

machine learning model PREDICTing CREDIT CARD DEFAULTS

CAPSTONE PROJECT

August 18, 2023

PRESENTED BY

SAMSON OGWUCHE AND CHARLES NTAMACK

Table of Contents

[**Executive Summary** 3](#_Toc143454984)

[**Introduction** 4](#_Toc143454985)

[Background 4](#_Toc143454986)

[Problem Statement 4](#_Toc143454987)

[Objectives and Metrics 5](#_Toc143454988)

[Assumptions and Limitations 5](#_Toc143454989)

[Opportunity 6](#_Toc143454990)

[Target Client 6](#_Toc143454991)

[**Model Governance** 7](#_Toc143454992)

[**Data Sources** 8](#_Toc143454993)

[Dataset Introduction 8](#_Toc143454994)

[Exclusions 8](#_Toc143454995)

[Data Dictionary 8](#_Toc143454996)

[**Data Exploration, Preparation, and Feature Engineering** 10](#_Toc143454997)

[Variable Level Monitoring 10](#_Toc143454998)

[Model Build Variable Level Statistics (e.g., mean, median, std, or distribution of categories) 13](#_Toc143454999)

[Acceptable Ranges 15](#_Toc143455000)

[Caps & Floors 16](#_Toc143455001)

[Missing Values 18](#_Toc143455002)

[Variable Drift Monitoring Tolerance 19](#_Toc143455003)

[**Model Exploration** 19](#_Toc143455004)

[Model Monitoring, Health & Stability 19](#_Toc143455005)

[Models Building 20](#_Toc143455006)

[Decision Trees. 21](#_Toc143455007)

[Random Forest 25](#_Toc143455008)

[Model Sensitivity to Key Drivers 30](#_Toc143455009)

[Logistic Regression 31](#_Toc143455010)

[Neural Networks. 32](#_Toc143455011)

[Model Comparison 34](#_Toc143455012)

[**Initial Model Fit Statistics** 35](#_Toc143455013)

[Risk Tiering (e.g., no action, report, refit, rebuild) 36](#_Toc143455014)

[**Model Assumptions and Limitations** 37](#_Toc143455015)

[**Recommendations** 38](#_Toc143455016)

[**Conclusions** 40](#_Toc143455017)

[**References** 41](#_Toc143455018)

# **Executive Summary**

In recent times, banks and other lending institutions have witnessed a surge in consumer credit debt including credit card debt. Canada’s total consumer debt as of the end of Q1 2023 stood at $2.37 trillion while that of the United States of America (USA) stood at $17 trillion. The total credit card debt of Canada as of the end of 2022 was $100 billion while that of the USA was $1 trillion.

Our project is to construct an advanced analytics model aimed at predicting credit card default. It serves as a valuable tool for banks and credit lending institutions, offering insights into the key risk factors to consider in managing existing customers and before credit card approval decisions for new customers. By leveraging this predictive model, potential factors contributing to repayment defaults can be proactively identified and managed, contributing to more informed lending practices. This work is a pivotal component of our journey towards attaining a postgraduate certificate in Business Analytics and Insights from Centennial College's esteemed Business School, situated within the context of the Canadian financial landscape. The culmination of rigorous research and practical implementation, this capstone project demonstrates our proficiency in applying cutting-edge analytical techniques to real-world scenarios. For this endeavor, we harnessed a comprehensive dataset of credit card customers in Taiwan, thoughtfully curated from open-source repositories. With meticulous attention to detail, we meticulously navigated each stage of the model-building process, encompassing data preprocessing, feature engineering, algorithm selection, model training, and validation. This project encapsulates the essence of modern analytics practices and their application to a critical financial domain. Through this document, we aim to articulate the intricate nuances of our approach, unveil the transformative insights gained, and underscore the profound implications for credit risk management. This endeavor stands as a testament to our dedication, skills, and ability to contribute meaningfully to the field of business analytics, all while fulfilling the requirements for graduation from Centennial College's Business School.

# **Introduction**

## Background

Credit card default risk has become a serious concern to lenders and the financial system at large. This project recommends that banks and other credit lending institutions implement predictive analytics models to manage and mitigate credit risk. We intend to build a model that can help to predict credit card payment default by customers thereby posing credit risk to the financial system. The model will help banks and other lenders identify risk factors in credit card customers and how to manage the risk.

## Problem Statement

As Canada’s economy continues to face difficult times, the number of individuals and corporations who apply to obtain and use credit cards across the country continues to rise. According to Equifax, as of the end of Q1 2023, Canada’s total consumer debt climbed to $ 2.37 trillion as consumers lean on credit cards. In the same vein, the number of credit cards issued in Q4 increased by 15.3% year-on-year, and almost 1.4 million new credit cards were issued. The rise in total credit card debt comes with the problem of customers’ default on payment for the amount used thereby posing a credit risk to the financial industry and the economy at large.

## Objectives and Metrics

The cost of living is on the rise and the total household debt in debt stock continues to pile and the total number of Canadians with access to credit increased to about 30.6 million as of the end of Q1 2023. The number of subprime consumers who are in the risk tiers increased by 8.3% and this constituted the largest growth. Total outstanding credit balances increased by 5.6% in Q1 2023. Credit card consumer balances grew by 11.36% in Q1 2023 representing the highest growth in a long time. However, prime consumers still constitute about three-quarters of the consumers indicating a relatively healthy distribution of credit consumers. Despite this, the delinquency rate on credit balances increased by 5.99% YoY in Q1 2023 and this therefore becomes imperative to make predictions to proactively identify consumers who are potential defaulters and manage the credit risk associated with such consumers.

