LOAN PREDICTION SYSTEM USING PYTHON

A project report
submitted in partial fulfillment of
the requirements for the award of the degree of
BACHELOR OF TECHNOLOGY

in

INFORMATION TECHNOLOGY

from

APJ Abdul Kalam Technological University

Submitted By

AYISHA NAILA (MEA20IT004)

LIYANA SHIRIN (MEA20IT009)

NIDHA (MEA20IT018)

NOUFAL IQBAL (MEA20IT020)

MEA ENGINEERING COLLEGE
DEPARTMENT OF INFORMATION TECHNOLOGY
VENGOOR P.O, PERINTHALMANNA, MALAPPURAM, KERALA-679325
AUGUST 2023

MEA ENGINEERING COLLEGE

PERINTHALMANNA-679325

DEPARTMENT INFORMATION TECHNOLOGY

CERTIFICATE

*This is to certify that the PROJECT report entitled "***LOAN PREDICTION SYSTEM USING PYTHON** *" is a bonafide record of the work done by* **AYISHA NAILA(MEA20IT004),LIYANA SHIRIN(MEA20IT009), NIDHA(MEA20IT018),NOUFA IQBAL(MEA20IT020)** *under our supervision and guidance. The report has been submitted in partial fulfillment of the requirement for award of the Degree of* **Bachelor of technology** *in* **INFORMATION TECHNOLOGY** *from APJ Abdul Kalam Technological University for the year 2023.*

Hanooja	**Deepa M**
Professor	*Head Of The Department*
Dept.of Information Technology	*Dept.of Information Technology*
MEA Engineering College	*MEA Engineering College*

ACKNOWLEDGEMENT

An endeavor over a long period may be successful only with advice and guidance of many well wishers. I take this opportunity to express my gratitude to all who encouraged me to complete this project. I would like to express my deep sense of gratitude to my resepcted **Principal Dr.G. Ramesh** for his inspiration and for creating an atmosphere in the college to do the project.

I would like to thank **Dr. Deepa M. , Head of the department, Information Technology** for providing permission and facilities to conduct the project in a systematic way. I am highly indebted to **Ms Hanooja T, Professor in information Technology** for guiding me and giving timely advices, suggestions and whole hearted moral support in the successful completion of this project.

My sincere thanks to project co-ordinator **Ms. Safna A.K , Associate Professor,Information Technology** wholehearted moral support in completion of this project.

Last but not least, I would like to thank all the teaching and non-teaching staff and my friends who have helped me in every possible way in the completion of my project.

DATE: Group.4

ABSTRACT

The concept and implementation of a Python-based loan prediction system are presented in this study. Financial organisations want more precise and effective tools to assess applicants' creditworthiness in light of the exponential growth of data. The project's objective is to use customer data to automate the loan qualifying procedure.

Using Python, a language recognised for its powerful data analysis and machine learning capabilities, we created a prediction model. The model makes use of a number of Python modules, including Scikit-Learn for implementing machine learning methods, NumPy for doing numerical computations, and Pandas for processing data. It employs a number of machine learning algorithms to forecast whether a loan should be given to an applicant using prior loan application data.

The model showed great accuracy in forecasting loan acceptance during testing using real-world data, enhancing the efficiency of financial institutions and their capacity to assess risk. The findings show that the predictive model may greatly shorten the time and resources needed for loan approval procedures and help to make lending judgements that are more well-informed and unbiased.

List of Abbreviations

SVM	**S**upport **v**ector **M**achine
ML	**M**achine **L**earning
NB	**N**aive **B**ayes
DT	**D**ecision **T**ree
KNN	**K** **N**earest **N**eighbour
CART	**C**lassification **A**nd **R**egression **T**ree
PCA	**P**rincipal **C**omponent **A**nalysis
ROC	**R**eceiver **O**perating **C**urve
AUC	**A**rea **U**nder **C**urve
MAE	**M**ean **A**bsolute **E**rror

CHAPTER 1

INTRODUCTION

As the data are increasing daily due to digitization in the banking sector, people want to apply for loans through the internet. Artificial intelligence (AI), as a typical method for information investigation, has gotten more consideration increasingly. Individuals of various businesses are utilizing AI calculations to take care of the issues dependent on their industry information. Banks are facing a significant problem in the approval of the loan. Daily there are so many applications that are challenging to manage by the bank employees, and also the chances of some mistakes are high. Most banks earn profit from the loan, but it is risky to choose deserving customers from the number of applications. One mistake can make a massive loss to a bank.

Loan distribution is the primary business of almost every bank. This project aims to provide a loan [1, 8] to a deserving applicant out of all applicants. An efficient and non-biased system that reduces the bank's time employs checking every applicant on a priority basis. The bank authorities complete all other customer's other formalities on time, which positively impacts the customers. The best part is that it is efficient for both banks and applicants. This system allows jumping on particular applications that deserve to be approved on a priority basis. There are some features for the prediction like- 'Gender', 'Married', 'Dependents', 'Education', 'Self employed', 'Applicant income', 'Loan Amount', 'Loan Amount Term','Credit History', 'Property area', 'Loan Status'.

1.1 Objective

The main objectives of this proposed technique are

- Develop a loan prediction model using Python.
- Use Historical loan application data for training and testing the model.
- Preprocess the data to handle missing values and outliers.
- Perform exploratory data analysis to understand data patterns and relationships.
- Extract relevant features like age, income, credit score, etc.
- Select appropriate machine learning algorithms for loan approval classification.
- Train the model on the training data and evaluate its performance using metrics like accuracy.
- Fine-tune the model's hyper parameters for optimal performance.
- Implement the trained model into a user-friendly system. Continuously monitor the model's real-world performance and update as needed.
- Continuously monitor the model's real-world performance and update as needed.

LITERATURE REVIEW

.

1.2 Bank Loan Prediction System using Machine Learning. [3]

Bank Loan prediction machine learning use a machine learning technique that will predict the person who is reliable for a loan, based on the previous record of the person whom the loan amount is accredited before. This work's primary objective is to predict whether the loan approval to a specific individual is safe or not. [3].
The use of machine learning techniques in the banking sector for loan approval can offer several benefits:

1.Efficiency: Machine learning algorithms can process large volumes of loan applications quickly and efficiently. This automation reduces the manual workload and speeds up the decision-making process, leading to faster loan approvals.

2.Risk Assessment: Machine learning models can analyze historical loan data and identify patterns related to creditworthiness and default rates. This helps banks assess the risk associated with each applicant accurately, leading to better-informed lending decisions.

3.Improved Accuracy: By leveraging historical data and patterns, machine learning models can make more accurate predictions about an applicant's likelihood of repaying the loan. This can lead to reduced instances of default and better loan portfolio management.

4.Consistency: Unlike human decision-makers, machine learning models are not influenced by emotions or biases. They consistently apply the same criteria to evaluate loan applications, ensuring fairness and objectivity in the process.

5.Cost Savings: Automating the loan approval process with machine learning can lead to cost savings for banks by reducing the need for manual review and approval, as well as minimizing the risk of bad loans.

