Predictive Credit Risk Modeling:
Leveraging Decision Trees, XGBoost, and Random Forests

Zulfikar Azaria Rahman ID/X Partners (Rakamin Academy)

Hello everyone,

My name is Zulfikar Azaria Rahman. I am passionate about data and skilled in business strategy. Today, I am excited to present my project on **Predictive Credit Risk Modeling** using **Decision Trees, XGBoost, and Random Forests**.

This project leverages advanced machine learning techniques to enhance credit risk prediction. By applying these models, I aim to provide valuable insights for financial decision-making through rigorous data analysis and strategic considerations.

LinkedIn

Table of Contents

Business & Data Understanding

What is Credit Risk ?

Credit risk refers to the likelihood that a borrower will be unable to repay a loan or meet credit obligations, potentially causing financial losses for the lender.

Understanding and managing credit risk is crucial for financial institutions because it directly impacts their profitability, financial health, and operational stability.

The objective of the credit risk prediction project is to develop a model that can assess the likelihood of borrower default. This helps in reducing the rate of loan defaults, improving credit decision-making, and optimizing the loan portfolio.

About Dataset

This dataset includes various information about consumers who apply for loans. These data contain borrower profile data, credit history, and loan status etc.

75 Column

466.285 Row

→

36 Column

438.679 Row

Data Understanding

Missing Value

