Loan Approval Prediction Using Classification Algorithm

Mohammed Jasim, Birla institute of technology and science, Pilani, Dubai campus
Email : f20200051@dubai.bits-pilani.ac.in
Prithvi Vinod Nair, Birla institute of technology and science, Pilani, Dubai campus
Email : f20200231@dubai.bits-pilani.ac.in
Haneen Noushad, Birla institute of technology and science, Pilani, Dubai campus
Email : f20200218@dubai.bits-pilani.ac.in

Abstract

For financial organizations and lenders, predicting loan approval is a critical responsibility in risk management and loss minimization. Machine learning algorithms have become very popular in recent years and are frequently used to forecast loan approvals based on historical data. This task's objective is to correctly categorize requests for loans as approved or rejected based on a variety of factors, including income, credit score, loan amount, employment history, and more.

Pre-processing of data, feature engineering, model selection, and evaluation are typically steps in the method of loan approval prediction. For predicting loan acceptance, several machine learning methods have been utilized, including logistic regression, decision trees, random forests, and neural networks. The calibre and volume of the data utilized for training and testing these algorithms affects how well they work.

Accurate loan approval prediction can give lenders possibilities to acquire financing while also assisting them in identifying dangerous applicants and preventing losses. By offering a more unbiased and based on information method, it can also aid in lowering prejudice in financing choices.

Keywords

Statistical Analysis - the method of gathering and studying massive amounts of data in order to spot trends and get insightful knowledge.

Regularization - refers to methods for calibrating machine learning models that aim to reduce the adjusted loss function and avoid either over or under fitting.

Multicollinearity - a statistical idea in which a model's several independent variables are correlated.

Variable selection - To pick acceptable variables from a comprehensive list of variables by deleting those that are unnecessary or redundant, or to choose among numerous variables which to include in a certain model.

High Dimensional data- High-dimensional data are those where the number of features (observed variables), p, is near to or greater than the number of observations n.

Statistical Modelling-The mathematical connection between random and non-random variables is how it is commonly defined.

Shrinkage Method-Some of the parameters are intended to be shrunk to zero.
Computational Algorithms-A computational algorithm comprises several control parameters that are deterministic in the first part of the algorithm, constitute a fixed component of the program, completely determine the computational process, and guarantee the first part of the algorithm is adaptable to a specific machine.

Data Visualization-Data visualization is the graphical representation of information and data.

Predictive Modeling- Data-mining technology called predictive modeling solutions creates a model by studying past and present data and using it to forecast future results.

Decision Tree- It is a hierarchical model in machine learning that uses a tree-like structure to represent a sequence of decisions and their possible consequences, leading to a final prediction or decision.

Random Forest- It is an ensemble machine learning algorithm that combines multiple decision trees to improve the accuracy and stability of predictions.

Introduction

Applying machine learning, loan approval prediction entails creating a classification model that can correctly categorize loan applications as accepted or refused based on a variety of input features. These attributes can be pre-processed and tailored to extract the data that is most closely related to judgments about loan acceptance.

The selection of algorithm depends on the needs of the task, such as accuracy, interpretability, and speed. Each algorithm has strengths and disadvantages of its own.

Several performance metrics, including accuracy, precision, recall, and F1-score, can be used to assess how well the machine learning model is performing. The task-specific needs, such as reducing false positives or false negatives, must be taken into consideration while selecting an assessment metric. To check that the algorithm can generalize to new data, it can also be tested using a holdout dataset, a portion of the data that was not utilized during training.

Machine learning has several benefits over conventional methods for predicting loan acceptance, including increased accuracy and less bias. Machine learning models can discover patterns and connections in the data that conventional statistical methods might miss. They can also deal with complicated interactions between parameters and enormous amounts of data. This can lower the risk of failure or scam and help lenders make more informed decisions about loan approvals.

Accurate loan approval prediction can give lenders possibilities to acquire financing while also assisting them in identifying dangerous applicants and preventing losses. By offering a more unbiased and based on information method, it can also aid in lowering prejudice in financing choices.