6.Real-time Decisions: By integrating machine learning models into web applications or banking systems, loan approval decisions can be made in real-time, providing customers with instant feedback on their applications.

7.Personalization: Machine learning can also help banks tailor loan offerings to individual customers based on their credit history and financial behavior, creating a more personalized and customer-centric experience.

Overall, the use of machine learning in the banking sector for loan approval can lead to more efficient, accurate, and customer-friendly lending processes, benefiting both the financial institutions and their customers.

1.2.1 System Model

The specific features used in the machine learning model for loan approval are not explicitly mentioned. However, we can infer some potential features that are commonly used in such models based on the context of loan prediction. Here are some typical features that could be used:

1.Employment Status: Whether the applicant is employed, unemployed, self-employed, or retired can affect their loan repayment ability.

2.Debt-to-Income Ratio: This ratio represents the proportion of the applicant's income that goes towards debt repayment, indicating their financial health.

3.Loan Amount Requested: The amount requested by the applicant can influence the risk associated with the loan.

4.Loan Term: The duration of the loan also impacts the risk and repayment likelihood.

5.Number of Dependents: The number of dependents may affect the applicant's disposable income.

6.Marital Status: Marital status can provide insight into the applicant's financial stability and responsibilities.

7.Credit Card Utilization: The utilization of credit cards reflects the applicant's credit management practices.

8.Property Ownership: For secured loans, property ownership details are relevant.

1.2.2 Limitations

Here are some of the limitations of the project:

• Data Bias: The model's predictions may be influenced by biased historical data, leading to unfair or discriminatory lending decisions.

• Overfitting: Machine learning models may become overly complex and fit too closely to the training data, resulting in poor generalization to new loan applications.

• Model Interpretability: Some machine learning algorithms lack transparency, making it difficult to explain the reasons behind loan approval or rejection decisions. • Privacy and Security Concerns: Managing and securing sensitive customer data used for loan predictions can be challenging, raising privacy and security risks.

• High Development and Maintenance Costs: Building and maintaining a robust machine learning system requires skilled expertise and ongoing monitoring, leading to higher costs for financial institutions.

1.3 Hybrid ML Classifier for Loan Prediction System [4]

Banks make the majority of their income from loans. A lot of individuals apply for loans, and it is difficult to choose the real candidate who will repay the loan. A lot of misunderstandings may occur when selecting the real applicant when the process is done manually. As a result, a loan prediction system based on machine learning is developed, in which the system will automatically identify the qualified candidates. This is beneficial to both the bank personnel and the applicant.

The loan approval process will be greatly shortened. The loan data is predicted by using the hybrid model of Naive Bayes (NB) and Decision Tree (DT) algorithms. First, the dataset is given to the three classification algorithms– Support Vector Machine (SVM), NB and DT Algorithms and the prediction is done with these three algorithms. The accuracy of each of these three is used to assess performance. The creation of the hybrid model increases accuracy.

The dataset is given to NB for training and the prediction of NB is given to DT Algorithm for training. Test data are sent to the model for prediction after training. The model is evaluated, and the performance is measured in terms of different metrics form sklearn metrics. This prediction of loan range is useful for bank staff to give the loan amount accordingly. The NB algorithm checks for equality and independence of all the features in the dataset. In DT algorithm, the tree is constructed based on the information gain value. The attribute with high information gain value is placed as the root node and also the other nodes are constructed based on information gain value. The proposed hybrid model predicts - yes or no, and based on the prediction, whether the loan is to be sanctioned or denied for the applicant is specified.

1.3.1 Features

Here are certain key points that could be explored in the paper:

- Integration of Multiple Models: The paper could focus on combining different machine learning models or algorithms, such as decision trees, logistic regression, support vector machines, neural networks, etc., to create a more accurate and robust classifier.

- Ensemble Techniques: The use of ensemble methods like bagging, boosting, or stacking to improve loan prediction accuracy by combining the predictions of multiple base classifiers.

• Feature Selection and Engineering: Exploration of feature selection methods or feature engineering techniques to identify the most relevant and informative features for loan prediction, which can enhance the classifier's performance.

• Handling Imbalanced Data: If the loan dataset suffers from class imbalance (e.g., a significant difference in the number of approved and rejected loans), the paper could discuss techniques to address this issue and ensure balanced model performance.

• Performance Evaluation: Rigorous evaluation of the hybrid ML classifier's performance using appropriate metrics like accuracy, precision, recall, F1-score, and ROC-AUC, demonstrating its effectiveness in loan prediction. [4].

1.3.2 Limitations

Here are some of the limitations of the project:
While a "Hybrid ML Classifier for Loan Prediction System" can offer advantages, it also has some limitations that researchers and practitioners need to be aware of:

1.Complexity: The hybrid model may be more complex than individual models, making it harder to interpret and understand its decision-making process.

2.Training Time: Combining multiple models can increase the training time, especially if the dataset is large or if the models are computationally intensive.

3.Model Maintenance: As the hybrid model involves multiple components, updating and maintaining it over time may be more challenging compared to single-model systems.

4.Data Dependencies: The performance of the hybrid model heavily relies on the quality and representativeness of the data used for training. If the data is biased or incomplete, the model's predictions may suffer.

5.Overfitting Risk: Combining multiple models could increase the risk of overfitting if not properly tuned or regularized, leading to poor generalization to new, unseen data.

It's essential to consider these limitations while developing and deploying the hybrid ML classifier for a loan prediction system. Conducting thorough testing and evaluation, as well as implementing proper model monitoring and maintenance practices, can help mitigate these challenges. Additionally, selecting appropriate ensemble and hybridization strategies can help strike a balance between accuracy and complexity in the model. [4].

1.4 Prediction of loan status in commercial bank using machine learning classifier [5]

The objective of this paper is to create a credit scoring model for credit data. In this paper, we propose a machine learning classifier based analysis model for credit data. We use the combination of Min-Max normalization and K-Nearest Neighbor (K-NN) classifier. The objective is implemented using the software package R tool. This proposed model provides the important information with the highest accuracy. It is used to predict the loan status in commercial banks using machine learning classifier.

1.4.1 Features

Improved Loan Approval Process: Using a machine learning classifier for loan prediction streamlines the loan approval process in commercial banks. The model can quickly assess loan applications, making it faster and more efficient for applicants to receive decisions.

1. Higher Accuracy: Machine learning classifiers can analyze large datasets and identify complex patterns, leading to more accurate predictions of loan statuses. This reduces the risk of bad loans and helps banks make better lending decisions.

2. Data-Driven Decision Making: By leveraging historical loan data, the model makes decisions based on objective patterns and historical performance, reducing the influence of human bias and subjectivity.

3. Enhanced Customer Experience: Faster loan approvals and more accurate decision-making contribute to an improved customer experience, increasing customer satisfaction and loyalty.

4. Risk Management: Accurate prediction of loan statuses helps commercial banks better manage their risk exposure. It allows them to identify high-risk applicants and make informed decisions to mitigate potential losses.

5. Cost Savings: With a more efficient loan approval process and reduced risk of bad loans, commercial banks can save costs associated with manual underwriting and loan defaults.