The metrics that will be used to define the success of this analytics project will be geared toward achieving the objective of minimizing the financial losses arising from credit card default and improving collection rates. The following will be the primary metrics for measuring the success of the project upon implementation:

* Reduction in credit card default (delinquency) rate
* Increase in credit card utilization and collection rate

## Assumptions and Limitations

It is assumed that the necessary data related to credit card defaults, customer behavior, and financial performance is available for analysis and decision-making. Sufficient and accurate data is crucial for effective risk assessment, identification of default risks, and evaluation of the impact of strategies used.

It is assumed that the financial institution has the necessary resources, including financial, technological, and human resources, to effectively address credit card defaults. A sufficient budget, staffing, and infrastructure are essential for implementing strategies and initiatives to achieve the objectives.

It is assumed that financial institutions have access to appropriate technological tools and systems to support credit risk assessment, data analysis, reporting, and customer communication. Adequate technological capabilities enable efficient and accurate implementation of strategies and initiatives.

## Opportunity

The business case for this project is to prevent credit losses associated with default on credit card consumers by proactively identifying the risk and managing them. This project will provide lending institutions with the opportunity to protect their financial assets and have more liquidity for investment in better business opportunities that meet their business strategic objectives.

## Target Client

Our target clients are financial institutions and other lenders such as auto leasing companies, mortgage lenders, grocery stores, etc.

Retail banks are credit institutions that finance the economy, and this is extremely linked to its proper functioning. A retail bank is therefore an economic player whose main activities are:

* The provision of means of payment to customers;
* Collection and administration or investment (savings) of funds deposited by clients;
* The granting of loans.

As a result of these activities, Retail banking is subject to several types of risks (credit risks, operational risks, and market risks). For this capstone project, we are going to focus on credit risks. Credit risk can be defined as the risk that the customer will no longer be able to meet the commitments he has made to the bank. This problem poses significant challenges for both cardholders and financial institutions, leading to adverse effects on personal finances, credit ratings, profitability, and economic stability. The primary issue contributing to credit card defaults is inadequate credit risk management by lenders. Addressing this problem requires implementing effective measures to improve risk assessment and mitigation strategies.

# **Model Governance**

The process of constructing predictive models encompasses a series of intricate steps, including data preparation, rigorous data quality checks, feature reduction, model building, predictive analysis, and critical examination of model outcomes.

Model monitoring refers to the continuous process of observing and tracking the performance of a model deployed to ensure that the model performance has not materially changed from the initial prediction. It ensures that the solution continues to function well and delivers accurate results in real-world scenarios. Governance involves establishing guidelines, policies, and ethical considerations for the development, deployment, and use of an analytics model. Validation on the other hand involves assessing the performance and accuracy of the model or project. The model needs to be monitored to ensure that the return on investment by the organization is maximized. Governance helps to ensure that necessary features, steps, and procedures are followed in building the model for optimal performance and reducing potential risks of the analytics model.

Validation, Monitoring, and Governance are integral parts of this capstone project to ensure that our results are reliable, effective, and responsible for the development and after its deployment. These practices contribute to building trust in our project's outcomes and contribute to its long-term success.

# **Data Sources**

## Dataset Introduction

To build the predictive analytics model that will help to predict credit card default, we sourced the dataset from open sources and obtained a credit card default dataset for 2005 in Taiwan from Kaggle. The dataset has 30,000 observations and 25 variables as below.

**Variables:** ID**,** Limit\_Bal, Sex, Education, Age, Pay\_0, Pay\_2, Pay\_3, Pay\_4, Pay\_5, Pay\_6, Bill\_Amt1, Bill\_Amt2, Bill\_Amt3, Bill\_Amt4, Bill\_Amt5, Bill\_Amt6, Pay\_Amt1, Pay\_Amt2, Pay\_Amt3, Pay\_Amt4, Pay\_Amt5, Pay\_Amt6, Default.

## Exclusions

Excluded ID from the dataset during preprocessing because it is not relevant in building our model.

## Data Dictionary

ID: Represents a unique key that identifies each customer.

PAY\_0: Repayment status in September 2005 (-1=pay duly, 1=payment delay for one month, 2=payment delay for two months,8=payment delay for eight - months, 9=payment delay for nine months and above)  
PAY\_2: Repayment status in August, 2005 (scale same as above)  
PAY\_3: Repayment status in July, 2005 (scale same as above)  
PAY\_4: Repayment status in June, 2005 (scale same as above)  
PAY\_5: Repayment status in May, 2005 (scale same as above)  
PAY\_6: Repayment status in April, 2005 (scale same as above)  
BILL\_AMT1: Amount of bill statement in September, 2005 (NT dollar)  
BILL\_AMT2: Amount of bill statement in August, 2005 (NT dollar)  
BILL\_AMT3: Amount of bill statement in July, 2005 (NT dollar)  
BILL\_AMT4: Amount of bill statement in June, 2005 (NT dollar)  
BILL\_AMT5: Amount of bill statement in May, 2005 (NT dollar)  
BILL\_AMT6: Amount of bill statement in April, 2005 (NT dollar)  
PAY\_AMT1: Amount of previous payment in September, 2005 (NT dollar)  
PAY\_AMT2: Amount of previous payment in August, 2005 (NT dollar)  
PAY\_AMT3: Amount of previous payment in July, 2005 (NT dollar)  
PAY\_AMT4: Amount of previous payment in June, 2005 (NT dollar)  
PAY\_AMT5: Amount of previous payment in May, 2005 (NT dollar)  
PAY\_AMT6: Amount of previous payment in April, 2005 (NT dollar)

# **Data Exploration, Preparation, and Feature Engineering**

## Variable Level Monitoring

Variables level monitoring involves a thorough review and monitoring of each feature in the dataset to ensure data quality assurance, feature drift detection, feature importance, model performance tracking, bias, and fairness monitoring. We monitored the quality of the variables in the dataset checking for outliers and some skewed variables. We noticed outliers and skewness in the dataset. We handled the outliers and skewness, however, six of the variables still have outliers and skewness but are reduced to the barest minimum. There were no missing values, no Null, and no duplications in the dataset.

