**GBUS-738 FINAL PROJECT**

**George Mason University-Fall 2023**

**Krishna Reddy- G01398102**

**Introduction:**

As we become proficient in data analysis and modelling our purpose of data analysis becomes more profound and objective to get the most of real-life insights that help customers and businesses. Analyzing bank personal loans is important because it helps lenders assess borrower risk, set appropriate interest rates, comply with regulations, prevent fraud, and tailor loan terms to individual needs. It also aids in portfolio management, economic assessment, risk mitigation, and informed decision-making based on historical data.

Some additional purposes and benefits of analyzing loans from national banks:

**Supporting Economic Growth**: National banks play a crucial role in facilitating economic growth by providing loans to individuals and businesses. By analyzing loans, they can ensure that credit is available to those who need it, which, in turn, supports economic development and job creation.

**Assessing Economic Health**: Loan analysis can serve as an economic indicator. The volume and type of loans being extended can provide insights into the overall financial health of a nation. For example, an increase in business loans might indicate growing entrepreneurial activity.

**Credit Market Stability:** Monitoring loans and their performance helps national banks assess the stability of the credit market. It allows them to detect signs of credit bubbles or overheating, which can lead to financial crises if left unaddressed.

**Consumer Protection**: By analyzing loans, national banks can ensure that lending practices are fair and transparent, protecting consumers from predatory lending and other unfair practices.

**Loan** **Dataset**:

The dataset contains data on over 3,500 consumers who secured a personal loan in 2017 from a national bank.

The objective is to perform analysis in a way we can determine and implement a machine learning algorithm on how many of the customers will default the payments in the future. In essence, develop a predictive analysis model.

Below are some basic steps as to how we start initializing setting up our dataset and exploring it.

Set Your Working Directory: (Sample)

Before loading a dataset, it's a good practice to set your working directory to the folder where your dataset is located. You can use the setwd() function for this. For example:

setwd("path/to/your/dataset/folder")

Load the Dataset:

Depending on the data format, you can use specific functions to load the dataset. Here are some common examples:

CSV File:

To load a dataset from a CSV file, you can use the read.csv() function:

mydata <- read.csv("your\_dataset.csv")

Excel File:

To load a dataset from an Excel file, you can use the readxl or openxlsx package:

# Using readxl package

library(readxl)

mydata <- read\_excel("your\_dataset.xlsx")

CODE:

Setting up a directory and reading the rds file and writing it into csv file :

Summary :

A computer screen shot of a math problem

Description automatically generated

PLOT:

A graph with a bar and a number of squares

Description automatically generated with medium confidence

INSIGHTS:

There are several insights that may be made from the examination of loan default rates by various loan terms. The default percentages for a range of loan terms are displayed in the bar chart. Based on the loan period, the graph clearly shows that there are notable variations in default rates. When compared to longer-term loans, loans with shorter terms—like 36 months—generally have lower default rates. This implies that borrowers with shorter loan terms may be more likely to be able to make their loan repayments on schedule, which lowers the likelihood of default. Loans with longer durations, such as 60 months, on the other hand, show greater default rates, suggesting that borrowers with longer payback terms may be more at danger of default.

DATA ANALYSIS:

Question: How does the total credit lines impact loan default rates?\*\*

CODE:

A screen shot of a computer code

Description automatically generated

GRAPH:

A graph showing a number of blue rectangular objects

Description automatically generated

Insights:

-Increasing Trend:The loan default rates show a consistent upward trend with an increase in the total credit lines. This suggests that applicants with a higher number of credit lines are more likely to default on their loans.

- Risk Assessment: Lenders should be cautious when approving loans for applicants with a large number of existing credit lines, as they might pose a higher default risk. Proper risk assessment strategies, such as stricter eligibility criteria or closer scrutiny, could be implemented for such applicants.

Question: Is there a relationship between loan default and missed payments in the last 2 years?