6. Scalability: The machine learning classifier can handle a large volume of loan applications, making it scalable to meet the demands of a growing customer base.

1.4.2 Limitations

Here are the main limitations of the "Prediction of Loan Status in Commercial Banks using Machine Learning Classifier":

- Data Quality Dependency: The model's performance heavily relies on the quality and completeness of historical loan data. Inaccurate or biased data can lead to unreliable predictions.
- Limited Generalization: The model may not generalize well to new, unseen loan data or different economic conditions, potentially reducing its effectiveness in real-world scenarios.
- Model Interpretability: Some machine learning classifiers lack interpretability, making it difficult to understand and explain the reasoning behind the loan predictions, which may be a concern for stakeholders and regulators.
- Changing Loan Criteria: The model might not adapt quickly to changes in loan approval criteria, regulatory updates, or shifting market conditions, affecting its accuracy over time.
- Data Privacy and Security: Using sensitive financial data for model training raises privacy and security concerns. Safeguarding customer information is crucial to avoid potential breaches or misuse.

  Addressing these limitations requires rigorous data preparation, ongoing monitoring, model interpretability techniques, and regular updates to maintain the model's accuracy and compliance with changing conditions and regulations.

[5].

METHODOLOGY

1.5 Problem Statement

Develop a loan prediction system that uses historical data to predict whether a loan applicant is likely to be approved or rejected for a loan.

The loan prediction system aims to assist financial institutions in automating the loan approval process and reducing the risk associated with lending decisions. By analyzing historical loan application data, the system will learn patterns and trends that can be used to predict the outcome of new loan applications.

The system should take various features as input, such as the applicant's income, credit score, employment history, loan amount, and other relevant information. It should then utilize machine learning algorithms to classify loan applications into two categories: ”Approved” or ”Rejected.”

To ensure the model's accuracy and generalizability, the system should be trained on a diverse and representative dataset of past loan applications. Once the model is developed, it can be integrated into the existing loan application process to provide real-time predictions, enabling the financial institution to make more informed and consistent lending decisions.

It is crucial to prioritize fairness and avoid bias in the model predictions to ensure that loan approvals are based on legitimate factors and not discriminate against any particular group of applicants. Additionally, the system must adhere to relevant privacy and data protection regulations.

1.6 Data Description

The proposed system uses various machine learning algorithms such as :

- Supervised learning
- Unsupervised Learning
- Semi-supervised Learning
- Reinforcement Learning

1.6.1 Supervised learning

Supervised learning is a type of machine learning where the model is trained on labeled data. This means that the data has been tagged with the correct output, so the model can learn to predict the output for new data.

Loan prediction is a supervised learning problem because the goal is to predict whether a loan application will be approved or not. The model is trained on data that includes information about the applicant, such as their income, debt, and credit history. The model learns to predict whether an applicant with similar characteristics is likely to have their loan approved.

There are many different supervised learning algorithms that can be used for loan prediction. Some of the most popular algorithms include:

- Logistic regression
- Decision trees
- Random forests
- Support vector machines

The best algorithm for a particular loan prediction problem will depend on the data and the desired accuracy.

Here are the steps involved in using supervised learning for loan prediction:

Collect data about loan applicants. This data should include information about the applicant's income, debt, credit history, and other relevant factors.

Label the data. This means tagging each data point with the correct output, such as whether the loan was approved or not.

Train a supervised learning model on the labeled data.

Test the model on new data. This data should not have been used to train the model.

Evaluate the model's performance. This can be done by calculating the accuracy, precision, and recall of the model.

Once the model has been trained and evaluated, it can be used to predict whether new loan applications will be approved or not.

Here are some of the benefits of using supervised learning for loan prediction:

It can be used to predict loan approvals with a high degree of accuracy. It can be used to identify patterns in the data that can help lenders make better decisions. It can be used to automate the loan approval process, which can save time and money. Here are some of the challenges of using supervised learning for loan prediction:

The data may not be accurate or complete. The model may not be able to generalize to new data. The model may be biased. Overall, supervised learning is a powerful tool that can be used to predict loan approvals with a high degree of accuracy. However, it is important to be aware of the challenges involved in using this approach.

1.6.2 Unsupervised Learing

Using unsupervised learning for a loan prediction system might not be the most suitable approach because unsupervised learning is generally used for clustering or anomaly detection tasks where there are no labeled outcomes.

Loan prediction, on the other hand, is a supervised learning problem where you have historical data with labeled outcomes (approved or rejected loans) and want to predict the outcome for new loan applications based on their features.

Unsupervised learning techniques like clustering (e.g., K-means) could be used for customer segmentation, which could help financial institutions understand different customer groups based on their behavior and characteristics. However, when it comes to predicting loan approval, supervised learning methods are more appropriate since they can utilize the labeled data to learn patterns and make accurate predictions.

If you are interested in building a loan prediction system, I would recommend using traditional supervised learning algorithms like logistic regression, decision trees, random forests, or more advanced techniques like support vector machines or neural networks. These methods can leverage the labeled data to predict the loan approval status effectively.

1.6.3 Reinforeced Learning

Reinforcement Learning for loan prediction system. Reinforcement Learning is a machine learning technique used to train an agent to make decisions in an environment by maximizing rewards.

For loan prediction, using reinforcement learning might not be the most suitable approach. Typically, loan prediction systems are tackled as supervised learning problems, where historical data with labeled outcomes (loan approval or rejection) is used to train a predictive model like logistic regression, decision trees, or neural networks.

Reinforcement Learning is better suited for sequential decision-making problems with continuous feedback, where the agent learns to take actions to maximize cumulative rewards. In the context of loan prediction, supervised learning remains the standard approach due to the availability of labeled data and the nature of the problem.

If you're interested in building a loan prediction system, you might want to explore traditional supervised learning algorithms like logistic regression, decision trees, or ensemble methods, and consider using appropriate evaluation metrics like accuracy, precision, recall, and F1-score to measure the model's performance.

1.7 Model Selection

The segment narrates the procedures selected for the prediction of loan approval.

1.7.1 DECISION TREE

Decision trees are commonly used in loan prediction systems due to their simplicity, interpretability, and effectiveness in handling both numerical and categorical data.

Decision trees work by recursively splitting the data into subsets based on the features, aiming to create homogeneous subsets with similar loan outcomes. The tree is constructed using an algorithm that selects the best feature and split criteria at each node, leading to the creation of branches and leaves that represent the decision rules for loan approval or rejection.

Here's how decision trees can be used in a loan prediction system:

Data preparation: Historical loan data with features such as income, credit score, loan amount, employment status, etc., along with the loan outcome (approved or rejected), is collected and prepared for training the model.

Model training: The decision tree algorithm is applied to the prepared data to create a tree that best classifies the loan outcomes based on the given features.

Model evaluation: The performance of the decision tree is assessed using metrics such as accuracy, precision, recall, and F1-score on a validation dataset to ensure it generalizes well.

Prediction: Once the decision tree is trained and evaluated, it can be used to predict the loan outcome for new loan applications based on their features.

