Financial Risk Analysis: Predictive Classification Model for Loan Eligibility

Project Overview

The objective of the Financial Risk Analysis project is to identify the factors influencing loan eligibility based on a customer's financial and demographic data, and customer segmentation in a financial institution. A comprehensive dataset containing customer demographics and transactions is used to run exploratory analyses to identify patterns and significant factors affecting customer behaviors. With the application of data analytics and classification model techniques like **exploratory data analysis**, **logistic regression**, **collaborative filtering**, **decision trees, naive bayes, and random forests,** the project aims to predict loan approval outcomes. These models will enhance decision-making by automatically classifying customers based on their likelihood of loan approval or default.

Step 1: Identifying Relevant Data Sets

For this project, we will use the **Financial Risk Dataset**, which comprises a rich set of 50 features related to individuals' personal, financial, and credit-related attributes. This dataset includes key variables such as *Credit Score, Annual Income, Loan Amount, Employment Status, Marital Status, and Debt-to-Income Ratio,* among others, which are crucial for assessing financial risk and determining loan eligibility. Additionally, it contains a binary target variable *(**`LoanApproved`**)*, indicating whether a loan was approved (1) or rejected (0).

Description of Variables in **Financial Risk Dataset**

Column Name	Definition
CreditScore	Credit score of the individual (300 to 850)
AnnualIncome	Annual income of the individual (in dollars)
LoanAmount	Requested loan amount (in dollars)
LoanDuration	Duration of the loan (in years)
Age	Age of the individual (in years)
EmploymentStatus	Employment status of the individual (Employed, Unemployed, Self-Employed)
MaritalStatus	Marital status of the individual (Single, Married, Divorced, Widowed)
NumberOfDependents	Number of dependents

EducationLevel	Education level of the individual (High School, Associate, Bachelor, Master, Doctorate)
HomeOwnershipStatus	Home ownership status (Own, Rent, Mortgage, Other)
MonthlyDebtPayments	Monthly debt payments (in dollars)
CreditCardUtilizationRate	Credit card utilization rate (0 to 1)
NumberOfOpenCreditLines	Number of open credit lines
NumberOfCreditInquiries	Number of credit inquiries in the last 6 months
DebtToIncomeRatio	Debt-to-income ratio (0 to 1)
BankruptcyHistory	Whether the individual has a bankruptcy history (0: No, 1: Yes)
LoanPurpose	Purpose of the loan (Home, Auto, Education, Debt Consolidation, Other)
PreviousLoanDefaults	Whether the individual has defaulted on a previous loan (0: No, 1: Yes)
InterestRate	Interest rate for the loan (0.01 to 0.3)
PaymentHistory	Years of payment history
SavingsAccountBalance	Savings account balance (in dollars)
CheckingAccountBalance	Checking account balance (in dollars)
InvestmentAccountBalance	Investment account balance (in dollars)
RetirementAccountBalance	Retirement account balance (in dollars)
EmergencyFundBalance	Emergency fund balance (in dollars)
TotalAssets	Total assets (in dollars)
TotalLiabilities	Total liabilities (in dollars)
NetWorth	Net worth (in dollars)
LengthOfCreditHistory	Length of credit history (in years)
MortgageBalance	Mortgage balance (in dollars)

RentPayments	Monthly rent payments (in dollars)
AutoLoanBalance	Auto loan balance (in dollars)
PersonalLoanBalance	Personal loan balance (in dollars)
StudentLoanBalance	Student loan balance (in dollars)
UtilityBillsPaymentHistory	Payment history for utility bills (0 to 1)
HealthInsuranceStatus	Health insurance status (Insured, Uninsured)
LifeInsuranceStatus	Life insurance status (Insured, Uninsured)
CarInsuranceStatus	Car insurance status (Insured, Uninsured)
HomeInsuranceStatus	Home insurance status (Insured, Uninsured)
OtherInsurancePolicies	Number of other insurance policies
EmployerType	Type of employer (Private, Public, Self-Employed, Other)
JobTenure	Job tenure (in years)
MonthlySavings	Monthly savings (in dollars)
AnnualBonuses	Annual bonuses (in dollars)
AnnualExpenses	Annual expenses (in dollars)
MonthlyHousingCosts	Monthly housing costs (in dollars)
MonthlyTransportationCosts	Monthly transportation costs (in dollars)
MonthlyFoodCosts	Monthly food costs (in dollars)
MonthlyHealthcareCosts	Monthly healthcare costs (in dollars)
MonthlyEntertainmentCosts	Monthly entertainment costs (in dollars)
LoanApproved	Whether the loan was approved (0: No, 1: Yes)

Step 2: Data Cleansing and Transformation

The project involves intensive data cleaning and transformation to ensure the readiness of the dataset for analysis. The steps include:

- Handling missing data by removing or replacing missing values.
- Normalization and standardization of data
- Data transformation by converting categorical variables to numerical representations
- Dimension reduction using Principal Components Analysis

Step 3: Pattern Discovery

In this step of the Financial Risk Analysis project, we aim to uncover meaningful patterns and trends in the dataset through predictive analysis. Our primary focus is to identify the following insights:

1. **Factors Influencing Loan Approval**:
    - Analyze the impact of various features (such as income level, credit score, employment status, debt-to-income ratio, and past loan repayment history) on loan approval outcomes using multiple data analytics algorithms. These analysis will quantify the significance of each factor and help identify high-risk customer profiles.
2. **Risk Factors Identification**:
    - Employ decision trees and random forests to identify key variables that contribute to loan defaults. This approach will allow us to visualize the decision-making process and understand which features are most predictive of risk.
3. **Predictive Modeling**:
    - Build and validate different classification models (logistic regression, naive Bayes, decision trees, and random forests) to predict loan approval outcomes. Each model will be evaluated based on performance metrics such as accuracy, precision, recall, and the ROC-AUC score.
4. **Behavioral Patterns**:
    - Explore customer behavior patterns related to loan applications, such as application frequency, loan amount requested, and repayment behaviors. Understanding these patterns will aid in predicting future behaviors and adjusting lending strategies accordingly.
5. **Visualization of Findings**:
    - Create visualizations (e.g., heat maps, bar charts, and scatter plots) to effectively communicate the insights discovered from the data. Visual representation of patterns and relationships will enhance the understanding of stakeholders.

By focusing on these areas, we aim to derive actionable insights that will inform lending strategies, improve risk assessment, and ultimately enhance the financial institution's decision-making processes. This comprehensive analysis will not only contribute to a better

understanding of customer behaviors but also support the development of tailored financial products that meet the diverse needs of the customer base.

Step 4: Visualization

The visualization includes the graphical representation of different variables in the table. The visualization includes but is not limited to:

- Distribution Plots
- Risk Group Visualization
- Transaction Trends

Step 5: Predictive Modeling

The next step includes using multiple predictive modeling algorithms to ensure the robustness and accuracy of the prediction. Following are 5 predictive models we will be leveraging in this project:

- Logistic Regression
- Linear Regression
- Decision Tree and Random Forest
- Collaborative Filtering
- Naive Bayes

Conclusion

This document outlines the steps required to identify the relevant data set for the project, clean/ transform the database, and analyze patterns in the dataset to provide recommendations and actionable insights that can help the financial institution analyze the factors influencing loan eligibility based on a customer's financial and demographic data