repayment. These relationships and associations can be used to uncover possibly hidden information that could be useful to potential investors [10]. We compared the performance of classifiers with two distinct datasets that shared the intention of predicting loan approval in order to locate classification models for the prediction of default: Perceptron, Stochastic Gradient Descent (SGD), Decision Tree (DT), Random Forest (RF), Extreme Gradient Boosting (XGB), LightGBM (LGBM), Gradient Boosting Classifier (GBC), Ridge, Bagging, Extra Tree (ET), AdaBoost, Voting (hard), and Voting (soft). We used precision, recall, the f1-score, the confusion matrix, and the ROC (AUC) curve to determine how well the classifiers performed.

Rest of the article organized as: The literature review is in section 2, the background of the experiment is in section 3, the experimental methodologies are in section 4, the experimental setup is in section 5, the result is in section 6, and the discussion and conclusion are in sections 7 and 8.

2. LITERATURE REVIEW

The application of various machine learning methods to the prediction of the success of the allocation of bank loans was the subject of several studies. The previous studies are presented below.

According to the research conducted by Addo et al. [11], two important key aspects in the management's processing when issuing the loan are the selection of the algorithms used to make a decision and the selection of variables to respond to business objectives.

Kurapati and Bhansali's research [12] demonstrates that the random forest algorithm is superior to other models like decision trees and gradient at identifying credit defaulters on loans and that a loan default prediction model can be used by as many people as possible if it has high accuracy.

Using data mining techniques, Jiang [13] investigated how to predict the success of bank telemarketing. The dataset came from the Machine Learning Repository at UCI. This data set contains 21 attributes and 4119 instances. SVM, LR, NB, a NN, and a DT were utilized by them. The LR algorithm proved to be the most accurate of the five. The logistic regression model had an accuracy of 92.03%.

Ilham et al. [14] claim that using methods from machine learning, proposed a model. They utilized various procedures: KNN, SVM, NN, and DT are all examples of logistic regression. The features of the dataset did not undergo any preprocessing; it makes use of a prepared dataset from the UCI repository directly. The metrics used to evaluate these models show that the SVM provides the most accurate foundation, which is 91.07 %.

Predicting Loan Approval Using Machine Learning Methods by [15] the primary objective of this paper is to determine whether it is safe to give a loan to a particular person. There are four parts to this paper: data collection, evaluation of the most

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

promising machine learning models based on the data, system training, and testing.

Exploring the use of a machine learning algorithm to predict the loan approval process, the author [16] argues that providing credit to businesses and individuals is inevitable for the smooth operation of expanding economies like India. It is extremely challenging for banks and NBFCs with limited capital to devise a standard resolution and safe procedure to lend money to its borrowers for their financial needs as an increasing number of customers apply for loans. Additionally, the stock price of NBFC inventories has decreased significantly in recent years. It has contributed to the spread of a contagion to other financial stocks, which has had a negative impact on the benchmark in recent times. In this paper, we try to reduce the risk of finding the right person who will be able to pay back the loan on time and keep the bank's nonperforming assets in check. This is accomplished by feeding a trained machine learning model with the customer's previous loan records, which could produce an accurate result. The great focal point of the paper is to decide if distributing the credit to a specific person will be protected. The sections of this paper are as follows: Data collection, data cleaning, and performance evaluation are all included. Evaluation found that the NB model performs better in experimental tests. In terms of loan forecasting, experimental tests revealed that the NB model performs better than other models.

Regarding Gautam et al. [17] they have used exploratory data analysis methods to handle the loan request or loan forecast. Two machine learning models—DT and RF—were utilized to address the issue of bank lending. The document can be expanded at a higher level in the future to make the program safer and more accurate. Finally, there were a number of computer failures, content faults, and the primary weight of the characteristics in the automated prediction system. Weight adjustments can be made safer, more dependable, and more dynamic by modifying the program. The prediction module may be incorporated into the automated processing system module in the future.

Vangaveeti et al. [18] proposed a machine learning technique known as supervised learning model based on the LR model. They were able to predict whether or not the loan would be approved by employing the LR model. Using the logistic regression model, they were able to predict whether or not the loan would be approved. The output was produced by putting these various input variables to use. The output of the software is binary, or both 0 and 1, in the event that it receives the input data. The loan is approved if the output is 1, and "1" is displayed. It is displayed. If the output is zero, "0" will be displayed, and the loan will not be approved. The credit forecast framework has been made to help firms to pick the proper decision to support or reject clients' advance demands that will without a doubt help the bank area construct effective stock channels. In this model, the system for calculated relapse is applied. Implementation and testing of the domain with various methods that performs better than standard data mining techniques.

Aphale, [19], predicted the value of consumers' credit by analyzing bank credit data using machine learning. They looked into the bank credit dataset using a variety of machine algorithms to determine which approaches were best. All of the other algorithms performed well in terms of accuracy and other performance measurement techniques, with the exception of the closest Centroid and Gaussian NB algorithms. Between 76% and more than 80% precision was achieved by each of these algorithms. Customers' credit value is also impacted by the most important characteristics. In the event that all characteristics are utilized, these most significant characteristics were used, and some of the performance accuracy was compared to the specified algorithms. The experimental results showed no difference in their prediction accuracy or any other measurement. Using linear regression, they developed a predictive model with the main characteristics for predicting credit value. Create an automated method for assessing bank risk by anticipating the most important features of a consumer's credit worth that will be included in their loan approval.

Patel et al. [20] used data mining techniques to predict individuals who might default on a home loan application dataset. Various approaches to forecasting loan defaults have been used. The best

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

results were achieved through the use of logistic regression, the random forest, gradient boosting, and cat boost classification. Gradient boosting, in contrast to logistic regression, produces results that are comparable or better.

Sheikh et al. [21] investigated the issue of loan default forecasts using logistic regression, a crucial method in predictive analytics. Clearing and handling the data, imputation of missing values, the construction of the data set, and model experimental analysis for model evaluation and testing of test data were the first steps in the prediction process. Analyses of the data were done. The data set's highest accuracy was 0.811 on the first informational collection. Due to the higher likelihood that the loan amount will not be reimbursed, candidates with the lowest loan score will not receive loan approvals after analyzing the following findings. It stands to reason that applicants with higher incomes and fewer loan requests are more likely to be approved and to pay back their debts.

3. BACKGROUND

Artificial Intelligence (AI) is a calculation that permits programming applications to turn out to be all the more precisely unsurprising without being unequivocally modified [22]. The idea that systems can learn from data, recognize patterns, and make decisions that lead to the best solutions with little human intervention is supported by a subset of artificial intelligence.

3.1. Machine Learning Classifiers

Data analytics technique known as machine learning teaches computers to do things that humans and animals do naturally: learn from past mistakes [23]. Without relying on a predetermined equation as a model, machine learning algorithms learn information directly from data using computational methods [24]. As more samples are made available for learning, the algorithms adjust their performance. AI utilizes two sorts of procedures: Unsupervised learning finds hidden patterns or intrinsic structures in input data, while supervised learning trains a model on known input and output data to predict future outputs. Predicting the class of a set of data points is known as classification [25]. Targets and categories are other names for classes. Classification is a type of supervised learning in which target labels accompany the input data [26]. Classification has numerous applications in numerous fields.