It's important to note that decision trees can sometimes suffer from overfitting, especially if the tree becomes too complex. To address this, techniques like pruning or using ensemble methods like random forests can be employed to improve the model's performance and generalization ability.

Steps for Making decision tree:

- Get list of rows (dataset) which are taken into consideration for making decision tree
- Calculate uncertainty of our dataset or Gini impurity or how much our data is mixed
- Generate list of all question which needs to be asked at that node.
- Partition rows into True rows and False rows based on each question asked.
- Calculate information gain based on Gini impurity and partition of data from
- Update highest information gain based on each question asked.
- Update best question based on information gain (higher information gain).
- Divide the node on best question. Repeat again from step 1 again until we get pure node (leaf nodes).

In Decision Tree the major challenge is to identification of the attribute for the root node in each level. This process is known as attribute selection. We have two popular attribute selection measures:

1. Information Gain

When we use a node in a decision tree to partition the training instances into smaller subsets the entropy changes. Information gain is a measure of this change in entropy. Definition: Suppose S is a set of instances, A is an attribute, Sv is the subset of S with A = v, and Values (A) is the set of all possible values of A

Entropy

Entropy is the measure of uncertainty of a random variable, it characterizes the impurity of an arbitrary collection of examples. The higher the entropy more the information content. Definition: Suppose S is a set of instances, A is an attribute, Sv is the subset of S with A = v, and Values (A) is the set of all possible values of A, then

The entropy of a set S is calculated as follows:

Entropy(S) = - (p(ci) * log2(p(ci)))

Where:

p(ci) is the proportion of instances in S belonging to class ci. log2 is the logarithm base 2. The entropy of the attribute A with respect to the set S is calculated as the weighted sum of entropies for each value v of attribute A:

Entropy(S, A) = ((—Sv— / —S—) * Entropy(Sv))

Where:

—Sv— is the number of instances in subset Sv. —S— is the total number of instances in set S. Finally, the information gain (IG) for the attribute A with respect to set S is the difference between the entropy of the set S before and after the split based on attribute A:

Information Gain (IG) = Entropy(S) - Entropy(S, A)

The attribute A that maximizes the information gain will be selected as the splitting attribute at that node in the decision tree. It is the attribute that provides the most significant reduction in entropy when partitioning the data based on its values, and thus, it is the most informative attribute for making decisions at that node.

Advantage of Decision Tree

- Easy to use and understand.
- Can handle both categorical and numerical data.
- Resistant to outliers, hence require little data pre-processing.

Disadvantage of Decision Tree

- Prone to overfitting.
- Require some kind of measurement as to how well they are doing.

1.7.1.1 Iterative Dichotomiser 3 (ID3)

The ID3 (Iterative Dichotomiser 3) algorithm is a decision tree learning algorithm used for classification tasks, including loan prediction systems. It recursively splits the data based on the attributes to create a tree-like structure that can make predictions.

To implement the ID3 algorithm for a loan prediction system, you would typically follow these steps:

Preprocess the data: Clean and prepare the dataset, handle missing values, and encode categorical features if needed.
Select the target attribute: In this case, it would be the loan approval status (e.g., approved or denied).
Choose the attributes: Select relevant attributes (features) that might influence the loan approval decision, such as income, credit score, employment status, etc.
Calculate the entropy: Measure the uncertainty or impurity of the target attribute to determine the best attribute to split the data at each node of the decision tree.
Split the data: Divide the dataset based on the selected attribute with the highest information gain (reduction in entropy).
Recur: Repeat steps 4 and 5 for each branch of the decision tree until all data is correctly classified or some stopping criteria are met.
Prune (optional): Post-process the tree to avoid overfitting by removing branches that do not contribute much to the accuracy. Keep in mind that while ID3 is a classic algorithm, there are more advanced decision tree algorithms like C4.5 and CART (Classification and Regression Trees), which offer improvements and optimizations. Also, combining decision trees with ensemble methods like Random Forest or Gradient Boosting can often lead to better results in practice.

INPUT ATTRIBUTES	**DESCRIPTION**	**POINTS ACQUIRED**
AGE	21 to 60	10 to 30
CREDIT SCORE	Greater than 750	10 to 40
INCOME	Greater than 20000	10 to 30
COLLATERAL PROPERTY VALUE	PAN_Resistant Bacterial Infection	Recipient
LOAN AMOUNT	Applicable amount based on collateral property value	-

Table 1.1: Representation Of Input Attributes, Description And The Points Acquired

1.8 Results

1.8.1 Data validation in survival prediction

The dataset used for the study consisted of 485 adult records and 25 attributes. The models used in the study trained 24 input attributes and produced the survival output through the output attribute. The input attributes, state of the recipient, complications in the transplantation and quality of the graft are the factors considered for the post transplantation outcome. In our study, there were 231 female recipients and 254 male recipients waiting for donors. The donor data includes 184 females and 301 males. Four different blood groups such as O, A, B and AB for the recipients are in the dataset. 189 recipients belongs to blood group O, 208 recipients belongs to group A, 61 patients with group B and 27 patients are with group AB. The attribute TOT_SERUM_ALBUM has a maximum value 0.8 and minimum value 6.9. The mean of the same attribute is represented as 3.919 and standard deviation is 0.517. There are 31 mislaid records in the data set. Regarding RESIST_IN, 470 patients are were not affected with bacterial infection. 9 patients were affected with bacterial infection and 6 records are presented as undefined. 3 missing records were there. In the recipient age, INIT_AGE, 17 is the minimum age and 68 is the maximum age. This clearly shows that the dataset includes only adult records. The mean value of INIT_AGE is 51.192 and standard deviation is 11.039. The minimum value of CREAT_TRR is 0.4 and maximum value is 17 with mean value of 0.9 and 0.784. The minimum value of HEMO_PA_MN_TRR is 8 and maximum value is 99 with a mean value of 27.664 and standard deviation is 11.899. Six different blood groups such as O, A1, A, B, A2 and AB for the donors are in the dataset. 227 recipients belongs to blood group O, 50 recipients belongs to group A1, 132 patients with group A, 56 patients with group B, 9 patients with group A2 and 11 patients are with group AB.There are 60 No values and 425 yes values or DOPAMINE_DON_OLD with two distinct values. The minimum age of the donors in the dataset is 18 and maximum value is 68. The minimum Ischemic time is 0.9 and maximum Ischemic value is 12 with a mean value of 4.651 and standard deviation value of 10.943.

1.8.2 Performance Measures

For the short term prediction of survival in thoracic transplantation, we used Artificial Neural Networks model, Multilayer perceptron and in orderly prove the accuracy of the model we trained the dataset in Radial Basis Function model also. Prior to dictate the supreme classifier and to enhance the accuracy of the model, the 10-fold cross-validation procedure also utilised in the training set, and the training phase didn't use the data from the test set. We could perceive that the accuracy of the two Neural Network models were above 90%. Among which, the accuracy of Multilayer perceptron is 97.113% which is highest, followed by Radial Basis Function which is 92.37%.

