# Module 12 -Credit Risk Classification Report

The purpose of this challenge was to build a prediction model that would be able to analyze credit worthiness of borrowers for a bank and to identify potential loan defaults.

The dataset provided by the bank contained 77,536 entries with seven key indicators. These are listed as follows:

Size of the loan

Interest rate on the loan

Income

Debt\_to\_Income ratio

number of accounts

Derogatory marks on the borrower's credit report

total debt

loan status indicating a healthy loan with a 0 and a loan in default with a 1.

Based on the value counts of loan status ,there are 2500 out of 75036 in default.

The features and target variable were spearated , to split data in Train,test data respectively

To complete this analysis, the dataset was first split into training and testing datasets. After instantiating a logistic regression model from scikit-learn, the training data was fit to the model. Predictions of the test data were compared to the given labels.

To further analyse the dataset , Decision Tree and Randon Tree Forest were also used to classify .

## Results

\* Machine Learning Model 1:

\* Confusion Matrix for Logistic Regression

Predicted 0 Predicted 1

Actual 0 18673 86

Actual 1 32 593

Accuracy Score : 0.993912505158894

Classification Report

precision recall f1-score support

0 1.00 1.00 1.00 18759

1 0.87 0.95 0.91 625

accuracy 0.99 19384

macro avg 0.94 0.97 0.95 19384

weighted avg 0.99 0.99 0.99 19384

\* Machine Learning Model 2:

\* Confusion Matrix for Decision Tree

Predicted 0 Predicted 1

Actual 0 18681 78

Actual 1 119 506

Accuracy Score : 0.9898369789517127

Classification Report

precision recall f1-score support

0 0.99 1.00 0.99 18759

1 0.87 0.81 0.84 625

accuracy 0.99 19384

macro avg 0.93 0.90 0.92 19384

weighted avg 0.99 0.99 0.99 19384

\* Machine Learning Model 3:

\* Confusion Matrix for Random Tree Forest

Predicted 0 Predicted 1

Actual 0 18680 79

Actual 1 72 553

Accuracy Score : 0.9922100701609575

Classification Report

precision recall f1-score support

0 1.00 1.00 1.00 18759

1 0.88 0.88 0.88 625

accuracy 0.99 19384

macro avg 0.94 0.94 0.94 19384

weighted avg 0.99 0.99 0.99 19384

## Summary

Summarize the results of the machine learning models, and include a recommendation on the model to use, if any.

\* All the three models perform well in terms of accuracy . Out of the three Logistic Regression amd Random Tree Forest seem to perform better .

Precision is good for both , but Logistic Regression model does a better job in predicting the Actual -1 . So for this dataset , its recommended.

\* Since both models correctly predict healthy loans, it is important to have a model that correctly classifies more high-risk loans, even. The risk associated with misclassifying a healthy loan as high-risk is lower than the risk of not classifying a high-risk loan as such.