There Is 40 Columns That Has Missing Value

Duplicated Data

There Is No Duplicated Data

Categorical Data

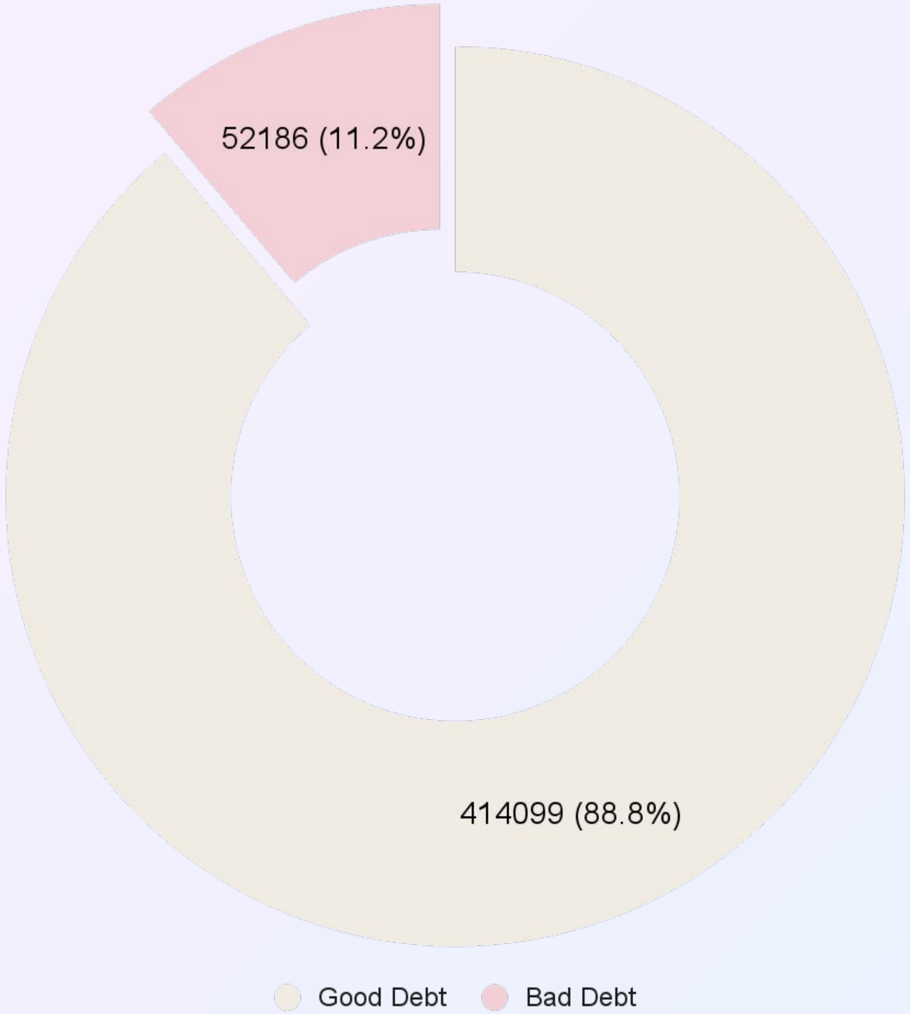
22
Categoricals Columns

Numerical Data

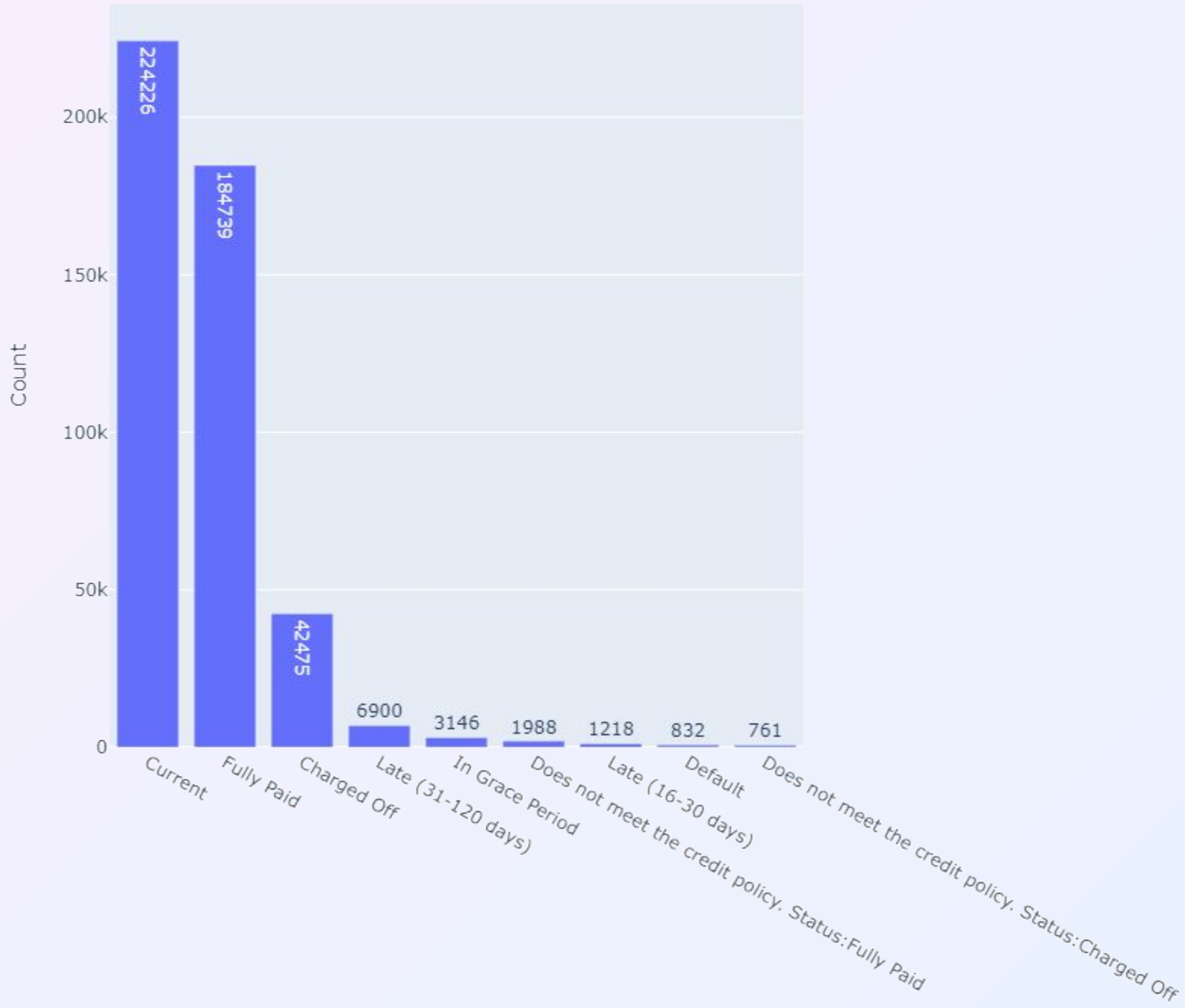
53
Numericals Columns

Exploratory Data Analysis

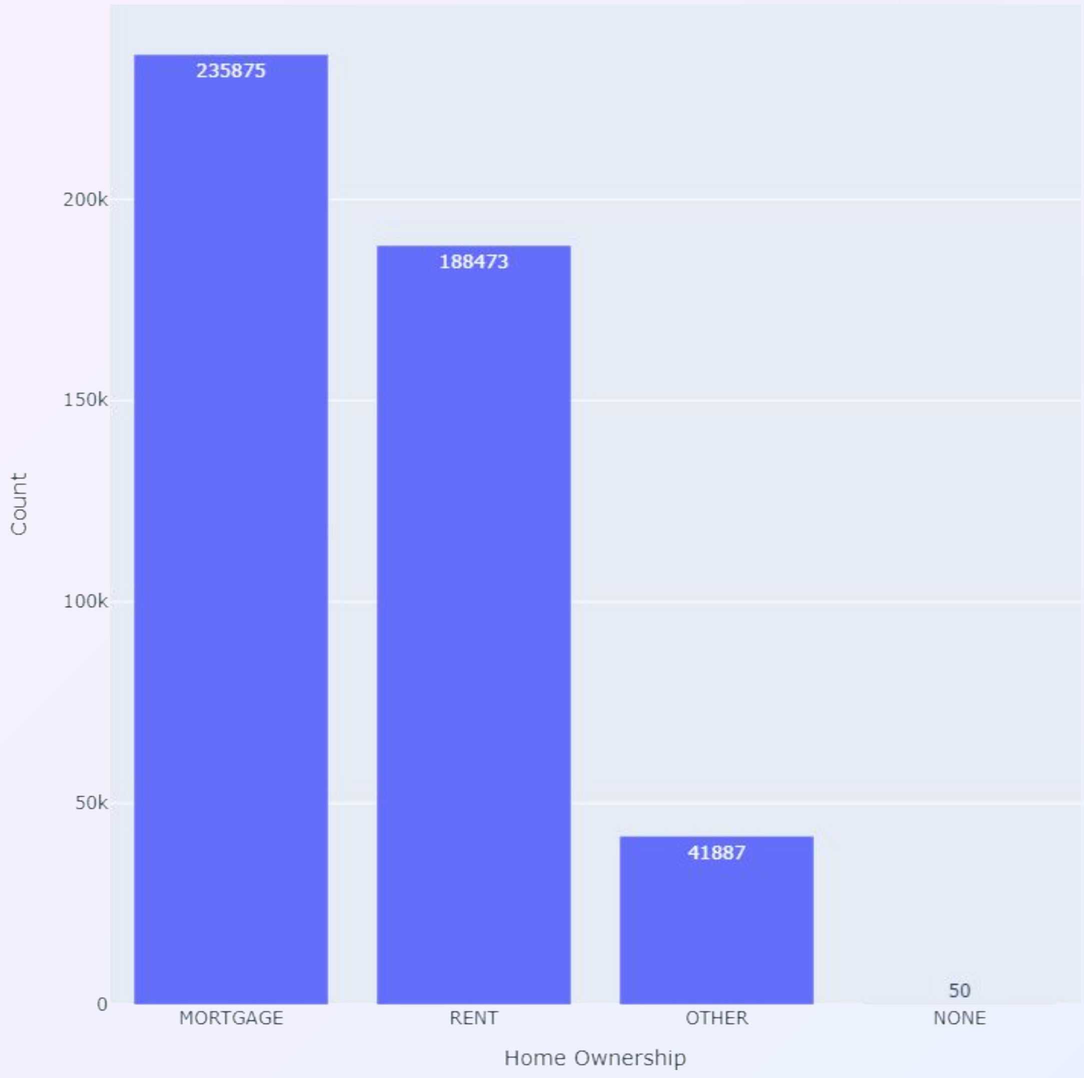
Loan Status Distribution



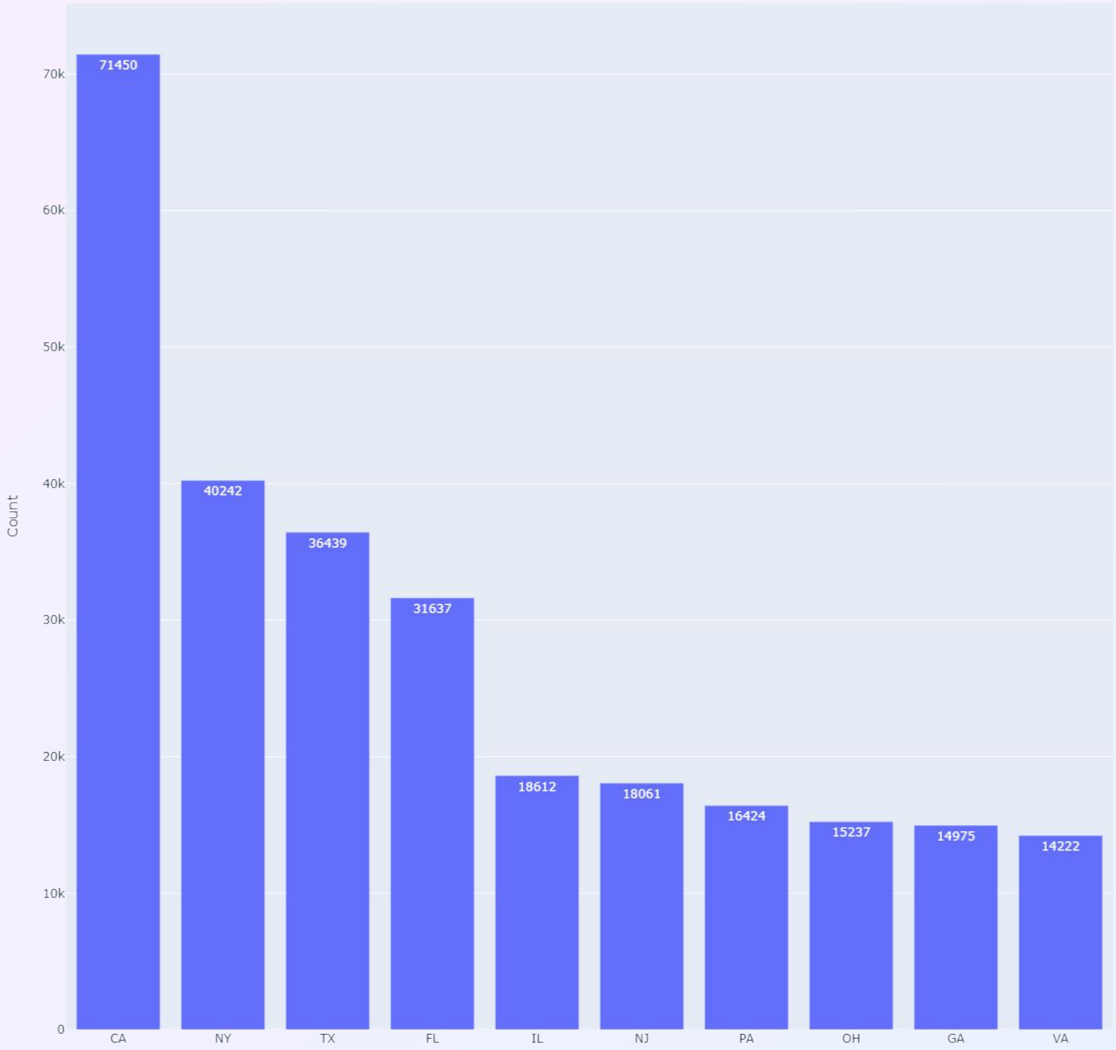
Loan Status Distribution

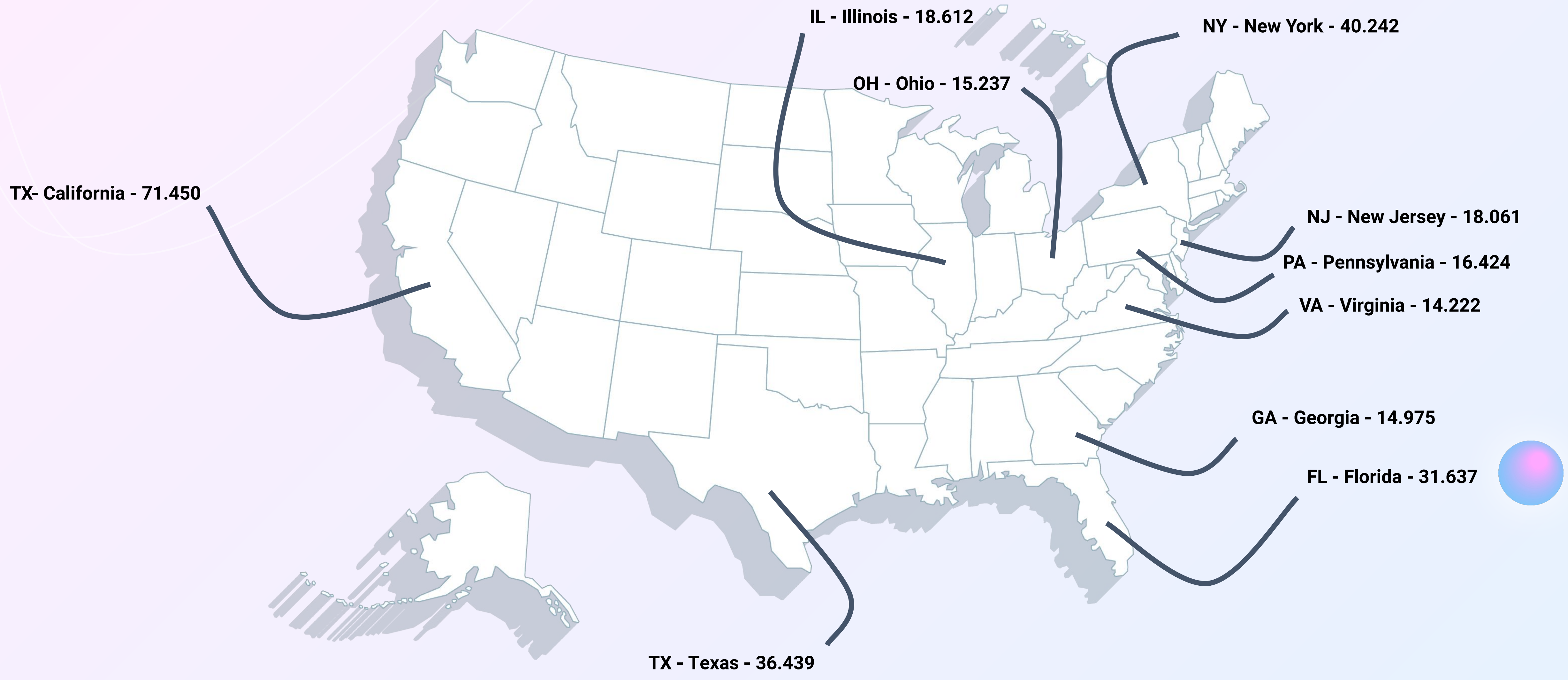


Home Ownership

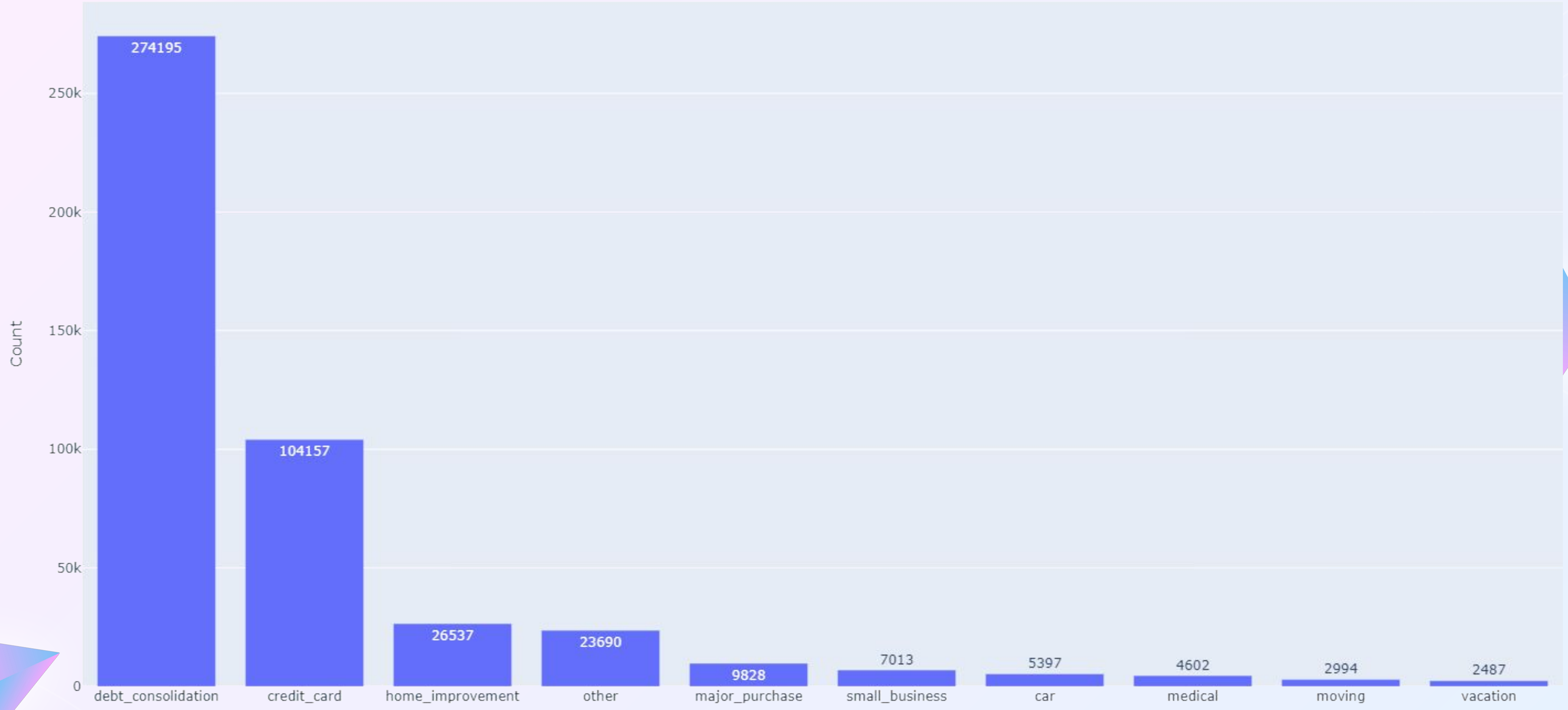


Top 10 Address State