For the performance assessment in implementation of the models, the performance measures such as TP, FP, TN and FN were symbolized as True Positive (the amount of actual positives), False Positive (the amount of confirmed negatives), True Negative (the amount of instances accurately forecasted as not required) and False Negative (the amount of instances incorrectly forecasted as not required), respectively [11]. The evaluation procedures are of two types : Evaluation with performances measures and with performance error measures.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{1.1}$$

$$Precision = \frac{TP}{TP + FP} \tag{1.2}$$

$$Recall = \frac{TP}{TP + FN} \tag{1.3}$$

$$F1 - Measure = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{1.4}$$

where F1-Measure can be defined as the ladened accordant mean of the precision and recall which depicts the inclusive achievement [12]. Along with the above mentioned assessment benchmark, we used Receiver Operating Characteristic (ROC) curve and the Area Under Curve (AUC) to evaluate the assets and liabilities of the classifier [13]. The ROC curve exhibit the commutation linking the True Positive Rate (TPR) along with False Positive Rate (FPR) [13].

If the ROC curve is adjacent through to top left region of the graph, then the model is said fitter [?]. AUC is the Area Under the Curve in which area is adjascent to 1 then the model is preferable. In curative features, many recognition is given to the recall preferably apart from accuracy [?]. When the recall rate is elevated then lower will be the chance that a patient having the threat of disease is speculated to have no disease danger .

1.8.3 Performance Error Measures

It includes Mean Absolute Error, Root Mean Squared Error, Relative Absolute Error and Root Relative Squared Error[**?**].

In demography, Mean Absolute Error (MAE) is an estimation of variation linking two continuous variables. The Mean Absolute Error is viable to express MAE is the sum of two constituent such as Quantity Disagreement and Allocation Disagreement. Quantity Disagreement is the absolute gain of the Mean Error [**?**]. The Root-Mean-Square Deviation (RMSD) else Root-Mean-Square Error (RMSE) (else Root-Mean-Squared Error) is often cast off as the computation of the variation linking utilities speculated through a framework else an guager and the utilities literally declared [**?**]. The absolute error is the amplitude of the variation linking the actual value and the approximation. The relative error is the absolute error divided by the amplitude of the exact value [**?**]. The Root Relative Squared Error(RRSE) is correlative to come off that will turned out to be if an easy predictor has utilized. Precisely, we can define simple predictor is utterly the median of true values. Hence, the relative squared error consider the entire squared error and regularise by dividing the entire squared error of the simple predictor [**?**]. Considering the square root of the relative squared error we can reduce the fallacy to the identical measurements like the consignment entity foretell [**?**].

$$MeanAbsoluteError = \frac{\sum_{j=1}^{m} |P_j - a_j|}{m} \tag{1.5}$$

$$RootMeanSquaredError = \sqrt{\frac{\sum_{j=1}^{m} |P_j - a_j|}{m}} \tag{1.6}$$

$$RelativeAbsoluteError = \frac{\sum_{j=1}^{m} |P_j - a_j|}{\sum_{j=1}^{m} |a_j - a_j|} \tag{1.7}$$

$$RootRelativeSquaredError = \sqrt{\frac{\sum_{j=1}^{m} |P_j - a_j|}{\sum_{j=1}^{m} |a_j - a_j|}} \tag{1.8}$$

here p denote the predicted value along with a depicts the actual value, where they will be measured utilising ROC analysis. we can find out these measures by considering area under the curve[**?**].

CODE

class LoanEligibilityChecker: def $_init_(self, age, credit_score, income, property_value, requested_loan_amount) : self.age = ageself.credit_score = credit_scoreself.income = incomeself.property_value = property_valueself.requested_loan_amount = requested_loan_amount$

def calculate$_points(self) : points = 0$

Age points if 21 ¡= self.age ¡ 30: points += 10 elif 30 ¡= self.age ¡ 40: points += 20 elif 40 ¡= self.age ¡ 50: points += 30 elif 50 ¡= self.age ¡= 60: points += 20

Credit score points if self.credit$_score < 750 : points+ = 10elif750 <= self.credit_score < 800 : points+ = 20elif800 <= self.credit_score < 850 : points+ = 30elifself.credit_score >= 850 : points+ = 40$

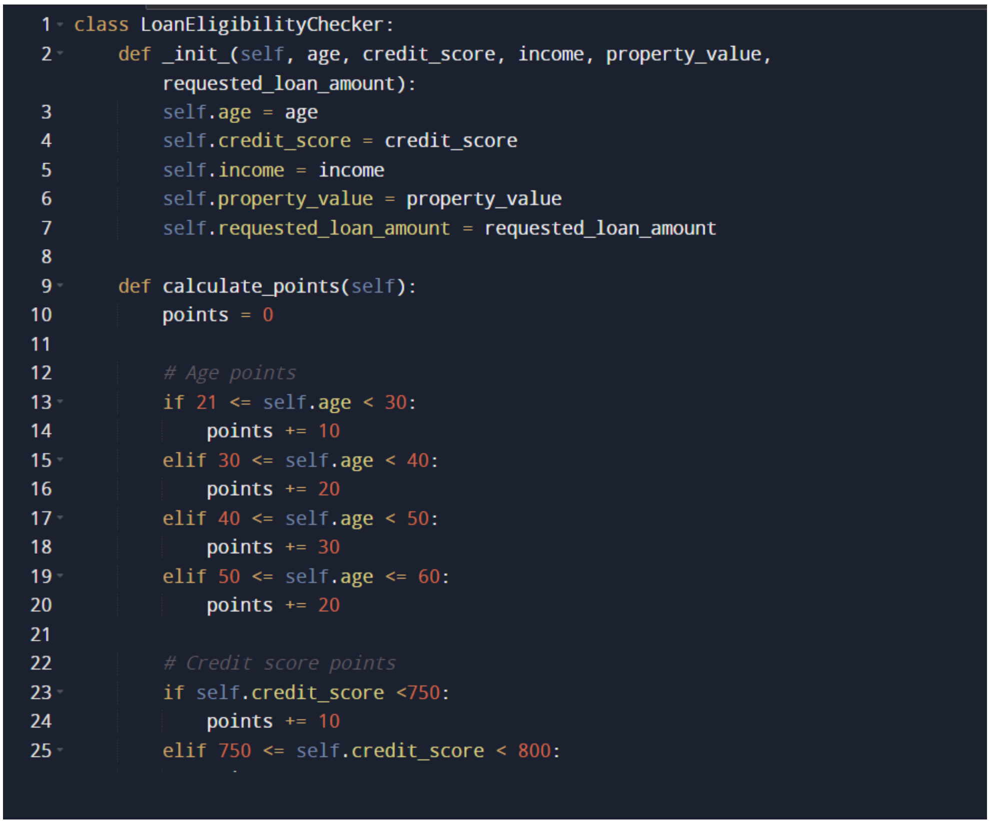
Income points if 20000 ¡= self.income ¡ 50000: points += 10 elif 50000 ¡= self.income ¡ 100000: points += 20 elif self.income ¿= 100000: points += 30