A few AI classifiers have been applied in our done trial like Perceptron, Stochastic Gradient Descent (SGD), Decision Tree (DT),Random Forest (RF), Extreme Gradient Boosting (XGB), LightGBM (LGBM), Gradient Boosting Classifier (GBC), Ridge, Bagging, Extra Tree (ET), AdaBoost, Voting (hard), and Voting (soft). Additionally, we have utilized two distinct loan approval datasets. The objective of using multiple classifiers is to determine which one performs better on which dataset. Figure 1 depicts the machine learning algorithms used in this experiment.

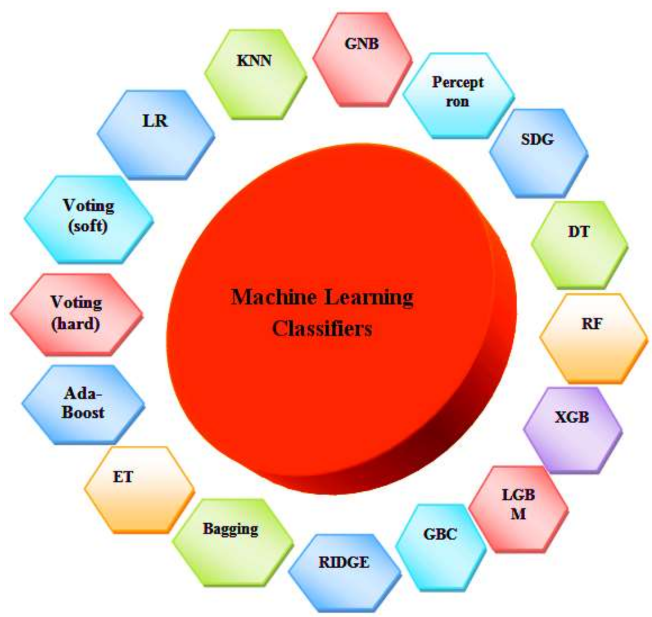

Figure 1: Experimental ML Classifiers

3.2. Accuracy of Classifiers

Because accuracy is such a well-known metric for assessing model performance in classification tasks, it is frequently used interchangeably with overall offline and online model performance [27]. Because it is one of the "if not the" easiest metrics to interpret and use in ML, accuracy gained this unique status. AI accuracy is the proportion of correctly classified predictions made by a trained machine learning model when divided by the total number of predictions made in each class. ACC is a common abbreviation for it [28]. ACC is measured using a scale that ranges from 0 to 100 or from 0 to 1 depending on the chosen scale. A classifier with accuracy of 0 always predicts the incorrect label, whereas one with accuracy of 1, or 100, always predicts the correct label. The fact that this metric is directly correlated with all confusion matrix values is a nice feature. These are the four mainstays of managed AI assessment: False negatives (FN), true positives (TP), false positives (FP), and true positives (TN). Accuracy is the ratio of the number of correct predictions to the total number of predictions. Right expectations are made out of TP and TN. All expectations are made out of the whole of positive (P) and negative (N) models. P is made up of TP and FP, while N is made up of TN and FN.

Therefore, accuracy may be defined as:

$$Accuracy = \frac{\text{Number of correct predictions}}{Total\ number\ of\ predictions} = \frac{TP+T}{TP+TN+FP+FN} \quad (1)$$

In addition, it is essential to emphasize that any metric evaluation of model accuracy should be based on a statistically significant number of predictions.

3.3. Classification Report (precision, recall and F1-score)

In machine learning, a metric for performance evaluation is a classification report. It is used to display the trained classification model's precision, recall, and F1 Score [29]. A classification-based machine learning model's performance evaluation metric is this one. The precision, recall, and F1 score of the model are shown. It provides a deeper comprehension of the trained model's overall performance.

The ratio of TP to the sum of true and FP is called precision.

$$Precision = \frac{TP}{TP+F} \quad (2)$$

The ratio of TP to the sum of TP and FN is called recall.

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

The precision and recall weighted harmonic mean is the F1. The model's expected performance is better the closer the F1 score value is to 1.0.

$$F1 = 2 \times \frac{precision \times recall}{precision + recall} \quad (4)$$

3.4. Confusion Matrix

There are multiple categorical outputs from a classification model. The majority of error measures will calculate our model's total error, but we are unable to identify individual errors in our model. Using a standard accuracy measure, we cannot see the model misclassifying more than one category. Also, let's say that the data show a significant class disparity. A model might be able to predict the majority class for all cases and have a high accuracy score in that scenario, which occurs when a class has more instances of data than the other classes; when it is not predicting the classes of the minority. Confusion matrices are useful in this situation. A confusion matrix facilitates visualization of the classification problem's outcomes by providing a table layout of the various prediction and result outcomes [30]. It displays a table of the classifier's predicted and actual values.

From a classifier's predicted and actual values, we can get four different combinations as in figure 2:

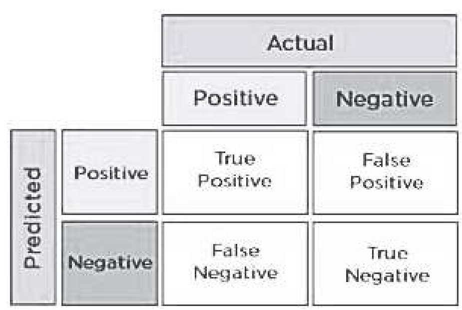

Figure2: Confusion Matrix

4. Feature Importance

The degree to which each feature aids in the prediction of the model is shown by the feature importance [31]. Essentially, it decides the level of convenience of a particular variable for an ongoing model and expectation. In general, we use a numerical value that we refer to as the score to represent the importance of features. The higher the

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

score value, the more significant a feature is. A feature importance score has numerous advantages. For instance, the relationship between independent and dependent variables can be determined. We would be able to identify and eliminate irrelevant features by analyzing variable importance scores. The model may run faster or even perform better if the number of not meaningful variables is decreased. Likewise, include significance is normally utilized as a device for ML model interpretability. It is possible to explain from the scores why the ML model makes particular predictions and how we can alter its predictions by manipulating features. There are many ways to determine a feature's importance, but the XGB and LGB classifiers are used in this manuscript to extract features.

3.5. ROC (AUC) Curve

The performance of the classification problems at various threshold settings is measured by the AUC-ROC curve. AUC is a measure of separability, whereas ROC is a probability curve [32]. It shows how well the model can differentiate between classes. The model is better at identifying 0 classes as 0 and 1 class as 1, the higher the AUC. By way of analogy, the model is better at identifying the occurrence of incident from non-occurrence it if the AUC is higher.

Plotting the ROC curve with TPR versus FPR is done with TPR on the y-axis and FPR on the x-axis in Figure 3.