Literature Survey

Reference	Objectives	Problem Statement	Dataset	Algorithm	Percent measure value
1.	The most reason of this plan is to guess. which clients will be reimbursed with a advance because the moneylender should expect the issue that the borrower won't be appropriate to reimburse the danger.	A calculated relapse demonstrate has been executed. The most elevated exactness gotten with the unique dataset is 0.811. Models are compared based on execution estimations such as affectability and specificity. As a result of dissecting, the taking after conclusions	Fannie Mae Single-Family Loan Performance Data.	Logistic Regression using stratified k-folds cross validation and Random Forest.	75.08
2.	The proposed model introduces random forest machine learning model, it is really wanting to transfigure every one of the straight out factors into figures. That	A Credit isn't limited to a single sort since it covers all sorts of scenarios empowering by and large scope for the expectation of the securing of advances. Each method has	Personal dataset with 1000 entries.	Random forest algorithm	accuracy rate ranging from 76% to much more above 80%

	involving the LabelEncoder in presenting the wholeness of the factors have figures that our models can comprehend.	numerous steps and a diagram for the endorsement of advances. The primary step begins with the client. In-order for the advance to urge prepared, the client must apply for the advance sort basing on the necessity.			
3.	The interest rate, together with other factors (such as the payback period), assesses the borrower's riskiness; the higher the rate of interest, the riskier the consumer. depending upon the interest rate, we will determine if the applicant is eligible for the loan.	The consumer first qualifies for a house loan, and the firm then verifies the customer's loan eligibility. The firm seeks to automate (in real-time) the loan qualifying procedure based on information supplied by customers while completing out online application forms.	In our feed forward Neural Network model, the dataset that we used consist of 5000 clients or the consumers (personal dataset)	Logistic Regression, SVM, J48, KNN, and Tree Model	The neural network model's accuracy rate within the normal cutoff hit 98 percent.
4.	The authors aimed to show the various ML algorithms utilized by	In their research, deployed various ensemble ML techniques	'Loan Eligible Dataset,' available on Kaggle	Random Forest, Gradient Boost, Decision Tree, Support	accuracy ranged from 80% to 76% respectively

	researchers for credit assessment of rural borrowers, especially those with inadequate loan history. Their finding showed that the ML algorithms we utilized in this research were widely used and showed great results.	such as AdaBoost, LogitBoost, Bagging, and Random Forest model to predict loan approval of bank direct marketing data.		Vector Machine, K-Nearest Neighbor, and Logistic Regression	
5.	This paper aims to detect users generating spam reviews or review spammers. We identify several characteristic behaviors of review spammers and model these behaviors so as to detect the spammers.	In the proposed system, each review goes through a tokenization process first. Then, unnecessary words are removed and candidate feature words are generated. Each candidate feature words are checked against the dictionary and if its entry is available in dictionary.	Amazon review dataset	TF-IDF, SVM Algorithm, Naïve Bayes Algorithm, Decision Tree Algorithm.	prediction using logistic regression accuracy is 82.70%
6.	build an ML application which can reduce the	by calculating the possibility of loan escape, the ideal	Kaggle dataset with 1010 entries.	Logistic Regression, Decision tree, Support Vector	the best accuracy is achieved for the XGBoost

	time required to approve a loan using ML based prediction model to approve the loan with minimal human intervention by filtering huge number of applications and forward very few applications for human verification.	customers to target for loan giving will be easily identified using a XGBoost model approach.		machine, Random Forest tree and XGB	model which is 0.82 .
7.	we strive to limit the uncertainty in the back of opting the authentic individual so that we can minimize the Bank's Human Resource, can tightly closed Banks assets, can decrease the length of mortgage get sanction.	This proposed method is highly preferred among the image with dynamic variations. The technique used in the paper is evaluated using 4500 instance of the MRI and 3000 instance of CT.	Google's KAGGLE	classification, logic regression, Decision Tree and gradient boosting.	all models can achieve up to 80% accuracy.The highest accuracy is 93%
8.	Customer data is collected based on various banks and accessing the customer	Analyze the data and provide the results based on the customer profile to grant the loans using machine	Dataset from public repository	decision tree, random forest, support vector machine, k-nearest neighbor,	78 to 82 percent on forest model.