Below are the screenshots of our variable-level monitoring checking outliers and skewness in the dataset before and after the treatment of outliers which are an essential part of variable-level monitoring in machine learning.

Figure 1: **Checking for Outliers Before Treatment**

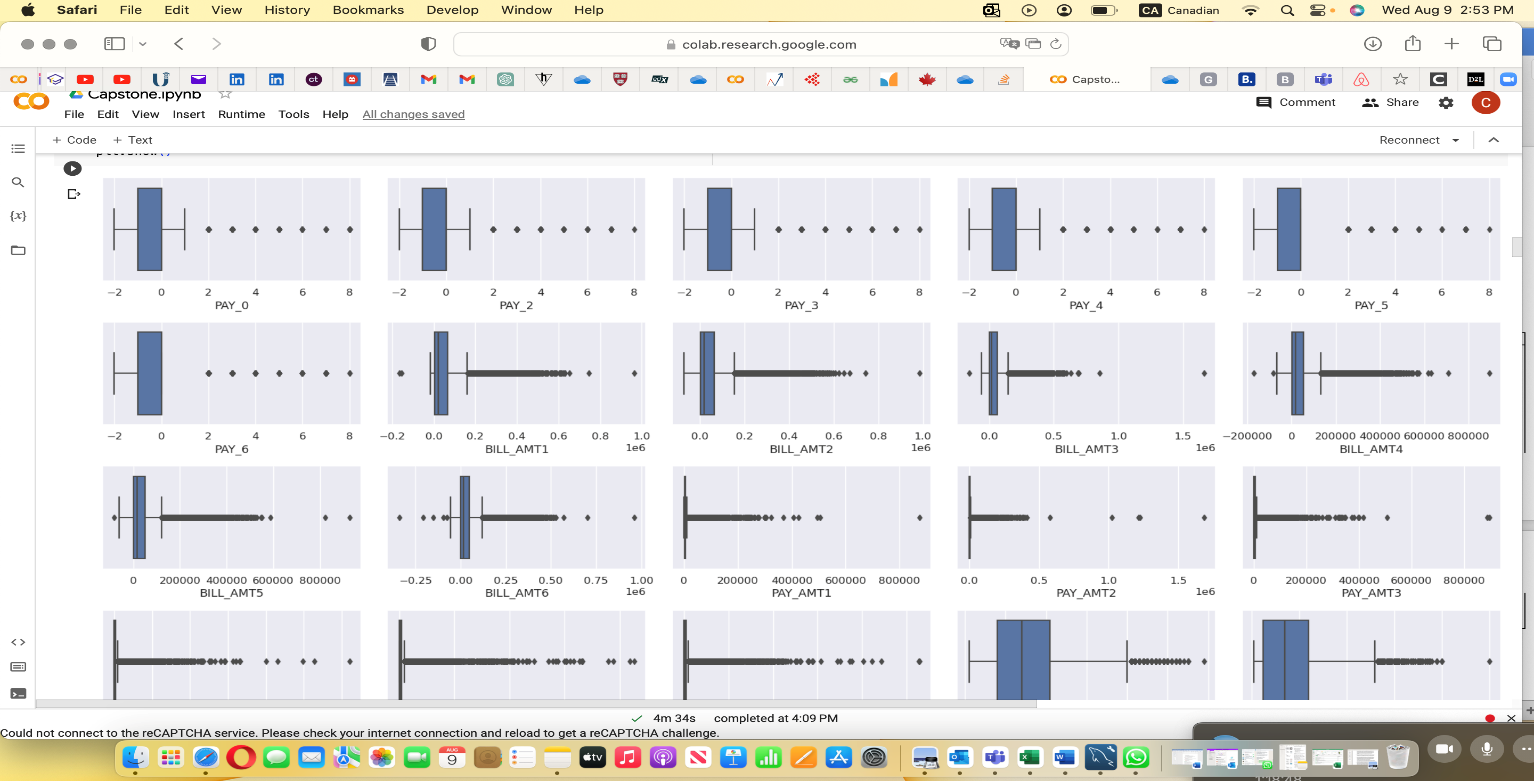


Figure 2: **Checking for Outliers After Treatment**

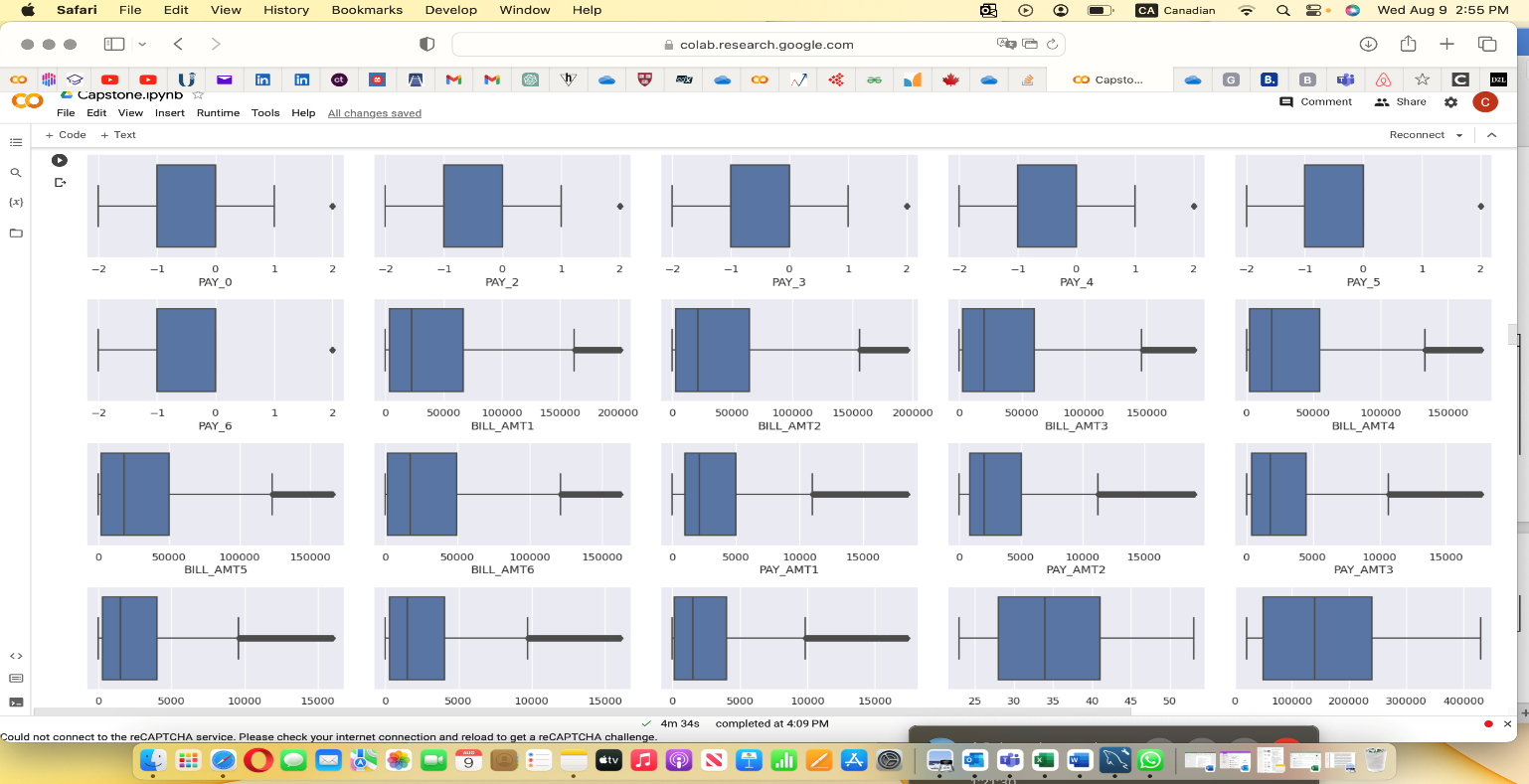


Figure 3:**Skewed Variables Before Treatment**

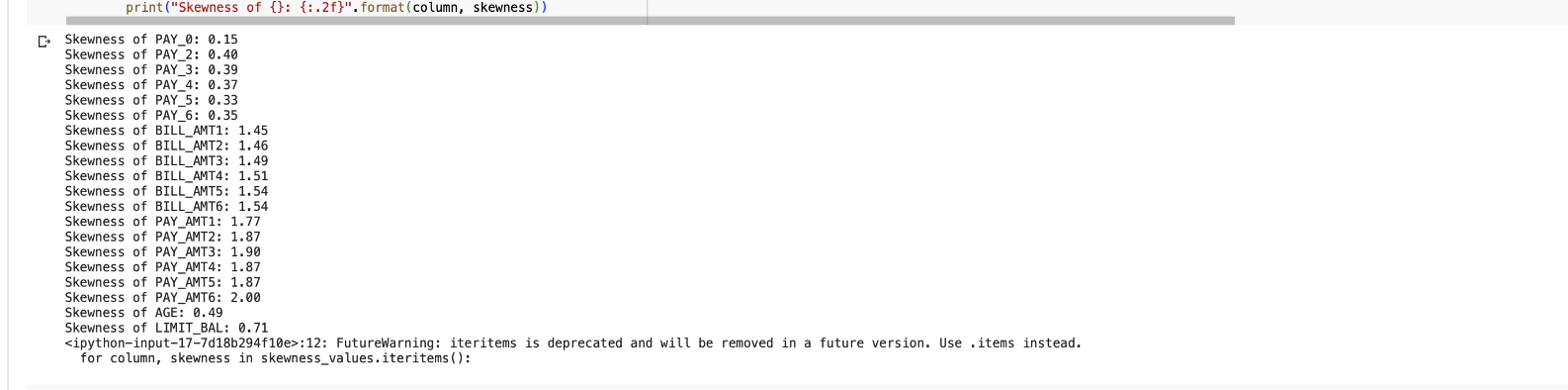
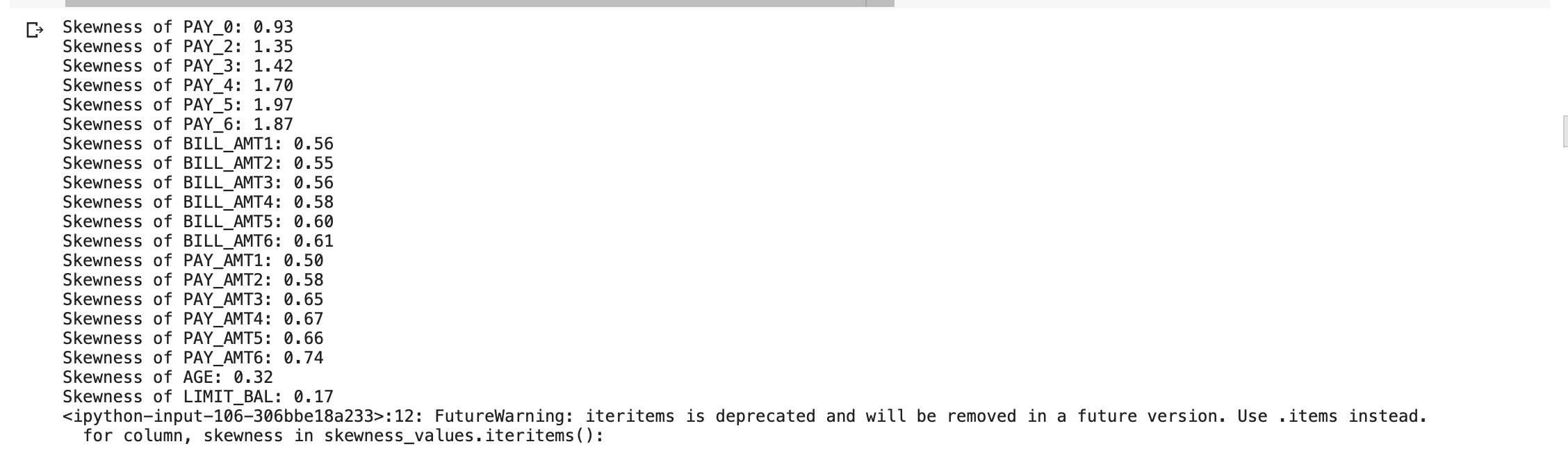