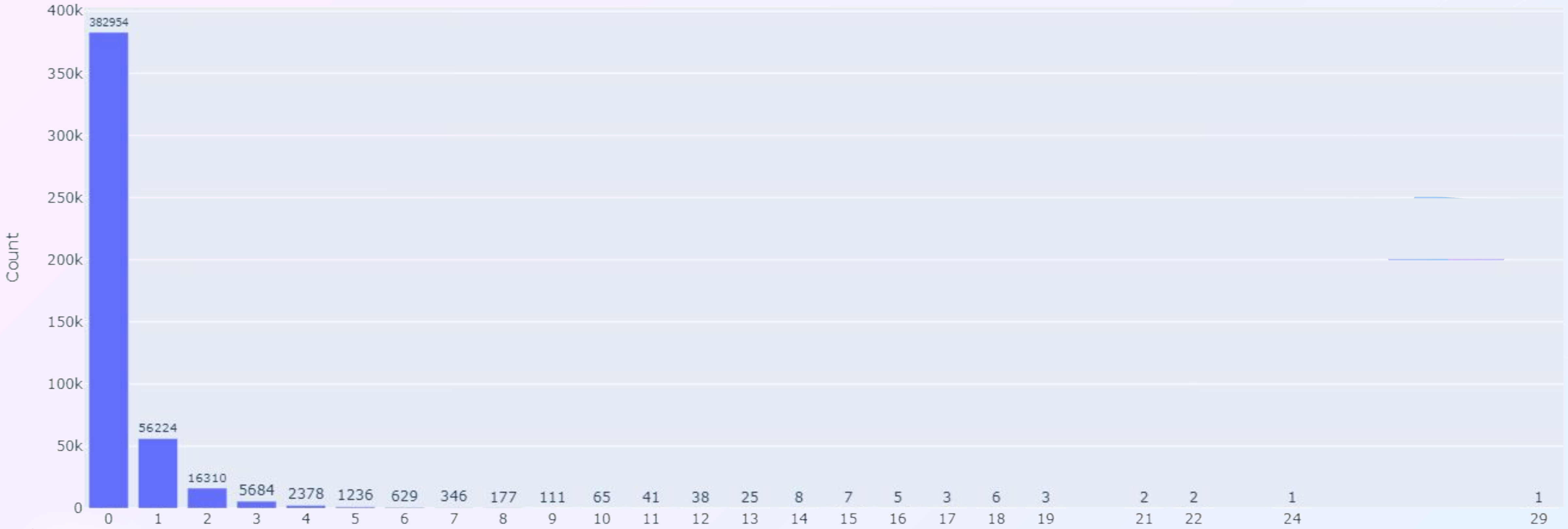



IL - Illinois - 18.612
NY - New York - 40.242
OH - Ohio - 15.237
TX- California - 71.450
NJ - New Jersey - 18.061
PA - Pennsylvania - 16.424
VA - Virginia - 14.222
GA - Georgia - 14.975
FL - Florida - 31.637
TX - Texas - 36.439

Top 10 Loan Purposes



Distribution of Delinquencies in The Last 2 Years



Data Preprocessing


Change the data type & Delete unnecessary columns.
Handling Imbalance
Missing Value Handling
Data Standardization & Data Splitting
Multicollinearity Study