	profiles to analyze the data based on the parameters which are essential to integrate with the machine learning techniques	learning approach is most advanced than traditional loan approval-based systems		and decision tree with adaboost are implemented in this work	
9.	In this study we have applied logistic regression as a tool to predict whether an applicant is eligible for the loan or not. The data is collected from the Kaggle for studying and prediction.	The logistic regression approach is used to forecast the loan's safety. Logistic regression is a type of supervised machine learning algorithm which uses labelled dataset	Kaggle and includes applicants of various ages and genders. Education, marital status, income, and other characteristics are included in the dataset.	Supervised Learning, Unsupervised learning, Reinforce -ment Learning	75%
10.	The main objective of this paper is to predict whether a new applicant granted the loan or not using machine learning models trained on the historical data set.	The experimental results conclude that the accuracy of Decision Tree machine learning algorithm is better as compared to Logistic Regression and Random Forest machine learning approaches.	Kaggle dataset	SVM, LR, Random forest	85 % for random forest (highest)

11.	The study aimed at evaluating the performance of five classification models used in predicting loan approval.	study concluded that Naïve Bayes is the best algorithms for predicting loan approval.	The dataset used consist of 12 variables and 689 data instances	Random forest,naïve bayes, LR	Random forest-83.2% and 79.2% Logistic-81.6% and 73.7%
12.	main goal is to determine which machine algorithm performs best at predicting whether a person is qualified for a loan.	y mining the Big Data belonging to the previously loan issued individuals, and based on this data, machine learning models were taught to produce the most accurate results.	Thirteen aspects make up the dataset we're using: Loan ID, Gender, Married, Dependents, Education,etc.	Adaboost classifier, Passive Aggressive Classifier, Random forest	73% to 78%
13.	(ML) algorithms are employed to extract patterns from a common loan-approved dataset and predict deserving loan applicants. Customers' previous data will be used to	The training data set was provided to the machine learning model, and the model was trained using that data set. Every new applicant's information entered on the application form serves as a test data set. In this	Bank of Kigali	Random Forest, XGBoost, Adaboost, Lightgbm, Decision tree, and K-Nearest Neighbor	highest accuracy of 92%

	undertake the study, including their age, income type, loan annuity, last credit bureau report, Type of organization they work for, and length of employment	paper, they used three machine learning methods to predict client loan approval			
14.	The aim of this project is to provide a quick, immediate and easy way to choose the deserving applicants.	Loan Prediction System can automatically calculate the weight of each feature field taking part in loan processing and compares the new data with its associated weight	Personal dataset	Random Forest	83%
15.	This research is focusing on application of machine learning (ML) techniques to predict customer eligibility for a credit card.	Many researchers have conducted machine learning applications on credit scoring and customer default predictions. Researchers' have concluded that SVM	m UCI Repository credit card defaulter	SVM, ANN	90% correctly

		(support vector machine) and ANN (Artificial Neural Network) performed better than other classifiers			
16.	aim of the work is to evaluate the accuracy and cross validation in predicting loan approval	. A study compares the accuracy of various data mining classification algorithms in loan approval prediction. It is important to analyze and compare various classification algorithms that provide better cross validation.	Kaggle dataset	Logistic Regression (LR), Random Forest (RF), Decision Tree (DT) and K-Nearest Neighbor (KNN)	72% to 78% (LR highest)
17.	To evaluate if a loan application was eligible for one, financial firms used highly competent personnel in the past	To retain the trust of those clients, they may also introduce several unique deals. This study boost all clustering techniques To gain the trust of clients/	dataset from the UCI	Agglomative Hierarchical Clustering, Decision Tree, Random forest	84 to 95%