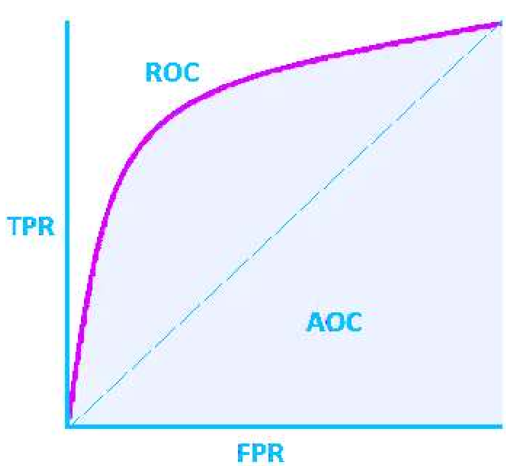

Figure3: Roc Curve

4. EXPERIMENTAL METHODOLOGY

For the purpose of analysis, loan datasets from various banks have been utilized. The named dataset-1 primarily contains categorical data, while the named dataset-2 primarily contains numerical data. In order to get the data ready for analysis, it is important to perform data preprocessing prior to analysis because good data can only lead to better outcomes [33]. The proposed system cleans imputations, normalizes, and transforms data during data preprocessing.

Null values and redundant attributes are removed from the dataset during the data cleaning process.

The proposed system employs Machine Learning Algorithms on this sampled data to determine which algorithm performs better and is suitable for prediction. This framework likewise thinks about the exactness of calculations in both datasets (dataset-1 and dataset-2) to choose the best calculation that predicts the qualifiers for credit endorsement in both datasets really. Figure 4 and the description of the proposed system's architecture can be found below.

The primary objective of this study is to select the most suitable model with high accuracy and low error.

Data Collection: The accuracy of our model is determined by the standard and quantity of the collected data (datasets 1 and 2). This step typically produces a representation of knowledge that will be used for training.

Exploratory Data Analysis (EDA): For dividing the data into training and evaluation sets and visually assisting in the detection of relevant relationships between variables, class imbalances, or other exploratory analyses.

Feature Engineering: Data engineering techniques like selecting relevant features, handling missing data, encoding the data, and normalizing it are all encapsulated in feature engineering.

Training and Testing: As a data analysis tool, Python will be used to clean the data and divide it into a training set and a test set, with the training set accounting for 80% of the data and the test set for 20%.

ML Classifiers: There are multiple algorithms used for various tasks.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Train the Model: Correctly answering a question or making a prediction is the objective of training.
Evaluate the Model: uses one or more metrics to objectively evaluate the model's performance.
Result Comparison: This errand of trial give an examination perspective on both datasets and anticipated score through calculations assists with picking ideal calculation to apply in certifiable element.

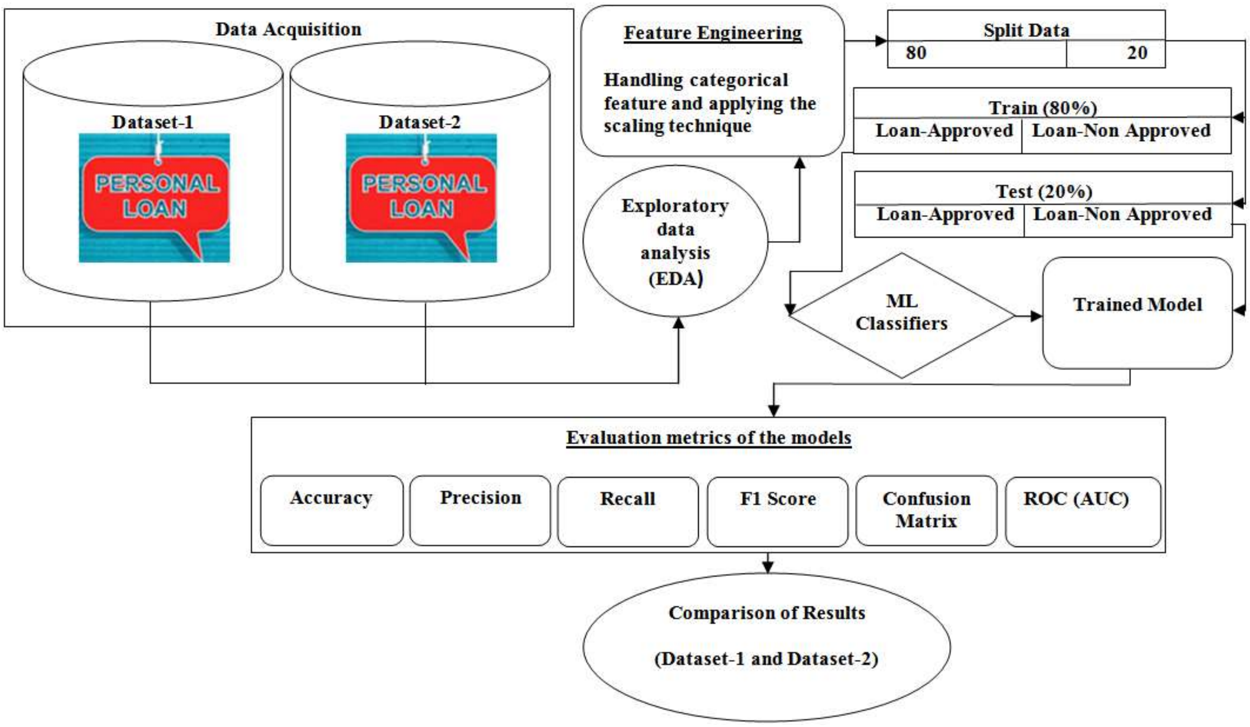

Figure 4: Architecture Of The Experiment

5. EXPERIMENTAL SETUP

The two dataset were used in this experiment, it contains all data for credit up-and-comers, and was used as discretionary data. In data analysis, a variety of machine learning techniques will be used to build models like base classifiers, ensemble classifiers, and voting (hard and soft) of classifiers.

Dataset-1 obtained from a hackathon, which was held by the "Univ. AI" and derived from the Kaggle data repository [34]. There are 13 attributes and 252000 instances in this dataset. Table 1 contains information about attributes, including the name attribute Id, CITY, STATE, which is irrelevant for analysis purposes. There are nine independent variables in the remaining ten attributes, one of which is the dependent variable “RISK_Flag”.

Table 1: Information About Attributes From Dataset-1

Attribute Name	Data Type	Description
Id	int	Customer id number
Income	int	Income of the user
Age	int	Age of the user
Experience	int	Professional experience of the user in years
Married/Single	string	Whether married or single
House_Ownership	string	Owned or rented or neither
Car_Ownership	string	Does the person own a car
Profession	string	Profession
CITY	string	City of residence
STATE	string	State of residence
CURRENT_JOB_YRS	int	Years of experience in the current job
CURRENT_HOUSE_YRS	int	Number of years in the current residence
Risk_Flag	int	Defaulted on a loan

Thera Bank provided dataset-2 (Bank_Personal_Loan_Modelling.xlsx), which was derived from the Kaggle data repository [35] in Table 2. There are 14 attributes and 5000 instances in this dataset. Twelve of these

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

attributes are used for additional analysis, but ID and ZIP are dropped because there is no mean for these attributes.