Figure 4:**Skewed Variables After Treatment**



Model Build Variable Level Statistics (e.g., mean, median, std, or distribution of categories)

Model variable statistics refers to the statistical analysis and interpretations of each variable in a dataset. This includes outlier detection, correlation, and relationship of variables. We built our variable statistics to detect the correlation between variables in the dataset using a correlation coefficient table and heatmap which show the extent of correlation between the variables. From our correlation coefficient, we did not notice the presence of any strong correlation between the variables. In our variable descriptive statistics, we have the count of 30,000 observations, mean, std, minimum, 1st quantile, 2nd quantile, 3rd quantile, and maximum values of each of the variables in the dataset. To detect the outliers in the dataset, we used a boxplot to visualize the points of deviation from the normal distribution in the dataset.

Figure 5: **Checking for Correlated Variables**

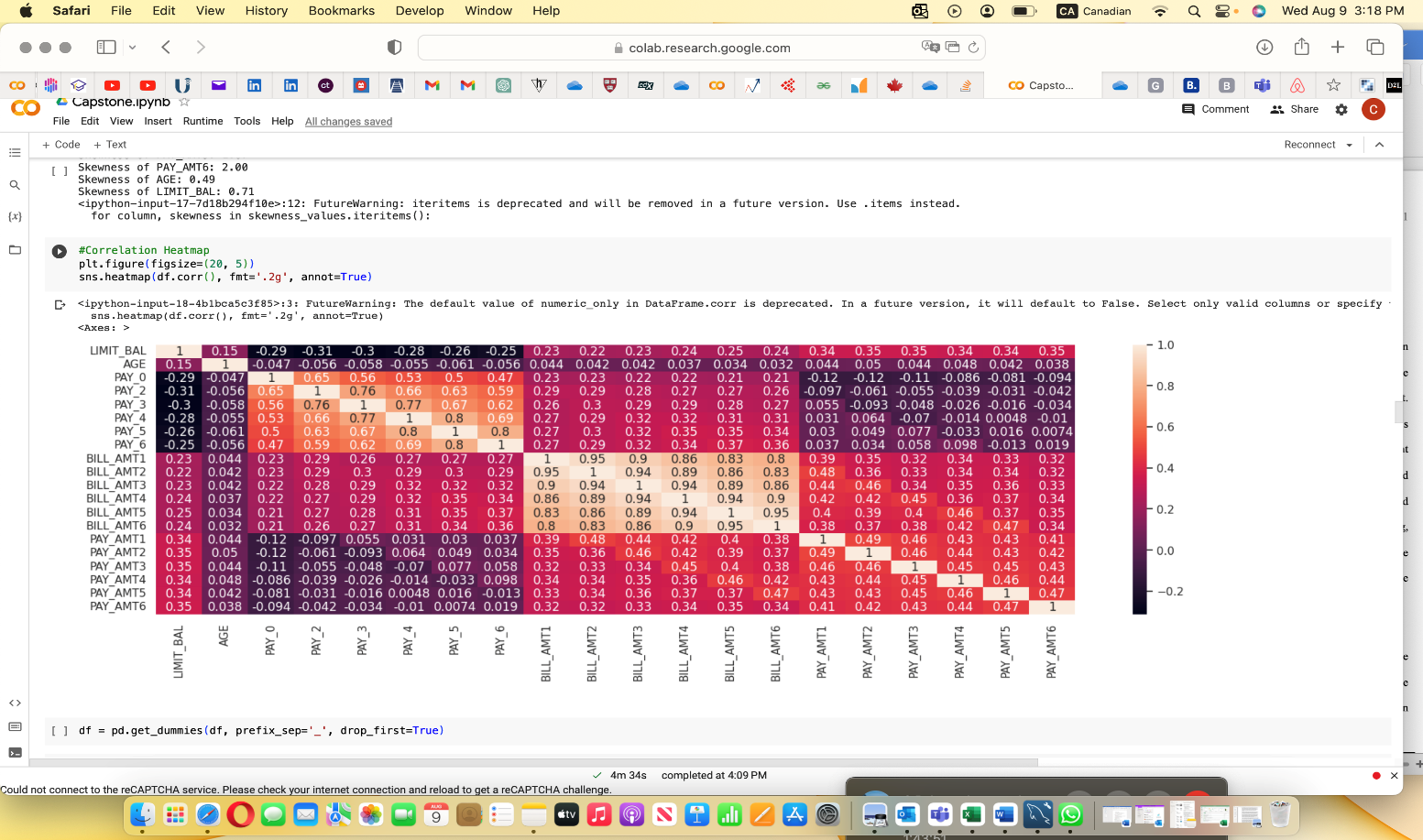
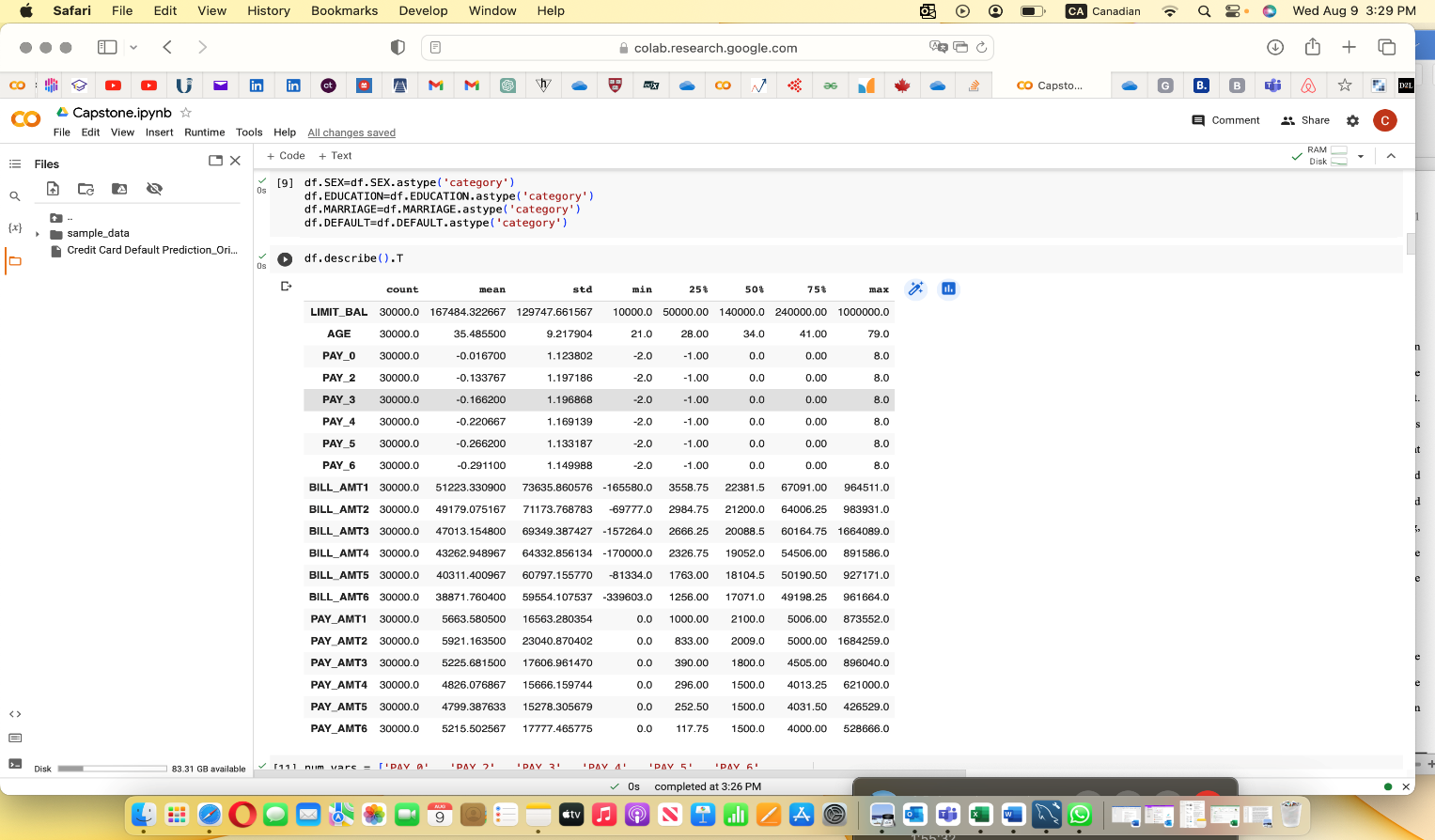