18.	To analyze the accuracy of Novel Random Forest (RF) and Linear Regression Algorithm (LR) algorithms used to approve bank loans	To overcome all the shortcomings of the existing algorithms, we compare and contrast the measures of accuracy of the Novel Random Forest (RF) over the traditional Linear Regression algorithm	credit card dataset downloaded from Kaggle	LR, Random forest	accuracy of RF is 70.5% which is higher than LR is 69.5%
19.	The main goal of this research is to create a proficient prediction of Supervised Machine learning algorithms for checking whether a person is eligible to get loan approval or not	implement the innovative approach of skillful Machine learning algorithms for innovative loan eligibility prediction and also to improve accuracy in existing algorithms like the Random Forest algorithm and Decision Tree algorithm	Kaggle dataset	Decision tree, Random forest	84 to 90% (RF highest)

20.	To implement machine learning techniques to enhance loan prediction technique.	The model created by the machine learning training process is employed while making the prediction. Loan prediction uses machine learning methods including Logistic Regression (LR), Random Forest (RF), and Decision Tree (DT).	k loan prediction system is drawn from the Kaggle	Logistic regression, Decision tree, Random forest	83 to 92%

Dataset Description

This dataset set is obtained from Kaggle. We have 615 different loan profiles as dataset and passing them over 10 different attributes and they are gender, married, dependents, education, self-employed, etc and loan statues is the target attribute.

Methodology

Loan approval predication is a classification problem. The model proposed in this study focuses on modelling implementation using classification techniques.

1. Classification Algorithm

   An example of a machine learning method that uses labelled data to predict the class of new, unlabelled data is a classification algorithm. They categorize data into one or more predetermined groups or classes using a collection of input features.Decision trees, support vector machines, k-nearest neighbours, and neural networks are a few examples of classification methods. Applications like picture recognition, spam detection, and fraud detection frequently use them.

   1.1 Random Forest Classifier

   A machine learning system called Random Forest constructs several decision trees and then aggregates the results to produce predictions. The final forecast is based on the consensus of the trees, each of which has been trained on a random selection of characteristics and data. For its high accuracy, resistance to noise and outliers, and capacity for handling sizable datasets, Random Forest is well-known. It frequently appears in applications like feature selection, regression, and classification.

   1.2 Support Vector Classifier

   A hyperplane is employed by the well-known classification technique Support Vector Classifier (SVC) to distinguish between several classes of data. It looks for the ideal hyperplane that minimizes classification mistakes while increasing the margin between classes. SVC can be expanded to handle several classes and is effective for data that can be separated in both linear and non-linear ways. Applications like text classification, bioinformatics, and picture classification frequently employ it.

1.3 KNN Classifier

The K-Nearest Neighbours (KNN) classifier is a non-parametric machine learning algorithm that predicts the class of a new sample based on the classes of its k-nearest neighbours in the training data. It is a simple and intuitive algorithm that is easy to implement and works well for low-dimensional data. However, it can be computationally expensive for large datasets and may suffer from the curse of dimensionality.

1.4 Decision Tree Classifier

A decision tree classifier is a kind of classification algorithm that model's decisions and potential outcomes using a tree-like structure to arrive at a final prediction or judgment. A hierarchical tree of decision rules is produced by recursively dividing the data depending on the most instructive feature at each node. The decision tree classifier can handle both numerical and categorical data and is straightforward and simple to understand. It is frequently employed in applications including customer segmentation, medical diagnosis, and credit scoring

Implementation

To build our algorithm, we must first understand the correlation between many user parameters caused by exercise. Since our datasets consist of numerous important variables and finally the calculated the loan status, it is challenging to identify patterns and correlations. We require the algorithm to be able to make decisions using those parameters. For this, a machine learning model is ideal. The significance of each characteristic will increase as more users train the machine learning model. Upon training, the algorithm will be able to predict the classification result based on the trained feature parameters.

