EXECUTIVE SUMMARY:

This project builds a machine learning model to predict whether a customer is likely to default on a loan within the next two years. Using financial and demographic data, the model helps assess credit risk, which can support better lending decisions. The workflow includes data cleaning, feature scaling, model training (Random Forest), evaluation using ROC AUC, and visualizing feature importance. This end-to-end solution reflects real-world applications in banking and risk management.

PROJECT OVERVIEW:

Problem: Identify high-risk customers likely to default within the next two years.
Goal: Build a machine learning model that accurately predicts loan default and previous interpretable insights into what features drive those predictions.
**Tools:** Python,Scikit-learn, pandas, Matplotlib, Random Forest, ROC AUC

METHODOLOGY:

- **Data Loading**
  Imported the credit risk dataset containing customer financial and demographic information.
- **Data Cleaning**
  Handled missing values using median imputation and removed irrelevant columns.
- **Feature Scaling**
  Standardized numeric features to prepare for machine learning.
- **Train-Test Split**
  Split the data into training and testing sets (70/30) with stratified sampling.
- **Model Training**
  Trained a Random Forest classifier to predict the likelihood of loan default.
- **Model Evaluation**
  Evaluated model performance using ROC AUC, confusion matrix, and classification report.
- **Feature Importance**
  Identified key features contributing to predictions to support business decisions.

RESULTS:

The Random Forest model achieved a ROC AUC score of ~0.84, indicating strong predictive performance. Key predictors of default include: Number of past-due payments, Revolving credit utilization, Age, and Monthly income.

Feature Importance
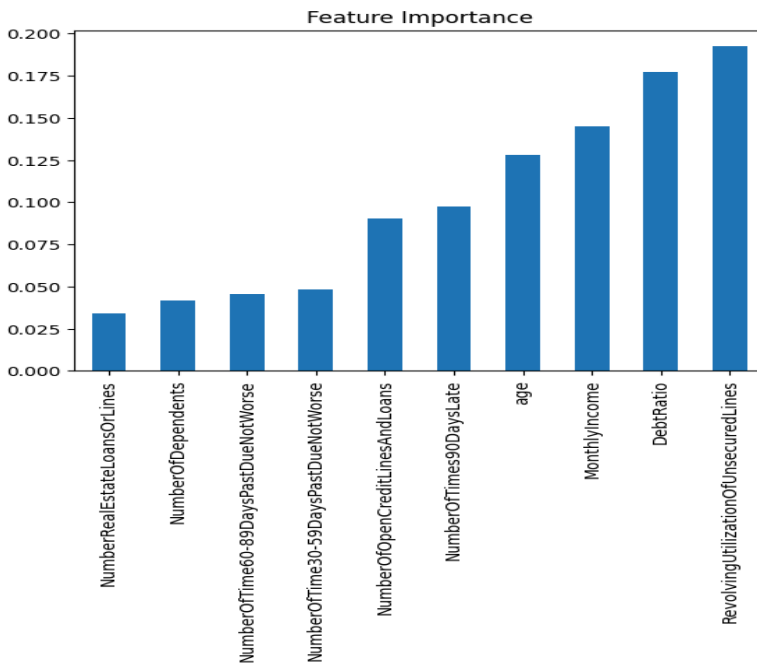
BUSINESS RECOMMENDATIONS:

1. Tighten lending criteria for applicants with high past-due payment history and high credit utilization.
2. Prioritize loan approvals for customers with steady income, lower debt ratios, and clean payment histories.
3. Use this model to automate initial credit risk screening and flag high-risk applicants for manual review.
4. Integrate model insights into customer risk dashboards for loan officers and risk managers.

BUSINESS RELEVANCE:

This project simulates a real-world application in financial services where accurate credit risk prediction can:

1. Improve loan approval decisions
2. Reduce default-related losses
3. Enhance compliance with regulatory requirements.