Table 2: Information About Attributes From Dataset-2

Feature Name	Feature Description
ID	Customer ID
Age	Age of the customer
Experience	Years of experience of customer has
Income	Annual Income of the customer
ZIP Code	Home Address ZIP code of the customer
Family	Number of family member of the customer
CCAvg	Avg. spending on credit cards per month
Education	Education level of the customer. 1→ Under Graduate 2→ Graduate 3→ Post Graduate
Mortgage	Value of House Mortgage
Securities Account	Does the customer have Security Account with bank or not?
CD Account	Does the customer have CD Account with bank or not?
Online	Does the customer have Online banking facility with bank or not?
CreditCard	Does the customer have a credit card issued by Bank or not?
Personal Loan	Target variable which indicates that the customer has token loan o

6. RESULT

The supervised learning methods were applied to the two customer loan-based datasets listed below. The test data set and the 10-fold cross validation were used to validate the classifier model and evaluate its accuracy as well as classifier's precision, recall, f1-score and confusion matrix.

6.1. Result based on dataset-1

- **Classifiers Accuracy, Classification Report and Confusion matrix**

The dataset-1 results are shown in Table 3, and the RF classifier has a significantly higher score than the other classifiers. For sure all classifiers have further developed accuracy over 85% however RF has accuracy 89.65% which is higher than any of different classifiers present in Table 3. In the process of analyzing the customers' bank loan dataset, ensemble classifiers and voting classifiers were also taken into consideration, but their accuracy had no effect on the final result. The classification report (precision, recall, and f1-score) and the score by confusion matrix also back up the RF score.

Table 3: Classification Report, Confusion Matrix, And Accuracy Of The Classifier From Dataset-1

Classifiers	Accuracy (%)	Classification Report			Confusion Matrix
			0	1	
LR	87.60	precision	0.875952	0.0	[[44148 0]
		recall	1.000000	0.0	[6252 0]]
		f1-score	0.933875	0.0	
KNN	88.97	precision	0.931749	0.560574	[[41638 2510]
		recall	0.943146	0.512156	[3050 3202]]
		f1-score	0.937413	0.535272	
GNB	87.60	precision	0.875952	0.0	[[44148 0]
		recall	1.000000	0.0	[6252 0]]
		f1-score	0.933875	0.0	
Perceptron	87.60	precision	0.875952	0.0	[[44148 0]
		recall	1.000000	0.0	[6252 0]]
		f1-score	0.933875	0.0	
SDG	87.60	precision	0.875952	0.0	[[44148 0]
		recall	1.000000	0.0	[6252 0]]
		f1-score	0.933875	0.0	
DT	88.15	precision	0.938686	0.520476	[[40846 3302]
		recall	0.925206	0.573257	[2668 3584]]
		f1-score	0.931898	0.545593	
RF	89.65	precision	0.935677	0.590663	[[41807 2341]
		recall	0.946974	0.540307	[2874 3378]]
		f1-score	0.941292	0.564364	
XGB	88.31	precision	0.888262	0.659312	[[43762 386]
		recall	0.991257	0.119482	[5505 747]]
		f1-score	0.936937	0.202302	

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

LGBM	87.68	precision recall f1-score	0.877357 0.998981 0.934227	0.659091 0.013916 0.027256	[[44103 45] [6165 87]]
GBC	87.61	precision recall f1-score	0.876119 0.999932 0.933940	0.769231 0.001599 0.003192	[[44145 3] [6242 10]]
Ridge	87.60	precision recall f1-score	0.875952 1.000000 0.933875	0.0 0.0 0.0	[[44148 0] [6252 0]]
Bagging	89.36	precision recall f1-score	0.934957 0.944233 0.939572	0.576539 0.536148 0.555611	[[41686 2462] [2900 3352]]
ET	89.46	precision recall f1-score	0.934099 0.946498 0.940258	0.583127 0.528471 0.554455	[[41786 2362] [2948 3304]]
AdaBoost	87.61	precision recall f1-score	0.876136 0.999932 0.933950	0.785714 0.001759 0.003511	[[44145 3] [6241 11]]
Voting(hard)	87.61	precision recall f1-score	0.876136 0.999932 0.933950	0.785714 0.001759 0.003511	[[44145 3] [6241 11]]
Voting(soft)	87.94	precision recall f1-score	0.880461 0.997848 0.935487	0.740437 0.043346 0.081898	[[44053 95] [5981 271]]

- **Feature Importance**

Measurement of feature importance is yet another method for comprehending the characteristics that better predict customer churn [36]. Out of nine independent features (Income, Age, Experience, Married/Single, House_Ownership, Car_Ownership, Profession, Current_Job_Years, and (Current_House_Years), neither the XGB nor the LGB classifiers take into account the customer's "profession." The features of income, age, experience, Current_Job_Years, Current_House_Years, Car_Ownership, married/single status, and House_Ownership are more relevant to the study of customer behavior who applied for a loan, in descending order of their score in dataset-1 (Figure 5).

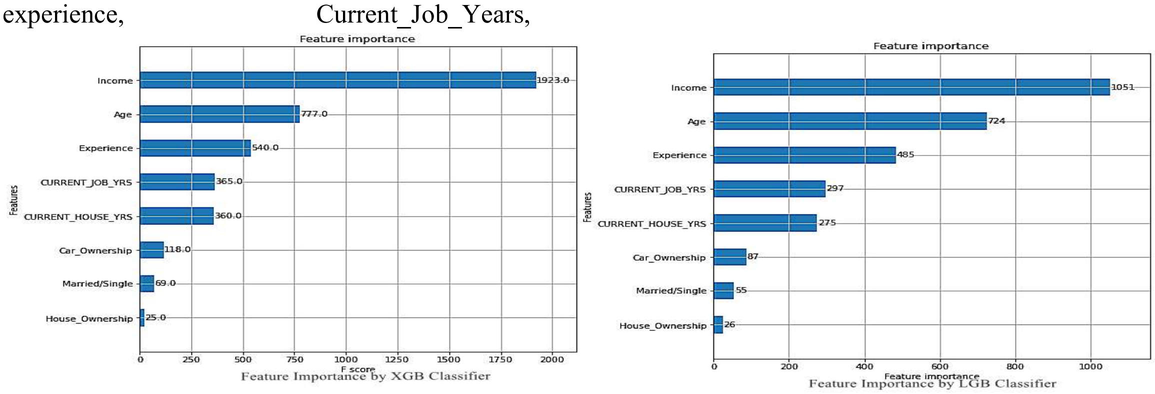

Figure5: Feature Importance By XGB And LGB Classifiers

- **ROC (AUC) Curve**

The AUC-ROC is the robust measure that we use as the optimizing performance metric because it is independent of the threshold used to determine the target class of instances [37]. Overall, the RF model brought about a higher ROC esteem. Although the ROC of the RF classifier is (0.936), it is statistically better than the ET classifier (0.934) in Figure 6.