Here, we are going to try and use four different Regression algorithms and analyse which among the four will give us the finest machine learning model. The four regression algorithms tested include.

1. Random forest classifiers
2. Support vector classifier
3. KNN classifier
4. Decision tree classifier

Steps to implement the model

1.Import required python libraries.
2. Import data set.
3.Drop unnecessary attributes.
4. Delete rows with null values.
5. Divide Dataset into X and Y.
6.Encode categorical values.
7.Split dataset into training and testing with 70% training.
8.Use various classifiers and predict the target attribute.
9.Find accuracy, recall, precision, f1-score and confusion matrix for each of the classifiers.

Results

DECISION TREE CLASSIFIER

```
[52] from sklearn.tree import DecisionTreeClassifier
     dt=DecisionTreeClassifier()
     dt.fit(X_train,Y_train)
     y_pred_dt = dt.predict(X_test)
     print("Accuracy of Decision Tree Classifier = ",metrics.accuracy_score(y_pred_dt,Y_test))
     print("Precision of Decision Tree  Classifier = ",metrics.precision_score(y_pred_dt,Y_test))
     print("Recall of Decision Tree Classifier = ",metrics.recall_score(y_pred_dt,Y_test))
     print("F1-Score of Decision Tree  Classifier = ",metrics.f1_score(y_pred_dt,Y_test))
     y_pred_dt
```

```
Accuracy of Decision Tree Classifier =  0.6319444444444444
Precision of Decision Tree  Classifier =  0.7052631578947368
Recall of Decision Tree Classifier =  0.7282608695652174
F1-Score of Decision Tree  Classifier =  0.7165775401069518
array([1, 0, 1, 0, 0, 0, 1, 1, 1, 0, 0, 1, 0, 1, 1, 1, 1, 1, 0, 1, 1, 1,
       1, 1, 0, 1, 1, 1, 1, 1, 0, 1, 1, 0, 1, 1, 1, 1, 1, 1, 0, 1, 1, 0, 0,
       1, 0, 1, 1, 0, 0, 1, 1, 1, 1, 0, 1, 0, 0, 1, 0, 1, 0, 1, 1, 1, 1,
       1, 1, 1, 0, 1, 1, 1, 0, 1, 0, 0, 1, 1, 0, 1, 1, 0, 1, 1, 1, 0, 1,
       0, 0, 0, 0, 1, 0, 1, 0, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 0, 1, 0, 1,
       1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 0, 0, 0, 0, 1, 0, 0, 0,
       1, 1, 0, 1, 0, 1, 1, 0, 1, 0, 0, 1])
```

CONFUSION MATRIX

```
[53] pd.crosstab(Y_test, y_pred_dt, rownames=[''], colnames=[''], margins=True)
```

	0	1	All
0	24	25	49
1	28	67	95
All	52	92	144

K-NEAREST NEIGHBORS CLASSIFIER

[50]
```
from sklearn.neighbors import KNeighborsClassifier
kn=KNeighborsClassifier(n_neighbors=10)
kn.fit(X_train,Y_train)
y_pred_kn = kn.predict(X_test)
print("Accuracy of K-Neighbors Classifier = ",metrics.accuracy_score(y_pred_kn,Y_test))
print("Precision of K-Neighbors Classifier = ",metrics.precision_score(y_pred_kn,Y_test))
print("Recall of K-Neighbors Classifier = ",metrics.recall_score(y_pred_rc,Y_test))
print("F1-Score of K-Neighbors Classifier = ",metrics.f1_score(y_pred_kn,Y_test))
y_pred_kn
```

```
Accuracy of K-Neighbors Classifier =  0.6527777777777778
Precision of K-Neighbors Classifier =  0.8631578947368421
Recall of K-Neighbors Classifier =  0.7652173913043478
F1-Score of K-Neighbors Classifier =  0.7663551401869159
array([1, 1, 0, 1, 0, 0, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
       1, 1, 1, 0, 1, 1, 1, 0, 1, 0, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1,
       1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 0, 1, 1, 0, 1, 1, 1, 1, 1, 1,
       1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1,
       1, 1, 0, 1, 1, 1, 0, 0, 1, 0, 0, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 0,
       1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 1, 1, 1, 1, 1, 1,
       1, 1, 1, 1, 1, 1, 0, 0, 1, 1, 1, 0])
```