Figure 6: **Variables Summary Statistics**



Acceptable Ranges

An acceptable range is a predefined interval or boundary within which all statistical measurements are considered normal or valid. Values outside the predefined interval are outliers. These are usually handled before building a model to avoid model overfitting and inaccurate model assessment results.

In our model, we set a predefined range for variable correlation of -0.69 to +0.69 as a normal correlation range while all correlation coefficients outside the predefined correlation range above are seen as strong correlations. Variables with strong correlation coefficients should normally be dropped from the model. However, in our dataset, the strong correlation coefficients are between the same class of variables, and as such, we did not drop them.

As stated earlier, there were outliers in the dataset, however, we tried two methods to handle the outliers and chose the best out of the two. Our best method for handling the outliers is the Cap and Floor method.

Caps & Floors

Cap and Floor is a strategic technique employed to confine predicted values within a predefined and acceptable range, fostering the maintenance of data integrity and precision. This dynamic approach delimits the upper and lower extremities of observations, assuring that most values fall within an optimal 95% to 99% range for upper values and a prudent 1% to 5% interval for lower values. In our model, we capped all values in the observations above 95% at the 95 percentiles while all values below 5% are floored at the 5 percentiles. This method was used to handle the outliers in the dataset. Although after using this method, we still have six of the variables that still have outliers. We have two choices, to leave the variables or remove them from the dataset if we consider them not to be important in the model. However, we chose to leave the variables because they are variables of importance in the model.

