**Project Goal:**

The goal of this project is to create and train a model(s) using a technique that would provide the best loan decision to approve or reject customer loan applications given the following scenarios:

Scenario 1: Approve or reject customer loan requests with the objective of maximizing the banks total return after 5 years at a fixed 4.25% interest rate. The total capital to be distributed is $1.4B.

Scenario 2: Approve or reject customer loan requests with the objective of maximizing the banks total return after 5 years at a fixed 4.25% interest rate. The total capital to be distributed is $450M.

Scenario 3: Approve or reject customer loan requests with the objective of maximizing the banks total return after 5 years at a customer offered interest rate. The total capital to be distributed is $1.4M.

**Details of Modeling Strategy**

**Step 1: From loss column:derived two new column**

Default: if loss > 0 then default = 1. If loss = 0 then default = 0.

Fraction Loss: the percentage of loss if the customer default. The calculation is: loss/100

**Step 2: Choose variables for models**

**Approach : Feature selection -** this was done by using Supervised learning method of lasso regression for the loss fraction prediction model . the top 91 variables were selected in the process on the basis of their absolute coefficient values being greater than zero.

And used stepwise elimination to choose variables for classifying **default** status of customer.

**Dimensionality Reduction :**The important variables , were then reduced using principal component analysis

**Step 3: Build models**

loss fraction prediction model model is regression model predicting the Fraction of loss to the bank for each customer . the model method was Random forest regression via the caret package in R .

Loan default model is a classification model predicting the probability of default or not in the form of labels as - defaulter and non- defaulters . Random forest classification is used for this model.

The output prediction of both models were then used combined to calculate the **expected loss** and the **expected gain**. The delta between the opportunity and Risk is then used for solving the Scenario 1 , 2 and 3

**Step 4: Use the models to predict**

We use the models that we built in step 3 to predict the following:

1. Probability of default using the first model.
2. Fraction of loss using the second model.
3. Expected loss = Amount of loan \* probability of default \* fraction of loss
4. Expected gain = Amount of loan \* (1 – probability of default) \* interest rate \* 5

**Step 5: Decisions in 3 scenarios**

**Scenario 1**

The interest rate is 4.32%. The bank has $1.4 billion of capital. input the results from both models and use these values to calculate following. First ,All customers observations are sorted on basis of the Profit Loss Ratios in a descending order(high to low). Next , the Cumulative Requested Loan for all customers is being updated to check against the total capital available with the bank (varying for all three scenarios).

* **Expected Loss** = Amount of loan \* Probability of default \* Fraction Loss
* **Expected Gain** = Amount of loan \* (1 – probability of default) \* Interest rate \* 5
* **Profit Loss Ratio** = Expected Gain / Expected Loss
* **Loan Approval = Profit Loss Ratio > 1**
* Cumulative Requested Loan values < $1.4 billion
* **S1\_approval** column is created 1 = approved & 0= rejected.

**Scenario 2**

The interest rate is 4.32%. The bank has $450 million of capital. impute the data in the provided dataset and calculate following things:

* **Expected Loss** = Amount of loan \* Probability of default \* Fraction Loss
* **Expected Gain** = Amount of loan \* (1 – probability of default) \* Interest rate \* 5
* **Profit Loss Ratio** = Expected Gain / Expected Loss

All customers observations are sorted on basis of the Profit Loss Ratios in a descending order(high to low). will only approve the loans to customers who have the Profit Loss Ratio > 1 and the Cumulative Requested Loan values are less than $450 million.

**Scenario 3**

Each customer offers their own interest rate. The bank has $1.4 billion in capital. impute the data in the provided dataset and calculate following things:

* **Expected Loss** = Amount of loan \* Probability of default \* Fraction Loss
* **Expected Gain** = Amount of loan \* (1 – probability of default) \* Interest rate offered by customer \* 5
* **Profit Loss Ratio** = Expected Gain / Expected Loss

all customers based on their Profit Loss Ratios from high to low and create the Cumulative Requested Loan for all customers. only approve the loans to customers who have the Profit Loss Ratio > 1 and the Cumulative Requested Loan values are less than $1.4 billion.