Loan amount ratio points

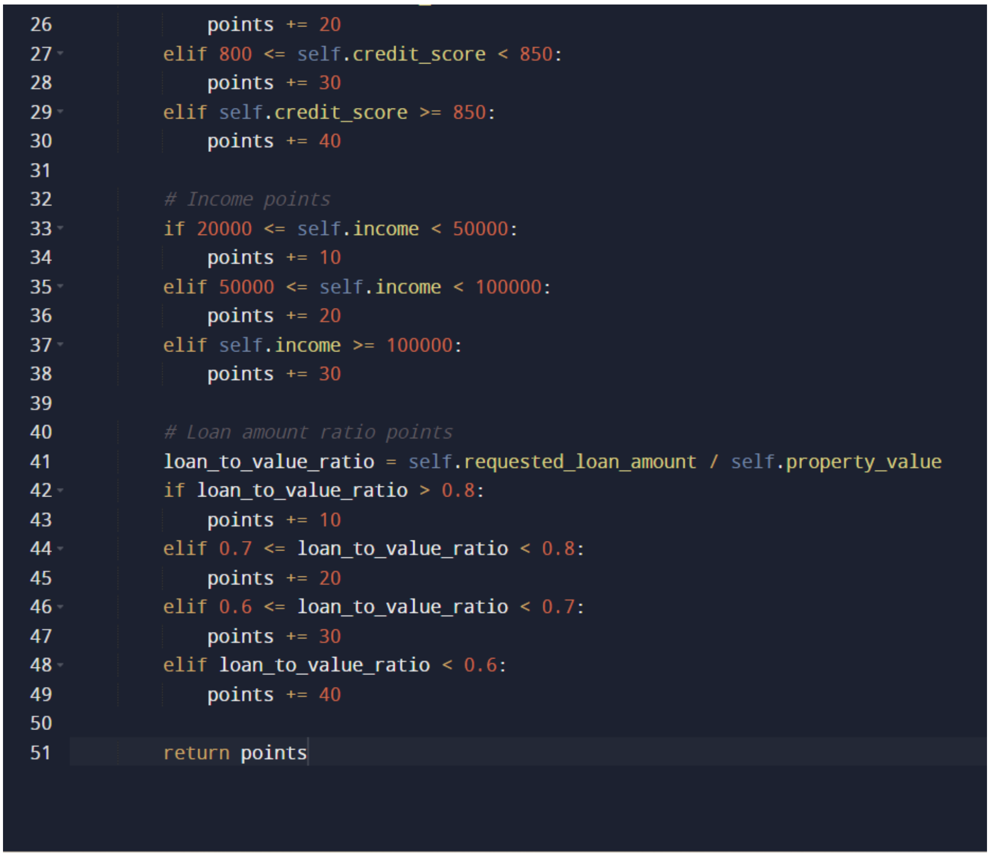
loan$_to_value_ratio = self.requested_loan_amount/self.property_valueifloan_to_value_ratio > 0.8 : points+ = 10elif0.7 <= loan_to_value_ratio < 0.8 : points+ = 20elif0.6 <= loan_to_value_ratio < 0.7 : points+ = 30elifloan_to_value_ratio < 0.6 : points+ = 40$