Figure 5: **Outliers before Cap & Floor**

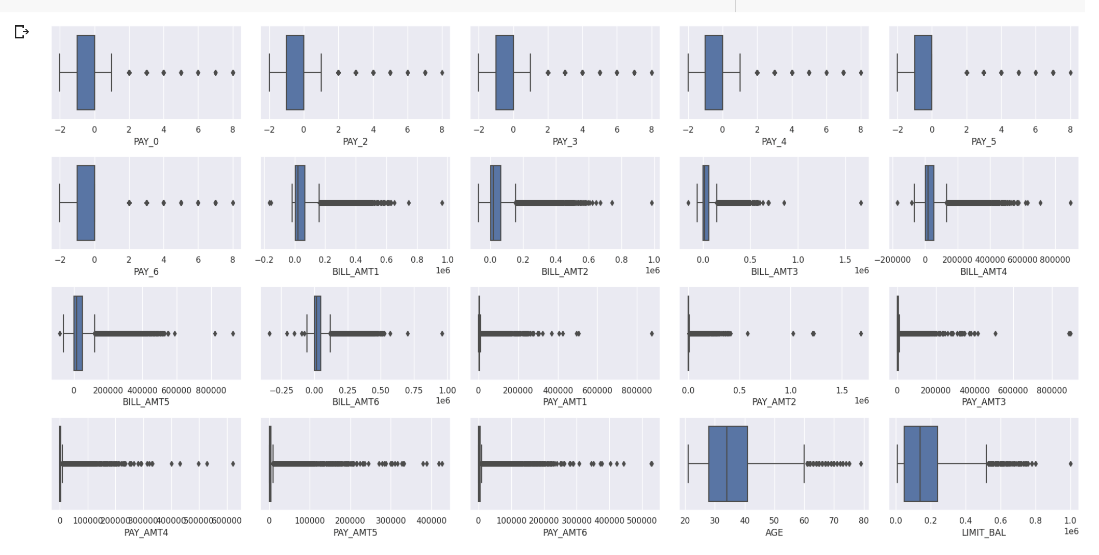
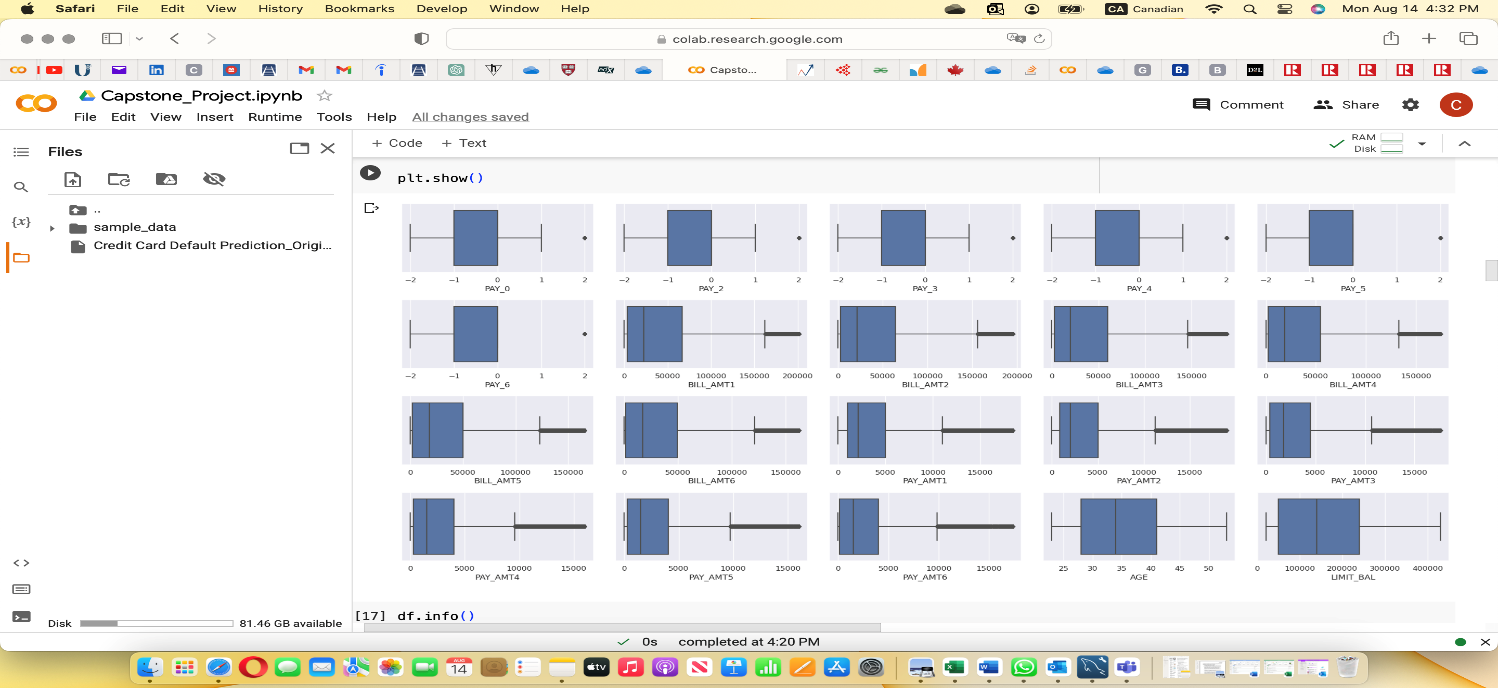
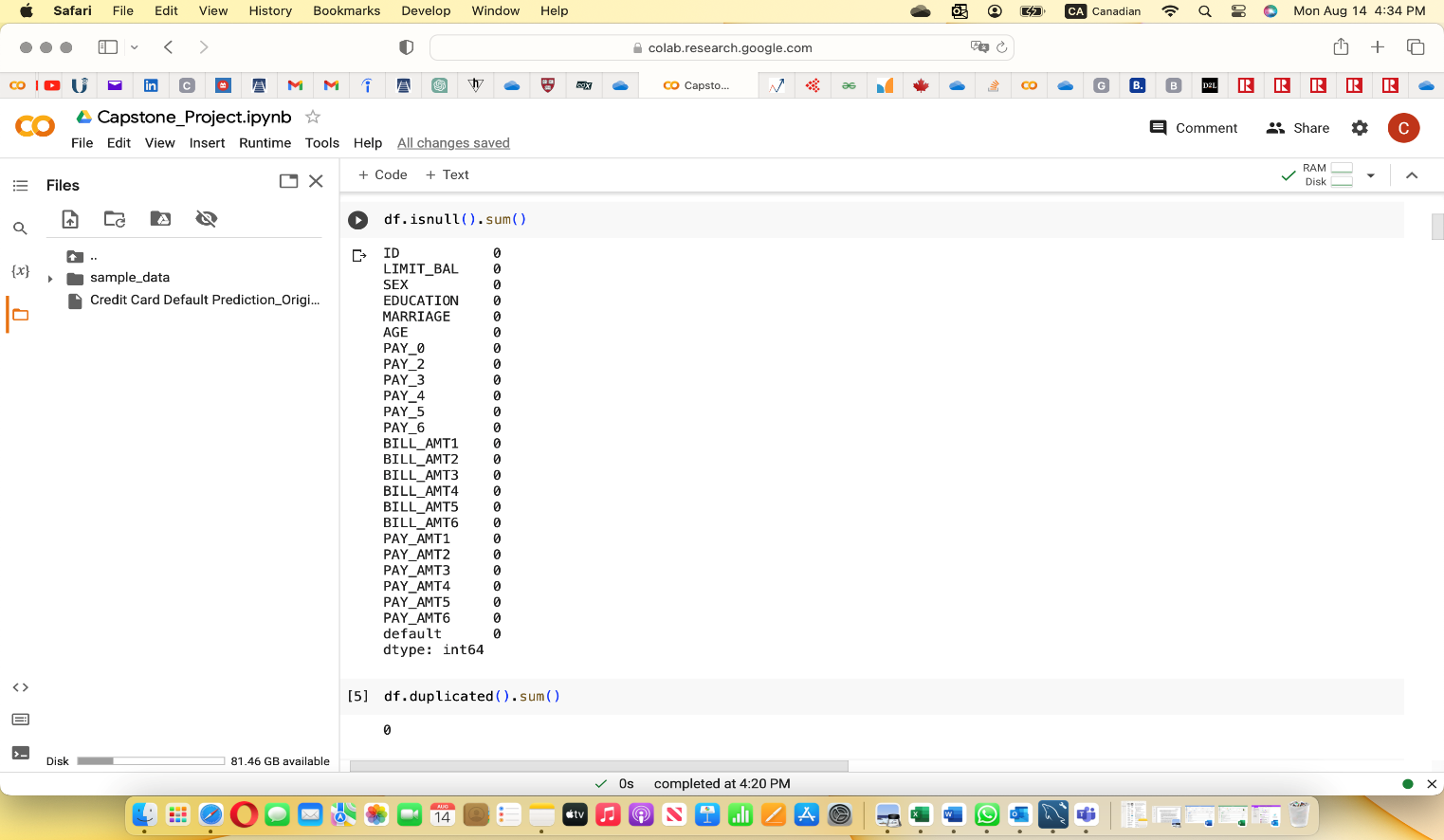