Modeling

Modeling

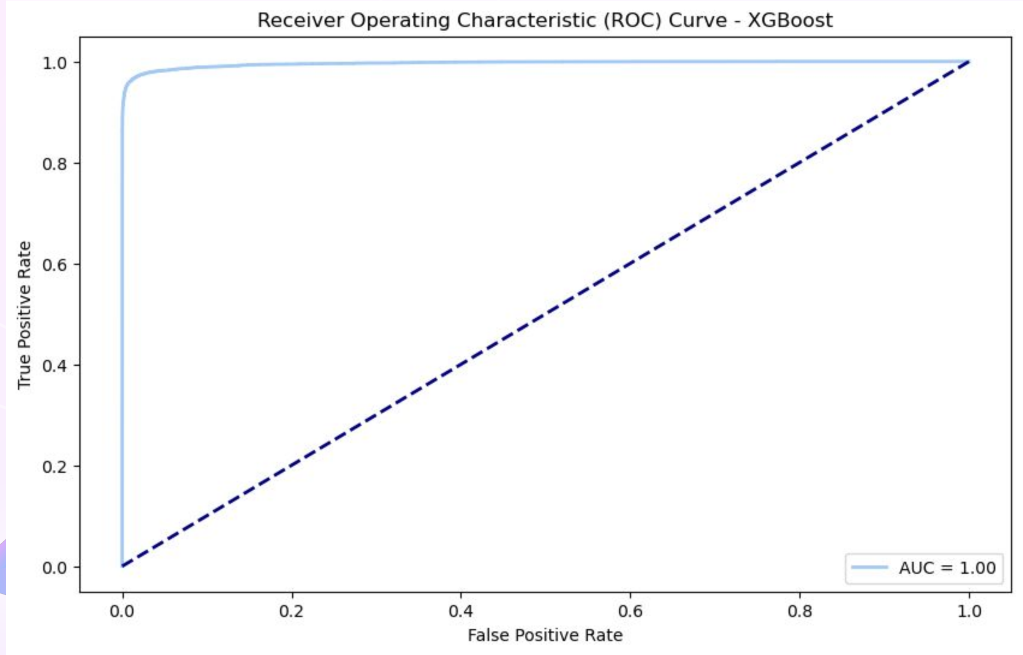
Training Data Result :

Model	Accuracy	Precision	Recall	F1 Score
XGBoost	98.90 %	99.35 %	90.36 %	94.64 %
Decision Tree	96.81 %	82.38 %	89.46 %	85.77 %
Random Forest	98.20 %	98.91 %	84.20 %	90.96 %

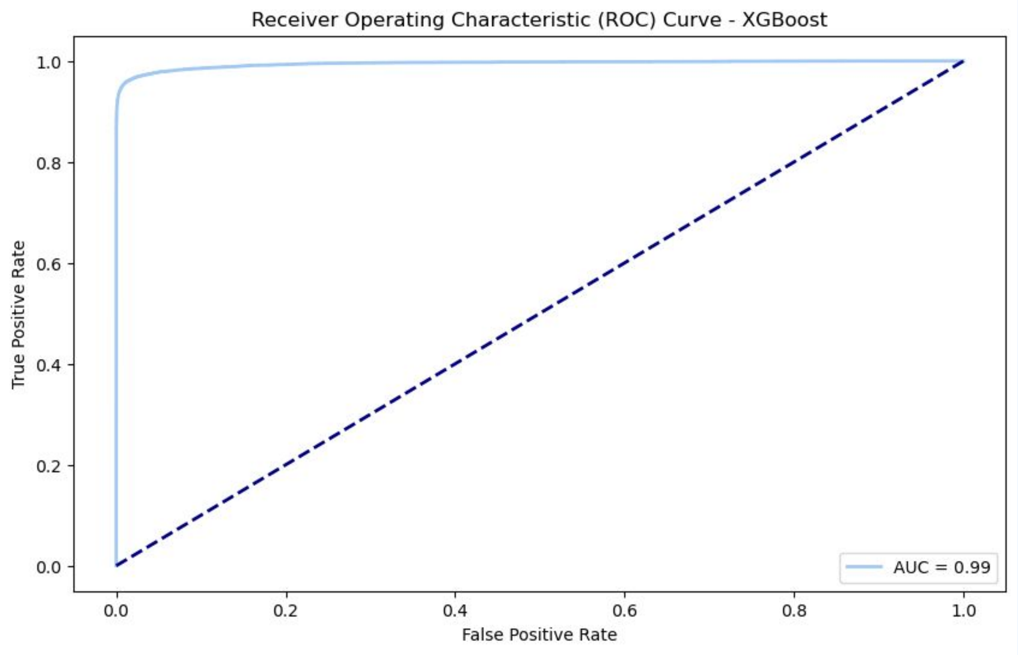
Testing Data Result :

Model	Accuracy	Precision	Recall	F1 Score
XGBoost	98.88 %	99.42 %	90.02 %	94.49 %
Decision Tree	96.81 %	82.00 %	89.63 %	85.64 %
Random Forest	98.20 %	98.96 %	83.95 %	90.84 %