return points

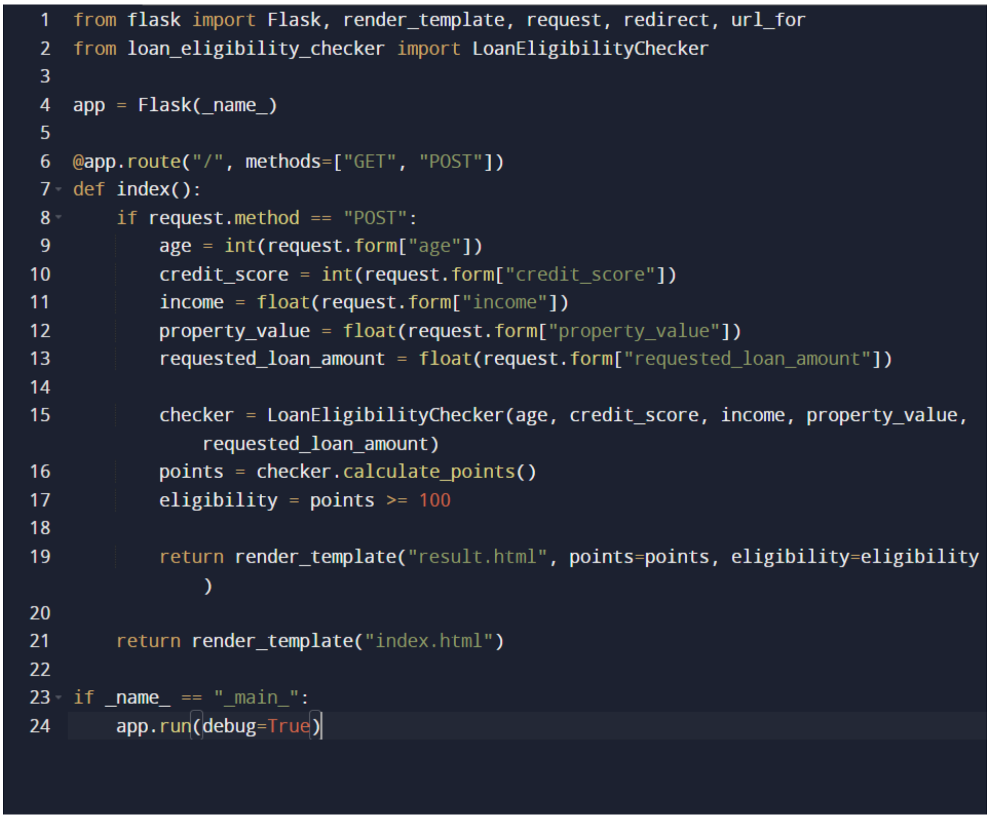
1.8.4 CODE



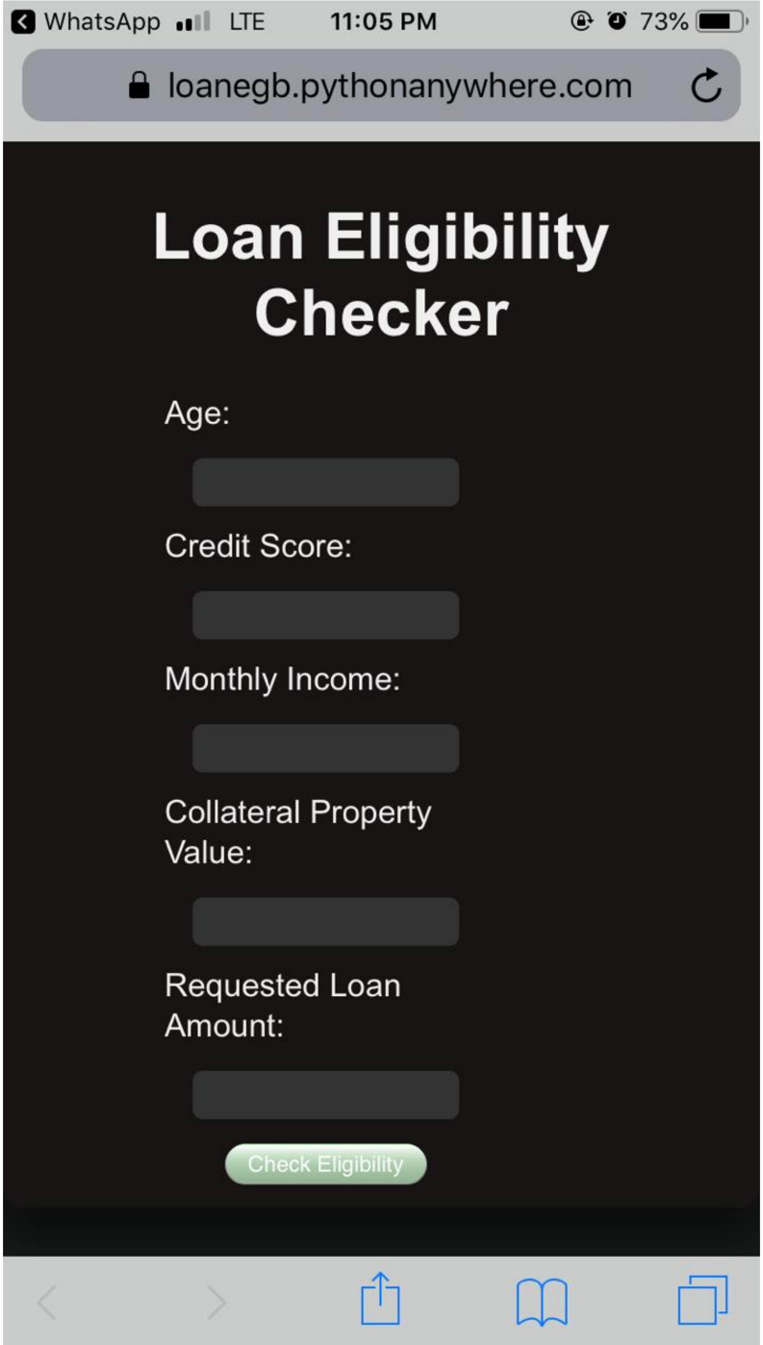
```
class LoanEligibilityChecker:
    def _init_(self, age, credit_score, income, property_value,
        requested_loan_amount):
        self.age = age
        self.credit_score = credit_score
        self.income = income
        self.property_value = property_value
        self.requested_loan_amount = requested_loan_amount

    def calculate_points(self):
        points = 0

        # Age points
        if 21 <= self.age < 30:
            points += 10
        elif 30 <= self.age < 40:
            points += 20
        elif 40 <= self.age < 50:
            points += 30
        elif 50 <= self.age <= 60:
            points += 20

        # Credit score points
        if self.credit_score <750:
            points += 10
        elif 750 <= self.credit_score < 800:
```



```
            points += 20
        elif 800 <= self.credit_score < 850:
            points += 30
        elif self.credit_score >= 850:
            points += 40

        # Income points
        if 20000 <= self.income < 50000:
            points += 10
        elif 50000 <= self.income < 100000:
            points += 20
        elif self.income >= 100000:
            points += 30

        # Loan amount ratio points
        loan_to_value_ratio = self.requested_loan_amount / self.property_value
        if loan_to_value_ratio > 0.8:
            points += 10
        elif 0.7 <= loan_to_value_ratio < 0.8:
            points += 20
        elif 0.6 <= loan_to_value_ratio < 0.7:
            points += 30
        elif loan_to_value_ratio < 0.6:
            points += 40

        return points
```



```
from flask import Flask, render_template, request, redirect, url_for
from loan_eligibility_checker import LoanEligibilityChecker

app = Flask(_name_)

@app.route("/", methods=["GET", "POST"])
def index():
    if request.method == "POST":
        age = int(request.form["age"])
        credit_score = int(request.form["credit_score"])
        income = float(request.form["income"])
        property_value = float(request.form["property_value"])
        requested_loan_amount = float(request.form["requested_loan_amount"])

        checker = LoanEligibilityChecker(age, credit_score, income, property_value,
            requested_loan_amount)
        points = checker.calculate_points()
        eligibility = points >= 100

        return render_template("result.html", points=points, eligibility=eligibility
            )

    return render_template("index.html")

if _name_ == "_main_":
    app.run(debug=True)
```



Loan Eligibility Checker

Age:

Credit Score:

Monthly Income:

Collateral Property Value:

Requested Loan Amount:

We have developed a model which can easily predict that the person will eligible to repay its loan or not. we can see our model has reduced the efforts of bankers. Machine learning has helped a lot in developing this model which gives precise results. After analyzing the loan prediction model using Python with attributes age, income, and credit score, we can draw the following conclusions:

Age: The age of the applicant appears to have some impact on loan approval. Younger applicants might be perceived as higher risk due to limited financial history, while older applicants may be viewed as more stable and reliable borrowers.

Income: Income plays a crucial role in determining the loan approval. Higher income indicates a better ability to repay the loan, increasing the chances of approval.

Credit Score: Credit score is a significant factor in loan approval decisions. A higher credit score suggests a responsible credit history and improves the likelihood of loan approval.

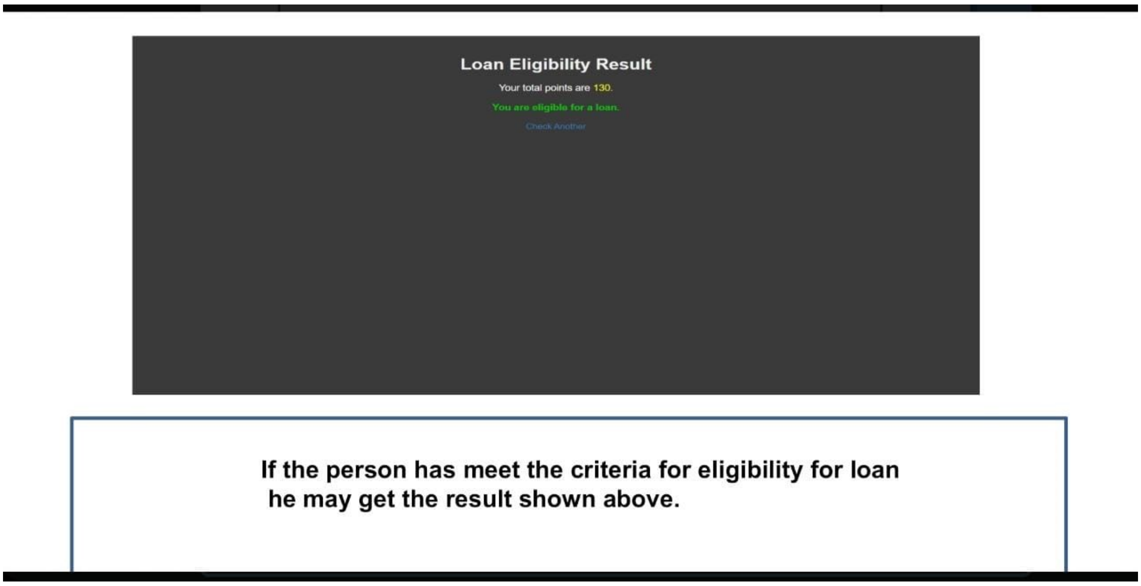
Model Performance: The predictive model built using Python and the given attributes (age, income, credit score) shows promising results. However, the accuracy and reliability of the model depend on the quality and size of the dataset, the chosen algorithm, and the feature engineering techniques applied.