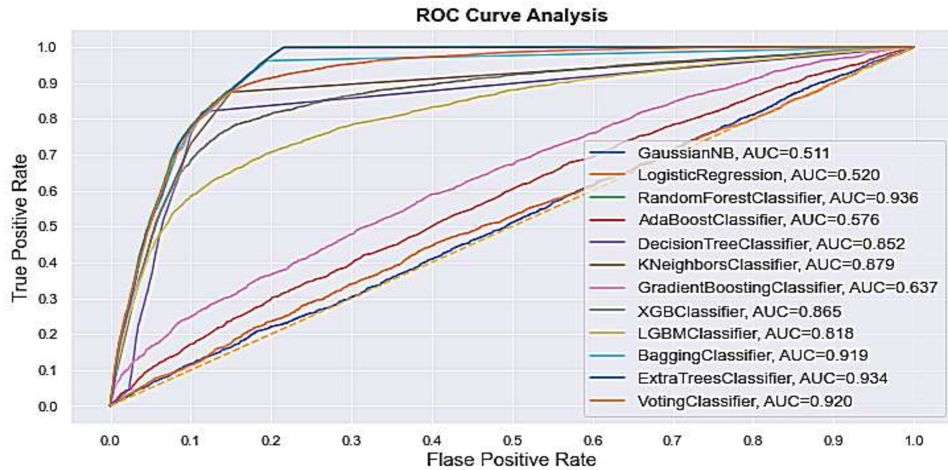

Figure 6: The ROC (AUC) Curve Of Classifiers

6.2. Result based on dataset-2

- **Classifiers Accuracy, Classification Report and Confusion matrix**

The accuracy score, which is a suitable metric for assessing model performance, is used to measure how well our proposed method performs [38]. On the other hand, the outcome that dataset-2 produced is shown in Table 4. As can be seen, the RF classifier prevails over all other classifiers once more. There is no other classifier that comes close to the RF classifier's accuracy of 99.20%. In addition, the confusion matrix score and the classification report (precision, recall, and f1-score) performed better than any of the other classifiers.

Table 4: Classification Report, Confusion Matrix, And Accuracy Of The Classifier From Dataset-2

Classifiers	Accuracy (%)	Classification Report			Confusion Matrix
			0	1	
LR	95.60	precision recall f1-score	0.966667 0.985746 0.976113	0.814286 0.647727 0.721519	[[899 13] [31 57]]
KNN	92.10	precision recall f1-score	0.944504 0.970395 0.957274	0.571429 0.409091 0.476821	[[885 27] [52 36]]
GNB	89.00	precision recall f1-score	0.952596 0.925439 0.938821	0.403509 0.522727 0.455446	[[844 68] [42 46]]
Perceptron	92.50	precision recall f1-score	0.935484 0.985746 0.959957	0.666667 0.295455 0.409449	[[899 13] [62 26]]
SDG	93.80	precision recall f1-score	0.951168 0.982456 0.966559	0.724138 0.477273 0.575342	[[896 16] [46 42]]

Journal of Theoretical and Applied Information Technology
15th October 2023. Vol.101. No 19

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

DT	98.00	precision	0.989035	0.886364	[[902 10]
		recall	0.989035	0.886364	[10 78]]
		f1-score	0.989035	0.886364	
RF	99.20	precision	0.992375	0.987805	[[911 1]
		recall	0.998904	0.920455	[7 81]]
		f1-score	0.995628	0.952941	
XGB	98.80	precision	0.992341	0.941860	[[907 5]
		recall	0.994518	0.920455	[7 81]]
		f1-score	0.993428	0.931034	
LGBM	98.90	precision	0.992350	0.952941	[[908 4]
		recall	0.995614	0.920455	[7 81]]
		f1-score	0.993979	0.936416	
GBC	98.70	precision	0.989119	0.962963	[[909 3]
		recall	0.996711	0.886364	[10 78]]
		f1-score	0.992900	0.923077	
Ridge	93.70	precision	0.938987	0.878788	[[908 4]
		recall	0.995614	0.329545	[59 29]]
		f1-score	0.966472	0.479339	
Bagging	98.60	precision	0.988043	0.962500	[[909 3]
		recall	0.996711	0.875000	[11 77]]
		f1-score	0.992358	0.916667	
ET	98.50	precision	0.985915	0.974026	[[910 2]
		recall	0.997807	0.852273	[13 75]]
		f1-score	0.991826	0.909091	
AdaBoost	97.20	precision	0.978355	0.894737	[[904 8]
		recall	0.991228	0.772727	[20 68]]
		f1-score	0.984749	0.829268	
Voting(hard)	98.40	precision	0.984848	0.973684	[[910 2]
		recall	0.997807	0.840909	[14 74]]
		f1-score	0.991285	0.902439	
Voting(soft)	98.00	precision	0.979570	0.985714	[[911 1]
		recall	0.998904	0.784091	[19 69]]
		f1-score	0.989142	0.873418	

- **Feature Importance**

XGB and LGB classifiers are used to extract significant attributes. In the following figure 7, 11 relevant attributes have been selected for each of the classifiers, all of which warrant further investigation. The XGB classifier selected the attributes Income, CCAvg, age, family, experience, education, mortgage, online, CD Account, CreditCard, and Securities Account in order of their decreasing score. The LGB classifier, on the other hand, selected the same important attributes as XGB, but scored “CD_Account” higher than “Online”.

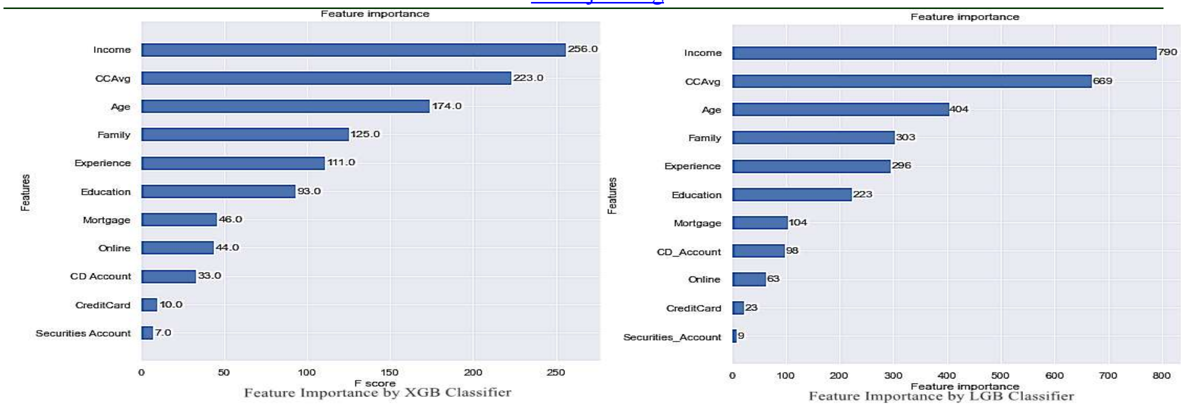

Figure 7: Feature Importance By XGB And LGB Classifiers

- **ROC (AUC) Curve**

Another metric used to determine the model's validation is the ROC (AUC) curve [39]. Figure 8 shows that the AUC scores of the GBC and LGBM classifiers are higher than any other classifiers. It took first place among the other classifiers we considered for our experiment with a score of 0.998, followed by RF and ET with scores of 0.995. The fact that the scores of all classifiers range from 0.900 to 0.998 indicates that they are all working in the right direction is a positive sign.