Modeling

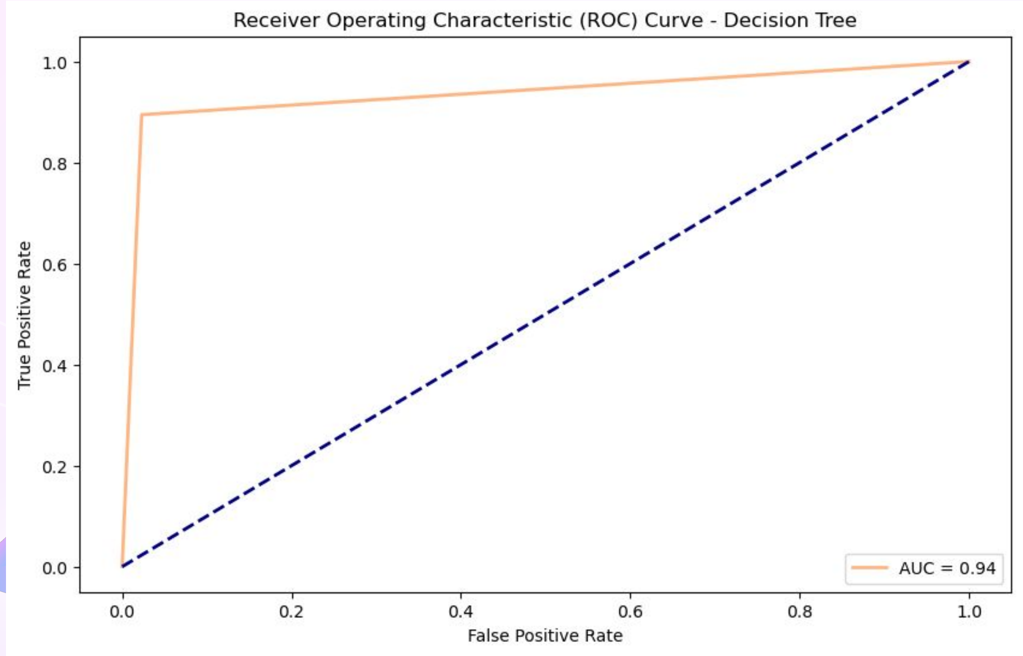


Training Data Result :

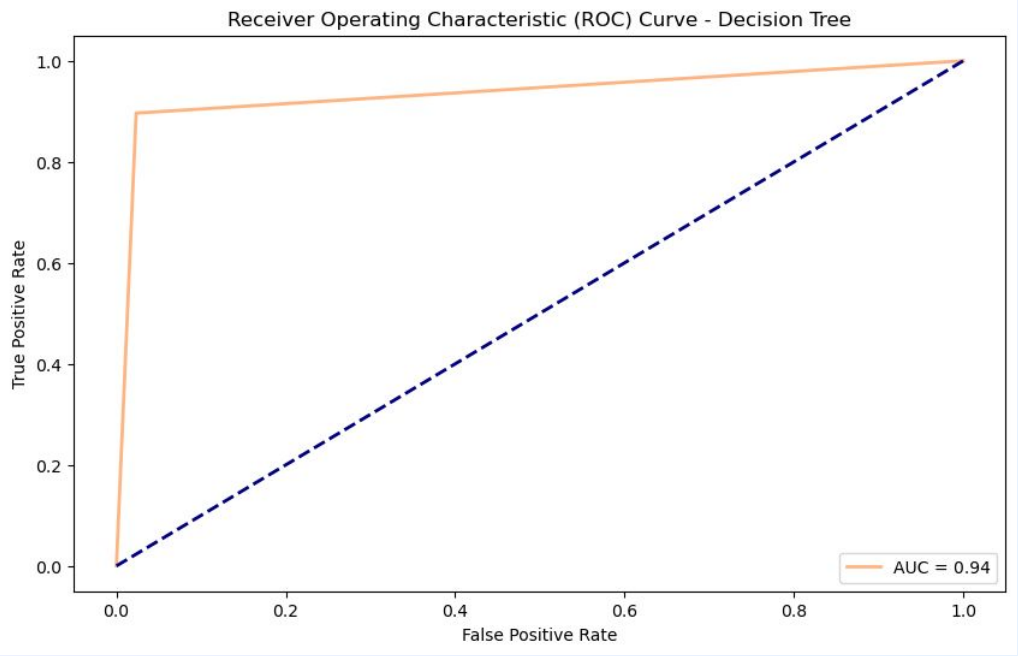


Testing Data Result :

Modeling

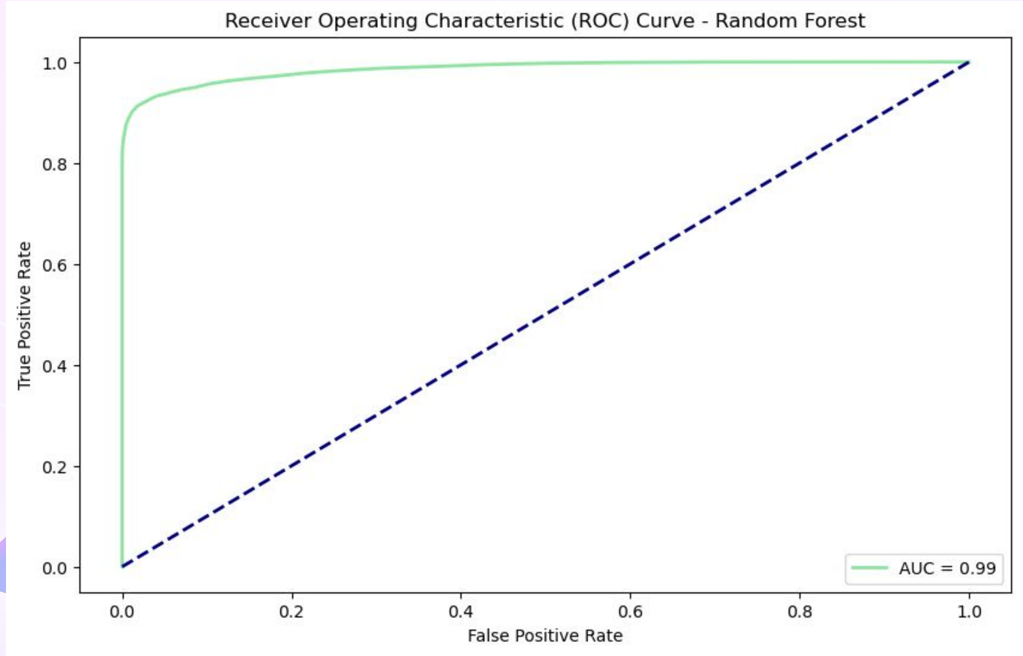


Training Data Result :