CODE:

A screenshot of a computer

Description automatically generated

GRAPH:

A graph with red and blue dots

Description automatically generated

Insights:

- Higher Defaults with Missed Payments: The default rates substantially rise for applicants who have missed payments in the last 2 years. This correlation indicates a significant relationship between recent missed payments and loan defaults.

- Risk Indicator: Missed payments in the recent past serve as a strong indicator of a higher default risk. Lenders should consider this factor seriously while evaluating loan applications. Applicants with a history of missed payments should undergo thorough assessments and possibly higher interest rates to compensate for the increased risk.

Question: How does debt-to-income ratio impact loan default rates?

CODE:

A screenshot of a computer program

Description automatically generated

GRAPH:

A graph of blue bars

Description automatically generated

Insights:

- Higher Debt-to-Income, Higher Defaults: The default rates increase notably as the debt-to-income ratio rises. Applicants with a higher debt-to-income ratio are more likely to default on their loans.

- Risk Mitigation: Lenders need to be cautious when approving loans for individuals with a high debt-to-income ratio. Implementing stricter eligibility criteria, lower loan amounts, or higher interest rates for these applicants can help mitigate the risk. Additionally, financial counseling services could be offered to help applicants manage their debt effectively and reduce the default risk.

PREDICTIVE ANALYSIS:

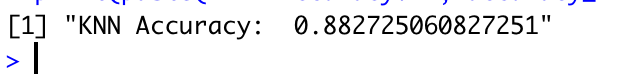
KNN:

CODE:

A screenshot of a computer program

Description automatically generated

OUTPUT:



QDA:

CODE:

A screenshot of a computer program

Description automatically generated

OUTPUT:

A screenshot of a computer code

Description automatically generated

LDA:

CODE:

A computer code with text

Description automatically generated with medium confidence

OUTPUT:

A screenshot of a computer code

Description automatically generated

**INSIGHTS**:

1. The KNN model provides a practical approach for lenders to predict loan defaults accurately. By considering multiple factors, including income and debt ratios, lenders can proactively manage potential risks.

2. By employing both QDA and LDA, lenders can assess the performance of different classification techniques. This comparison aids in selecting the most appropriate method for their specific lending scenario.

3. The dataset showcases a diverse range of borrowers with various loan purposes and homeownership statuses. Understanding these profiles is essential for tailoring lending strategies and products to meet diverse customer needs.

4. Utilizing these predictive models, lenders can implement risk mitigation strategies, adjust interest rates based on risk profiles, and make informed lending decisions. Additionally, analyzing default rates across different loan terms offers valuable insights into borrower behavior over varying repayment periods.

RECOMMENDATIONS TO IMPROVE:

Recommendation 1: Banks can use the diverse borrower profiles identified in the dataset to create targeted marketing strategies and tailored loan products. Understanding the specific needs of different customer segments allows for personalized offerings, leading to higher customer satisfaction and reduced default risks.

Recommendation 2.: Utilize the insights from predictive models like KNN, QDA, and LDA to implement risk-based pricing. Adjust interest rates and loan terms based on borrowers' risk profiles. Higher-risk borrowers could be offered loans with slightly higher interest rates, providing compensation for the increased risk.

Recommendation 3: Incorporate the predictive models into the bank's credit underwriting process. By considering various factors like income, debt ratios, and borrower profiles, banks can make more informed lending decisions. Strengthening the underwriting process ensures that loans are granted to individuals who are more likely to repay, minimizing defaults.

Recommendation 4.: Implement proactive default management strategies. Early identification of potential defaulters using predictive models allows the bank to reach out to these customers. Banks can offer financial counseling, temporary payment adjustments, or customized repayment plans to help borrowers avoid default. Timely intervention can significantly reduce default rates.

**SUMMARY**:

In summary, these insights highlight critical relationships between specific applicant attributes and loan default rates. Lenders can utilize this information to make informed lending decisions, thereby minimizing default risks and ensuring a healthier loan portfolio.