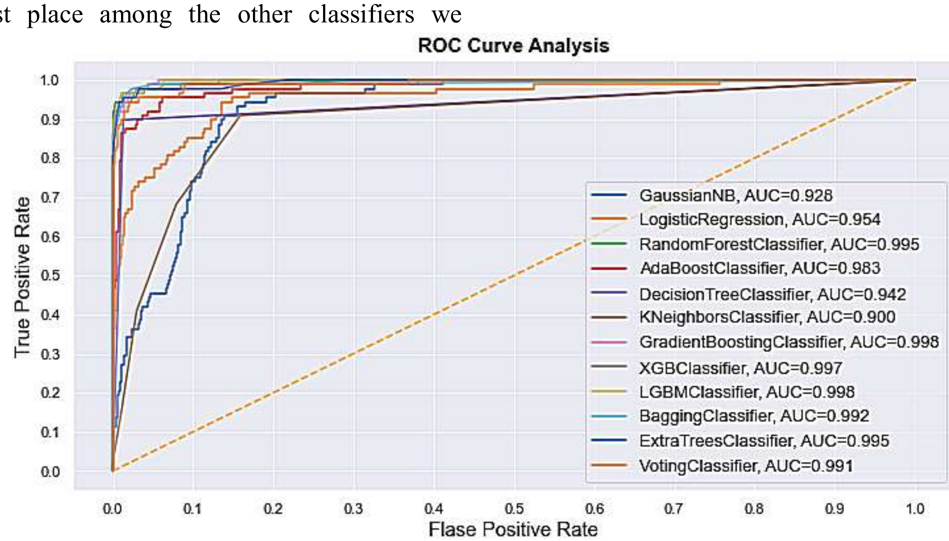

Figure 8: The ROC (AUC) Curve Of Classifiers

6.3. Comparison of Result (Dataset-1 and Dataset-2)

Table 5 and Figure 9 depict the comparison between the classifiers' results (accuracy) and avisual presentation. The RF classifier scored 89.65% in dataset-1 and 99.2% in dataset-2 for accuracy, respectively.

Journal of Theoretical and Applied Information Technology
15th October 2023. Vol.101. No 19

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 5: Classifier Accuracy Comparisons Between Datasets 1 And 2

Classifiers	Dataset-1 (% Accuracy)	Dataset-2 (% Accuracy)
LR	87.6	95.6
KNN	88.97	92.1
GNB	87.6	89
Perceptron	87.6	92.5
SDG	87.6	93.8
DT	88.15	98
RF	89.65	99.2
XGB	88.31	98.8
LGBM	87.68	98.9
GBC	87.61	98.7
Ridge	87.6	93.7
Bagging	89.36	98.6
ET	89.46	98.5
AdaBoost	87.61	97.2
Voting(hard)	87.61	98.4
Voting(soft)	87.94	98

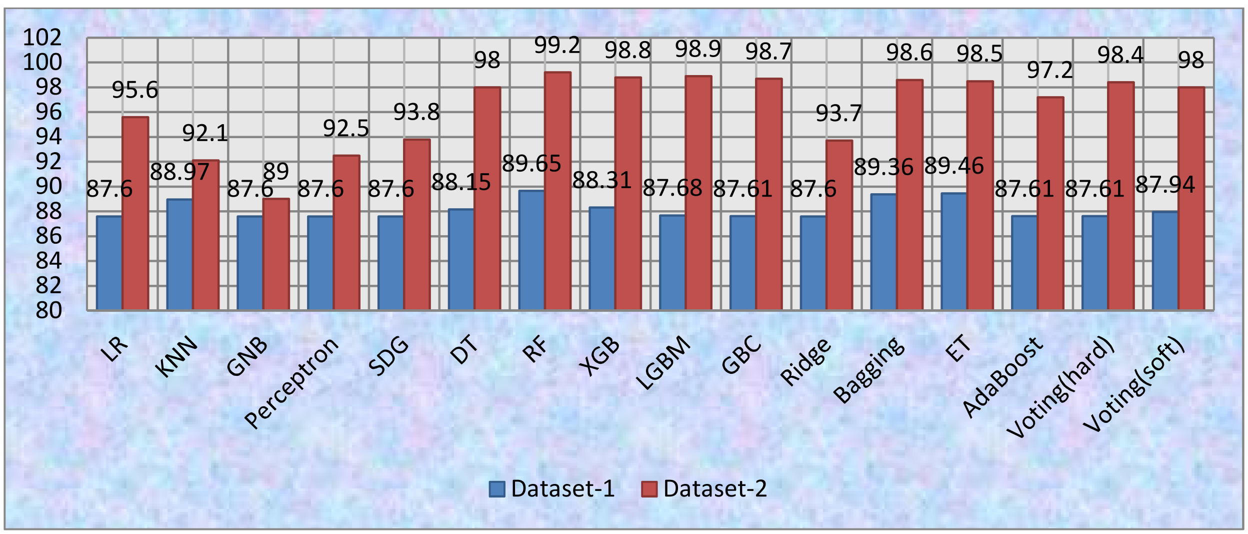

Figure 9: Diagrammatic Presentation Of Comparison Of Classifier's Accuracy

7. DISCUSSION

The usefulness of AI models and correlation in surveying loan endorsement in two bank crediting datasets were evaluated in this study.

In loan-credit, there is typically no central customer credit database and very little to none of a customer's credit history is available; this is the most common scenario. Because of this, it is challenging for banks to determine who should be denied bank loans. This paper demonstrates that machine learning algorithms are effective at extracting hidden information from the data set, which aids in assessing credit defaults, to overcome the drawback. The validation and test set served as the foundation for all of the performance metrics used in this paper. Several machine learning models were applied to the data set, but only those with an overall accuracy of 85 percent or higher on the validation set were included in this paper. The models in this paper are base classifiers, ensemble classifiers, and voting classifiers. This may be because our data set contains a lot of categorical features, and these classifiers have been shown to

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

generally perform better than other machine learning algorithms with such data sets. In both datasets (datasets 1 and 2), as well as in all ensemble classifiers (XGB, LGBM, GBC, Bagging, ET, and Adaboost) and voting (hard and soft) algorithms, the Random Forest (RF) model outperforms all other models reported in this paper. In a bank-credit environment, RF algorithms may be effective at predicting loan approval. On the validation set, the overall accuracy of all ensemble classifiers was at least 98%. As shown in Sections 6–6.3, the ensemble classifiers' ability to predict loan approval in bank-credit was also demonstrated by other performance measures. We used multiclass classification algorithms because they give us the additional advantage of having the average risk class, allowing us to further investigate customers who are predicted to be in that class before deciding whether to give them loans or not.