CONFUSION MATRIX

```
pd.crosstab(Y_test, y_pred_kn, rownames=[''], colnames=[''], margins=True)
```

	0	1	All
0	12	37	49
1	13	82	95
All	25	119	144

SUPPORT VECTOR CLASSIFIER

[47]
```
from sklearn.svm import SVC
svc = SVC()
svc.fit(X_train,Y_train)
y_pred_svc = svc.predict(X_test)
print("Accuracy of Support Vector Classifier = ",metrics.accuracy_score(y_pred_svc,Y_test))
print("Precision of Support Vector Classifier = ",metrics.precision_score(y_pred_svc,Y_test))
print("Recall of Support Vector Classifier = ",metrics.recall_score(y_pred_svc,Y_test))
print("F1-Score of Support Vector Classifier = ",metrics.f1_score(y_pred_svc,Y_test))
y_pred_svc
```

```
Accuracy of Support Vector Classifier =  0.6597222222222222
Precision of Support Vector Classifier =  1.0
Recall of Support Vector Classifier =  0.6597222222222222
F1-Score of Support Vector Classifier =  0.7949790794979079
array([1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
       1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
       1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
       1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
       1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
       1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
       1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1])
```

CONFUSION MATRIX

```
[48] pd.crosstab(Y_test, y_pred_svc, rownames=[''], colnames=[''], margins=True)
```

	1	All
0	49	49
1	95	95
All	144	144

RANDOM FOREST CLASSIFIER

```
[46] from sklearn.ensemble import RandomForestClassifier
     rfc = RandomForestClassifier(n_estimators=80)
     rfc.fit(X_train,Y_train)
     y_pred_rc = rfc.predict(X_test)
     print("Accuracy of Random Forest Classifier = ",metrics.accuracy_score(y_pred_rc,Y_test))
     print("Precision of Random Forest Classifier = ",metrics.precision_score(y_pred_rc,Y_test))
     print("Recall of Random Forest Classifier = ",metrics.recall_score(y_pred_rc,Y_test))
     print("F1-Score of Random Forest Classifier = ",metrics.f1_score(y_pred_rc,Y_test))
     y_pred_rc
```

```
Accuracy of Random Forest Classifier =  0.7638888888888888
Precision of Random Forest Classifier =  0.9263157894736842
Recall of Random Forest Classifier =  0.7652173913043478
F1-Score of Random Forest Classifier =  0.8380952380952381
array([1, 1, 1, 1, 0, 0, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 0, 1, 1, 1,
       1, 1, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1,
       1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 1, 0, 0, 1, 1, 0,
       1, 1, 1, 1, 0, 1, 1, 0, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 0, 1,
       1, 0, 0, 0, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1,
       1, 1, 0, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 0, 1, 0, 1, 1, 1, 1,
       1, 1, 1, 1, 1, 0, 1, 1, 1, 0, 0, 1])
```

CONFUSION MATRIX

```
[ ] pd.crosstab(Y_test, y_pred_rc, rownames=[''], colnames=[''], margins=True)
```

	0	1	All
0	22	27	49
1	7	88	95
All	29	115	144

Conclusion

The algorithms we used on this model are random forest classifier, support vector classifier, KNN classifier and decision tree classifier out of which random forest classifier gave us the best accuracy. With ongoing advancements in technology and data analysis, loan approval prediction using machine learning is expected to continue to evolve and improve in the years to come.

References

[1]. *Loan Prediction System Using Machine Learning*
Anant Shinde , Yash Patil , Ishan Kotian , Abhinav Shinde and Reshma Gulwani.