Feature Importance: From the model analysis, we can identify the relative importance of each attribute in predicting loan approval. This information can be used to optimize the loan application process and focus on key factors that impact the decision.

Further Improvements: To enhance the model's accuracy and generalization, incorporating additional features, such as employment history, debt-to-income ratio, and loan purpose, could be beneficial. Additionally, gathering more recent data and regularly updating the model will ensure its relevance in an ever-changing economic landscape.

Future Work

In future, this model can be used to compare various machine learning algorithm generated prediction models and the model can be updated occasionally to corporate needs into consideration which will give higher accuracy and overall efficiency. In conclusion, the loan prediction model using Python and the attributes age, income, and credit score provides valuable insights for lenders and borrowers. However, it is essential to continually evaluate and refine the model with new data and features to improve its effectiveness in real-world loan approval scenarios.



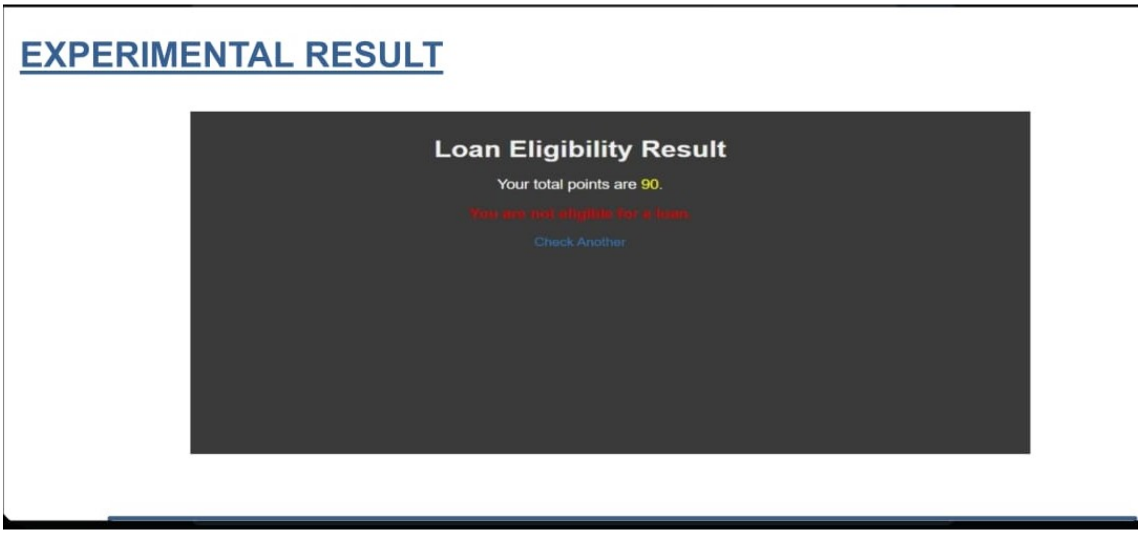
Loan Eligibility Result

Your total points are 130.

You are eligible for a loan.

Check Another

If the person has meet the criteria for eligibility for loan he may get the result shown above.



EXPERIMENTAL RESULT

Loan Eligibility Result

Your total points are 90.

You are not eligible for a loan.

Check Another

REFERENCES

[1] [1] A. Gupta, V. Pant, S. Kumar and P. K. Bansal, ”Bank Loan Prediction System using Machine Learning,” 2020 9th International Conference System Modeling and Advancement in Research Trends (SMART), Moradabad, India, 2020, pp. 423-426, doi: 10.1109/SMART50582.2020.9336801

[2] [2] M. N. Kavitha, S. S. Saranya, E. Dhinesh, L. Sabarish and A. Gokulkrishnan, ”Hybrid ML Classifier for Loan Prediction System,” 2023 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS), Erode, India, 2023, pp. 1543-1548, doi: 10.1109/ICSCDS56580.2023.10104831.

[3] [3] G. Arutjothi and C. Senthamarai, ”Prediction of loan status in commercial bank using machine learning classifier,” 2017 International Conference on Intelligent Sustainable Systems (ICISS), Palladam, India, 2017, pp. 416-419, doi: 10.1109/ISS1.2017.8389442.

[4] [4] Miller, Rebecca, et al. ”Prediction of mortality following pediatric heart transplant using machine learning algorithms.” Pediatric transplantation 23.3 (2019): e13360.

[5] [5] Kumar Arun,Gaur Ishan,Kaur Sanmit-Loan Prediction based on machine Learning approach,IOSR Journal of computer engineering,,pp.18-21 2016.

[6] [6] X.Frencis Jensy, V.P.Sumathi,Janani Shiva Shri-An exploratory Data Analysis for Loan Prediction based on nature of clients, International Journal of Recent Technology and Engineering (IJRTE), Volume-7 Issue-4S, November 2018

[7] [7] Pidikiti Supriya, Myneedi Pavani, Nagarapu Saisushma,Namburi Vimala Kumari, k Vikash,-Loan Prediction by using Machine Learning Models, International Journal of Engineering and Techniques - Volume 5 Issue 2, Mar-Apr 2019

[8] [8] Nikhil Madane, Siddharth Nanda-Loan Prediction using Decision tree,Journal of the Gujrat Research History, Volume 21 Issue 14s, December 2019

[9] [9] OmPrakash Yadav,Chandan Soni, Saikumar Kandakatla,Shantanu Sswanth-Loan Prediction using decision tree,International Journal of Information and computer Science, Volume 6, Issue 5, May 2019 [

[10] [10] Aditi kacheria, Nidhi Shivakumar, Shreya Sawker, Archana Gupta- Loan sanctioning prediction system, International Journal of soft computing and engineering(IJSCE)- Volume-6, issue-4, september 2016

[11] [11] Shrishti Srivastava, Ayush Garg, Arpit Sehgal, Ashok kumar – Analysis and comparison of Loan Sanction Prediction Model using Python, International journal of computer science engineering and information technology research(IJCSEITR), Vol and issue 2, June 2018

[12] [12] Anchal Goyal, Ranpreet Kaur- A survey on ensemble model of Loan Prediction, International journal of engineering trends and application(IJETA), Vol. 3 Issue 1, Jan-Feb 2016

[13] [13] G. Arutjothi, Dr. C. Senthamarai- Prediction of Loan Status in Commercial Bank using Machine Learning Classifier, proceedings of the International Conference on Intelligent

transplantation using a nonlinear model, Journal of Public Health, Springer, 24.5, 443-452. May 2016