The cutting edge examination has been made in the accompanying Table 6 which has been directed by a few specialists in ongoing day. Predictions are made by collecting bank datasets containing various attributes and instances. These datasets have been used to predict a bank loan using a variety of methods and classifiers. The number of scores generated by various classifiers varies, but our proposed methods outperform those of other researchers and achieve the highest accuracy (99.2%) of any random forest classifier. The RF classifier outperformed several other classifiers in both dataset analysis and performance.

Table 6: Comparison Of State Of The Art Research With Proposed Method

Authors	Dataset Used	Dataset Attributes	Classifier/Algorithms	Classifier's Accuracy (%)
Proposed method	Hackathon organized by "Univ.AI" and Kaggle (Thera Bank)	(Id, Income, Age, Experience, Married/Single, House_Ownership, Car_Ownership, Profession, CITY, STATE, CURRENT_JOB_YRS, CURRENT_HOUSE_YRS, Risk_Flag) and (ID, Age, Experience, Income, ZIP Code, Family, CCAvg, Education, Mortgage, Securities Account, CD Account, Online, CreditCard, Personal Loan)	LR, KNN, GNB, Perceptron, SDG, DT, RF, XGB, LGBM, GBC, Ridge, Bagging, ET, AdaBoost, Voting(hard), Voting(soft)	Dataset-1(RF)-89.65 Dataset-2 (RF)-**99.2**
Anand et. al [40]	Kaggle	Age, Educational Background Category, Employment Status(or Years of Experience), Address – Demographic Area converted to Numeric Equivalent, Income, Debt Income, Credit to Debt Rasio, Other Debt.	Multiple Logistic Regression, Decision Tree, Random Forests, Gaussian Naive Bayes, Support Vector Machines, and other ensemble methods	ET- 86.17, RF- 85.55, CatBoost-84.92, LGB-84.49, ExGB-83.87
Ali et. al [41]	University of Tennessee	Borrower's age, Value of loan (USD), Ratio of loan to home purchase price, Borrower's credit score, First time home buyer? (Y/N), Borrower's total monthly debt expense, Borrower's total monthly income, Appraised value of home at origination, Purchase price for house, Borrower debt to income ratio, Current loan status	Binary Logistic Regression, ANN-MLP	95 98
Rath et. al [42]	UCI machine repository	Applicant income, Co-applicant income, Loan amount, Credit history, Married, Education, Dependents	Logistic regression, Decision tree, SVM	79, 72, 64
Lemos	Brazilian	Attribute information link:	Random forests,	DT- 78.2, Knn-

et. al [43]	financial institution	https://link.springer.com/article/10.1007/s00521-022-07067-x/tables/1	Decision trees, k-nearest neighbors, Elastic net, Logistic regression, and Support vector machines	77.9, Elastic net-76.2, LR- 76.2, Svm- 80.3, RF-82.8
Khatir et. al [44]	UCI machine learning repository (German credit dataset)	chk acct, duration, credit his, purpose, amount, saving acct, present emp, installment rate, sex, present resid, property, age, other nstall, housing, n credit, job, n people, telephone, foreign, response	random forests, random forest recursive feature elimination and random oversampling	RF-RandOverSampling 0.9846 NN-Imbalanced 0.9724 LR-Imbalanced 0.9724

8. CONCLUSION

Machine learning and ensemble methods were used in this study to evaluate individuals' credit risk performance in a banking environment. Machine and ensemble models, which have been widely used and produce prediction results with greater precision, have taken the place of traditional methods that use models like linear regression to estimate with reasonable accuracy in today's world. We used data from two banks to compare the accuracy of various machine learning algorithms by conducting in-depth experimental analysis and categorizing loan requests into approved and unapproved categories.

Even in the absence of a central credit database and/or credit history, the analytical results showed that machine learning algorithms can be used to model credit risk in a banking environment. In general, the RF machine learning algorithm has performed better with our data than other algorithms, and ensemble classifiers consistently outperform random forest models. Bajari et al. [45]; Carbo-Valverde et al. [46]; Fernández-Delgado et al. [47] found that the most accurate prediction was made by the RF classifier. Prediction accuracy is comparable between random forest classifiers, as shown by our investigation of a specific data set. Generally prediction accuracy is no less than 90% (on the validation set) in ensemble classifiers on these datasets are extremely great. Mathematical elements for the most part have displayed to have higher relative significance while anticipating default on credits than features. In addition, "Income" and "Age" have been identified as among the top two most significant features for predicting loan approval in both datasets, and this will be one of our subsequent areas of investigation: to devise a strategy for predatory lending in a credit environment. In addition, lending institutions, even in developing nations, can easily adapt the algorithms used in our paper for credit scoring because they are more cost-effective to implement.

Like any other study, this one had its limitations. In future works, our experimental analysis will be based on a larger data set, even though our work focused on using data from two banks from a lending institution. Our results, the particular selection of features, etc., allow us to draw some general qualitative inferences regarding the significance of various features and the application of ensemble classifiers in lending scenarios may not be all around pertinent across different nations and different establishments. The utilization of a broad informational collection could support the model's exhibition and give more exact assessments. In a similar vein, by comprehending the limitations of machine learning algorithms, we might be able to control the number of outliers more effectively. Another promising area of research for the foreseeable future is incorporating the temporal aspects of credit risk.

REFERENCES:

[1] SINGH, S. K., BASUKI, B., & SETIAWAN, R. (2021). The effect of non-performing loan on profitability: Empirical evidence from Nepalese commercial banks. The Journal of Asian Finance, Economics and Business, 8(4), 709-716.
[2] Krugman, P. (1988). Financing vs. forgiving a debt overhang. Journal of development Economics, 29(3), 253-268.
[3] Moudud-Ul-Huq, S., Akter, R., & Biswas, T. (2020). Impact of financial crisis on credit risk:

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Pre-and post-financial crises in an emerging economy. FIIB Business Review, 9(2), 118-132.

[4] Aslam, U., Tariq Aziz, H. I., Sohail, A., & Batcha, N. K. (2019). An empirical study on loan default prediction models. Journal of Computational and Theoretical Nanoscience, 16(8), 3483-3488.

[5] Madaan, M., Kumar, A., Keshri, C., Jain, R., & Nagrath, P. (2021). 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.

[6] Khandani, A. E., Kim, A. J., & Lo, A. W. (2010). Consumer credit-risk models via machine-learning algorithms. Journal of Banking & Finance, 34(11), 2767-2787.

[7] Curak, M., Poposki, K., & Pepur, S. (2012). Profitability determinants of the Macedonian banking sector in changing environment. Procedia-Social and Behavioral Sciences, 44, 406-416.

[8] Dastile, X., Celik, T., & Potsane, M. (2020). Statistical and machine learning models in credit scoring: A systematic literature survey. Applied Soft Computing, 91, 106263.

[9] Gonzalez, L., & Loureiro, Y. K. (2014). When can a photo increase credit? The impact of lender and borrower profiles on online peer-to-peer loans. Journal of Behavioral and Experimental Finance, 2, 44-58.