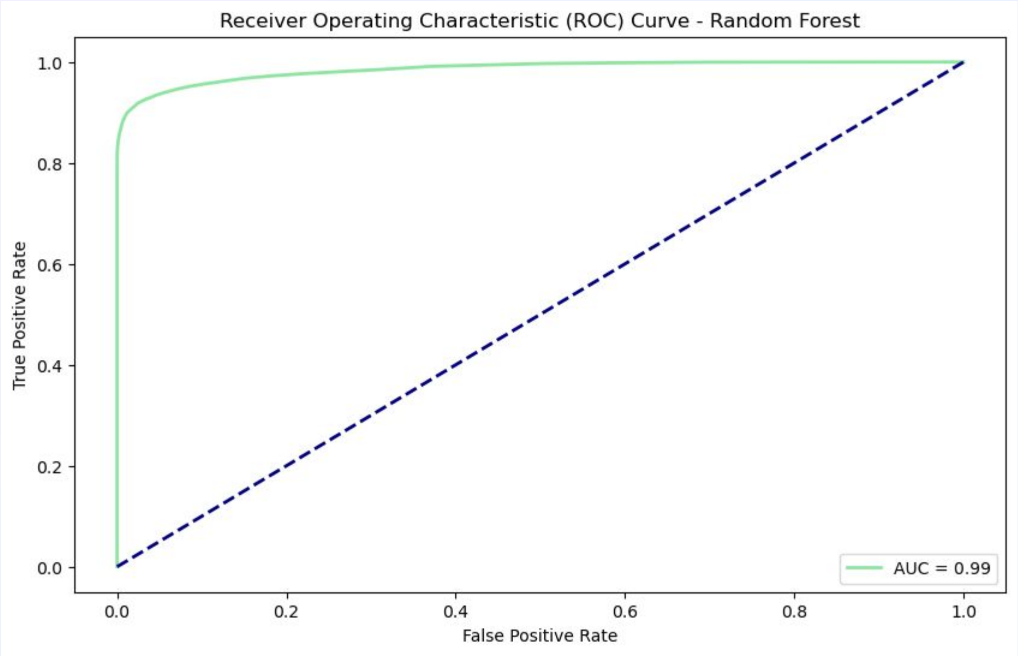


Testing Data Result :

Modeling



Training Data Result :



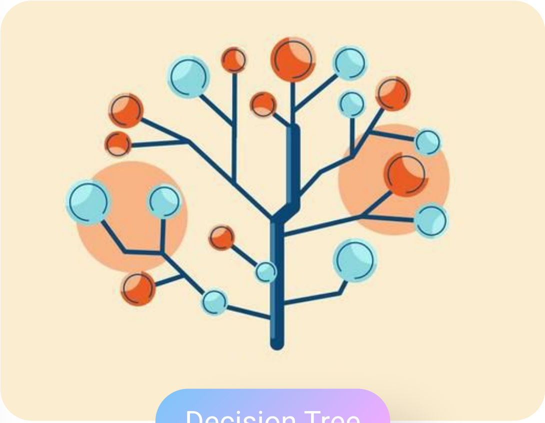
Testing Data Result :

Model Performance Summary

Model Performance Summary

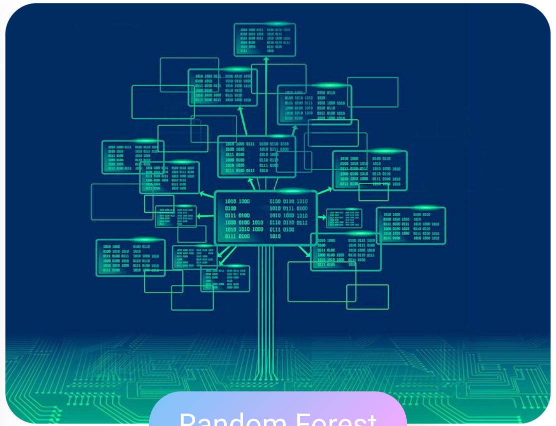
XGboost

The metrics for both training and test sets are very close, indicating no significant overfitting or underfitting. The model generalizes well to unseen data.



Decision Tree

The metrics for training and test sets are also quite close, suggesting no significant overfitting or underfitting. However, its overall performance is lower compared to XGBoost and Random Forest, indicating it might be underperforming.



Random Forest

The metrics for training and test sets are close, indicating good generalization. However, the recall is slightly lower compared to XGBoost, which might suggest a slight room for improvement.

Business Recommendation

Credit Approval Process Optimization

- Use the XGBoost model to speed up credit approval and improve accuracy in assessing creditworthiness
- Decline high-risk applications and impose stricter terms on medium-risk applicants to reduce non-performing loans (NPLs)

Personalized Product Offers

- Segment customers by risk profile. Offer aggressive credit products to low-risk customers and stricter terms or safer products to high-risk customers
- Provide competitive interest rates for low-risk customers and higher rates for high-risk customers to better price the risk

Credit Portfolio Management

- Use the model to regularly assess the risk of the existing credit portfolio and take early action on changing risk profiles
- Develop proactive policies to manage customers with increasing risk, such as offering loan restructuring programs

New Product Development

- Develop new credit products tailored to different risk segments, like microloans for low-risk customers
- Offer credit insurance to high-risk customers to mitigate the bank's risk

Customer Relationship Management (CRM)

- Identify low-risk customers at risk of leaving and offer incentives like increased credit limits or loyalty programs to retain them
- Proactively manage customers showing signs of financial distress before they default, such as offering loan restructuring or financial counseling

THANK YOU