[2]. *Machine Learning based Loan Eligibility Prediction using Random Forest Model*
Chenchireddygari Sudharshan Reddy, Adoni Salauddin Siddiq, N. Jayapandian.

[3]. *AN EXPLORATORY STUDY BASED ANALYSIS ON LOAN PREDICTION, ANKIT SHARMA*

[4]. *Machine Learning Models for Predicting Bank Loan Eligibility, Ugochukwu .E. Orji, Chikodili .H. Ugwuishiwu, Joseph. C. N. Nguemaleu, Peace. N. Ugwuanyi*

[5]. *LOAN APPROVAL PREDICTION Naga V Vara Prasad Mella , R . Rishitha Sai*

[6]. *AN APPROACH TO LOAN APPROVAL PREDICTION USING MACHINE LEARNING, B. Yamuna, B. Yamuna, D. Sai Nithin, E. Sri Ramya.*

[7]. *LOAN APPROVAL PREDICTION USING SUPERVISED MACHINE LEARNING ALGORITHM Dr.E.Neelima , Venkata Ayyappa Reddy Maruprolu , Sai Prasad Penta , Karthikeya Mallareddy*

[8]. *Customer Loan Eligibility Prediction using Machine Learning Algorithms in Banking Sector Ch. Naveen kumar , D. keerthana , M Kavitha , M Kalyani*

[9]. *Machine Learning Based Model for Prediction of Loan Approval, Bhanu Prakash Lohani, Mayank Trivedi, Ridhik Jeet Singh*

[10]. *PREDICTION OF MODERNIZED LOAN APPROVAL SYSTEM BASED ON MACHINE LEARNING APPROACH, Soumya Ranjan Jena, Vasantha S*

[11]. *AN IMPROVED PERFORMANCE EVALUATION OF CLASSIFICATION ALGORITHM FOR PREDICTION OF LOAN APPROVAL 1A.U. Farouk, A. 2Abdulkadir, A.Y. Abarshi*

[12]. *Customer Loan Eligibility Prediction using Machine Learning, Amjan Shaik, Kunduru Sai Asritha, Neelam Lahre*

[13]. *Predicting Bank Loan Eligibility Using Machine Learning Models and Comparison Analysis ,Miraz Al Mamun*

[14]. *EXPLORATORY DATA ANALYSIS FOR LOAN PREDICTION Vinuthna Ramavarapu, Bali Satya Durga, Basu Sindhu Devi, Chippada Satya Pravallika, Ms. D. Lalitha Kumari*

[15]. *Credit Card Approval Prediction by Using Machine Learning Techniques, M. P. C. Peiris*

[16]. *Prophecy of loan approval by comparing Decision Tree with Logistic Regression, Random Forest, KNN for better Accuracy. B. Aditya , V. Nagaraju*

[17]. *A Prediction Model for Bank Loans Using Agglomerative Hierarchical Clustering with Classification Approach Micheal Olaolu Arowolo , Oluwatosin Faith Adeniyi , Marion Olubunmi Adebiyi , Roseline Oluwaseun Ogundokun*

[18]. *A Novel Approach for Bank Loan Approval by Verifying Background Information of Customers through Credit Score and Analyze the Prediction Accuracy using Random Forest over Linear Regression Algorithm Ch.Venkata Sandeep , T. Devi*

[19]. *Higher Accuracy on Loan Eligibility Prediction using Random Forest Algorithm over Decision Tree Algorithm, Narra Rahul Kumar, L. Rama Parvathy*

[20]. *Dynamic Prediction of Loan Approval using Machine Learning Techniques ,Siva Adapa*

Contributions

Haneen: Dataset, Data pre-processing, SVC, ,KNN, result, implementation, methodology
Jasim: DT ,RF , 10 papers literature survey and reference, conclusion
Prithvi: evaluation metrics and confusion matrix, abstract, introduction, 10 papers literature survey and references