[10] Koh, H. C., & Tan, G. (2011). Data mining applications in healthcare. Journal of healthcare information management, 19(2), 65.

[11] Addo, M. P., Guegan, D., Hassani, B. (2018). Credit Risk Analysis Using Machine and Deep Learning Models.

[12] Kurapati, N., Bhansali, P.K. (2018). Predicting the credit defaulters using machine learning techniques. Int. J. Manag. Technol. Eng. 8(11), 6 (2018)

[13] Jiang, Y. (2018). Using Logistic Regression Model to Predict the Success of Bank Telemarketing. International Journal on Data Science and Technology, 4(1), 35.

[14] Ilham, A., Khikmah, L., & Iswara, I. B. A. I. (2019). Long-term deposits prediction: a comparative framework of classification model for predict the success of bank telemarketing. In Journal of Physics: Conference Series, 1175 (1), 1-6.

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

[16] Blessie, E. C., & Rekha, R. (2019). Exploring the Machine Learning Algorithm for Prediction the Loan Sanctioning Process. 1, 2714–2719. https://doi.org/10.35940/ijitee.A4881.119119

[17] Gautam, K., Singh, A. P., Tyagi, K., & Kumar, S. (2020). Loan Prediction using Decision Tree and Random Forest. 853–856.

[18] Vangaveeti, S. A., Venna, N. L., Naga, P., Ramyayajamanam, S., Marni, H., Satish, N., & Maganti, K. (n.d.). Logistic regression-based loan approval prediction. XIII(V), 319–325.

[19] Aphale, A. S. (2020). Predict Loan Approval In Banking System Machine Learning Approach for Cooperative Banks Loan Approval. 9(8), 991–995.

[20] Patel, B., Patil, H., Hembram, J., & Jaswal, S. (2020). Loan Default Forecasting using Dat2a Mining. 7–10.

[21] Sheikh, M. A., Goel, A. K., & Kumar, T. (2020). An Approach for Prediction of Loan Approval using Machine Learning Algorithm. Icesc, 490–494.

[22] Gupta, A., Gupta, R., & Kukreja, G. (2021). Cyber security using machine learning: techniques and business applications. Applications of Artificial Intelligence in Business, Education and Healthcare, 385-406.

[23] Müller, A. C., & Guido, S. (2016). Introduction to machine learning with Python: a guide for data scientists. " O'Reilly Media, Inc.".

[24] Muzumdar, P., Basyal, G. P., & Vyas, P. (2022). An empirical comparison of machine learning models for student's mental health illness assessment. arXiv preprint arXiv:2202.13495.

[25] Aggarwal, C. C., & Aggarwal, C. C. (2015). Data classification (pp. 285-344). Springer International Publishing.

[26] Sarker, I. H. (2021). Machine learning: Algorithms, real-world applications and research directions. SN computer science, 2(3), 160.

[27] Bahri, M., Bifet, A., Gama, J., Gomes, H. M., & Maniu, S. (2021). Data stream analysis: Foundations, major tasks and tools. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 11(3), e1405.

[28] Kourou, K., Exarchos, T. P., Exarchos, K. P., Karamouzis, M. V., & Fotiadis, D. I. (2015). Machine learning applications in cancer prognosis and prediction. Computational and structural biotechnology journal, 13, 8-17.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[29] Chaurasia, A. and chaurasia, V. (2023). Ensemble technique to predict heart disease using machine learning classifiers. Network Biology, 13(1), 1.

[30] Nieto, Y., García-Díaz, V., Montenegro, C., & Crespo, R. G. (2019). Supporting academic decision making at higher educational institutions using machine learning-based algorithms. Soft Computing, 23, 4145-4153.

[31] Belete, D. M., & Huchaiah, M. D. (2022). Grid search in hyperparameter optimization of machine learning models for prediction of HIV/AIDS test results. International Journal of Computers and Applications, 44(9), 875-886.

[32] Chaurasia, V., & Chaurasia, A. (2023). Detection of Parkinson's Disease by Using Machine Learning Stacking and Ensemble Method. *Biomedical Materials & Devices*, 1-13.

[33] Kumar, S., & Toshniwal, D. (2015). A data mining framework to analyze road accident data. Journal of Big Data, 2(1), 1-18.

[34] Kaggle data repository: https://www.kaggle.com/datasets/subhamjain/loan-prediction-based-on-customer-behavior?select=Training+Data.csv. (Accessed on 31st March 2023).

[35] Kaggle data repository: https://www.kaggle.com/code/pritech/bank-personal-loan-modelling/notebook. (Accessed on 31st March 2023).

[36] Verbeke, W., Martens, D., Mues, C., & Baesens, B. (2011). Building comprehensible customer churn prediction models with advanced rule induction techniques. Expert systems with applications, 38(3), 2354-2364.

[37] Chaurasia, V., & Pal, S. (2020). Machine learning algorithms using binary classification and multi model ensemble techniques for skin diseases prediction. International Journal of Biomedical Engineering and Technology, 34(1), 57-74.

[38] Kynkäänniemi, T., Karras, T., Laine, S., Lehtinen, J., & Aila, T. (2019). Improved precision and recall metric for assessing generative models. Advances in Neural Information Processing Systems, 32.

[39] Sofaer, H. R., Hoeting, J. A., & Jarnevich, C. S. (2019). The area under the precision-recall curve as a performance metric for rare binary events. Methods in Ecology and Evolution, 10(4), 565-577.

[40] Anand, M., Velu, A., & Whig, P. (2022). Prediction of loan behaviour with machine learning models for secure banking. Journal of Computer Science and Engineering (JCSE), 3(1), 1-13.

[41] Ali, S. E. A., Rizvi, S. S. H., Fong-Woon, L., Rao, F. A., & Jan, A. A. (2021). Predicting delinquency on Mortgage loans: an exhaustive parametric comparison of machine learning techniques. International Journal of Industrial Engineering and Management, 12(1), 1.

[42] Rath, G. B., Das, D., & Acharya, B. (2021). 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.

[43] de Lima Lemos, R. A., Silva, T. C., & Tabak, B. M. (2022). Propension to customer churn in a financial institution: A machine learning approach. Neural Computing and Applications, 34(14), 11751-11768.

[44] Hussin Adam Khatir, A. A., & Bee, M. (2022). Machine Learning Models and Data-Balancing Techniques for Credit Scoring: What Is the Best Combination?. Risks, 10(9), 169.

[45] Bajari, P., Nekipelov, D., Ryan, S. P., & Yang, M. (2015). Machine learning methods for demand estimation. American Economic Review, 105(5), 481-485.

[46] Carbo-Valverde, S., Cuadros-Solas, P., & Rodríguez-Fernández, F. (2020). A machine learning approach to the digitalization of bank customers: Evidence from random and causal forests. Plos one, 15(10), e0240362.

[47] Fernández-Delgado, M., Cernadas, E., Barro, S., & Amorim, D. (2014). Do we need hundreds of classifiers to solve real world classification problems?. The journal of machine learning research, 15(1), 3133-3181.