Figure 6: **Outliers After Cap & Floor**

Missing Values

Missing values occur when there are no values recorded or stored for variables in the observations. Generally, missing values in a dataset are handled by the replacement or imputation method depending on the nature of the dataset, the extent of the missing values, and the role of the analysis. The missing values can be replaced with the mean or median values of the variables without missing values. Our dataset fortunately does not have missing values. However, we would have used median values to replace missing values if there were missing values because median values are not affected by outliers.

Figure 7: **Resulting no Missing Value**



Variable Drift Monitoring Tolerance

Variable drift occurs when the statistical properties of the target variable, which the model is trying to predict change over time, and the predictions become less accurate and unreliable. Models are usually built using historical data. However, variables’ statistical properties change over time, and this will affect the accuracy of the model prediction built on historical data. Variable drift monitoring is continuous monitoring of the models’ variables’ behavior to ensure that the model accuracy remains as predicted. Variable drift monitoring tolerance is the acceptable limit within which variables’ statistical properties can drift. Any drift outside this limit is not acceptable and will require investigation or corrective action. Model drift monitoring is to ensure stable model accuracy over time.

The dataset used in this model is based on customers’ credit card history. However, some variables’ status may likely change over time thereby causing the variables and this may affect the prediction of this model in the future. For instance, the status of a customer with a variable education may change from university student to graduate in the future and the model prediction may become inaccurate. Similarly, the status of a customer under the variable marriage may change from single to married in the future and this may affect model prediction built on customer’s status as single.

We set a **3%** tolerance level for variable drift and for any drift outside this limit, the model needs to be re-run.

# **Model Exploration**

Model Monitoring, Health & Stability

Model monitoring involves closely tracking the performance of machine learning models in an operation. The goal is to detect any deviation in the model, analyze the model prediction accuracy, eliminate prediction errors, and tweak the model to ensure the best performance. Model health is the overall state and reliability of a machine learning model. Model stability is the consistency of a model’s performance over time and across different subsets of data. Monitoring model health and stability includes monitoring the data distribution drifts, model performance shifts, operation metrics, data integrity, performance by segment, and bias.

It is expected that when our model is deployed, the users will monitor the model performance against customers’ statuses to ensure that they remain the same over time, otherwise, the model needs to be updated based on the statuses of customers.

To monitor the model’s health and stability, we set a model accuracy limit of **80%.** If the accuracy goes below this limit, the model needs to be re-run.

Models Building

Model building is the process of creating and training machine learning or statistical models using a dataset to make predictions, classifications, or generate insights. It involves selecting the appropriate algorithm, preparing the data, tuning hyperparameters, and evaluating the model's performance.

In our effort to generate the best predictions for credit card default by customers, we used many algorithms to compare our prediction based on the models’ accuracies. We followed the steps below in building the models.

1. Defining the independent variables and the target variable by splitting the data into (X and Y).
2. Split the dataset into train and test (50:50)

The next step is to build and fit the models using the following algorithms.

## Decision Trees.

In building the decision trees, we chose the decision tree algorithm for classification using a decision tree classifier SKlearn. tree. To optimize the model’s performance, we performed a grid search with cross-validation to find the best hyperparameters. Hyperparameters are hyperparameters set before training that affect the model’s performance. The following are the parameters that we used for tunning the decision tree.

Max\_depth is the maximum depth of the tree.

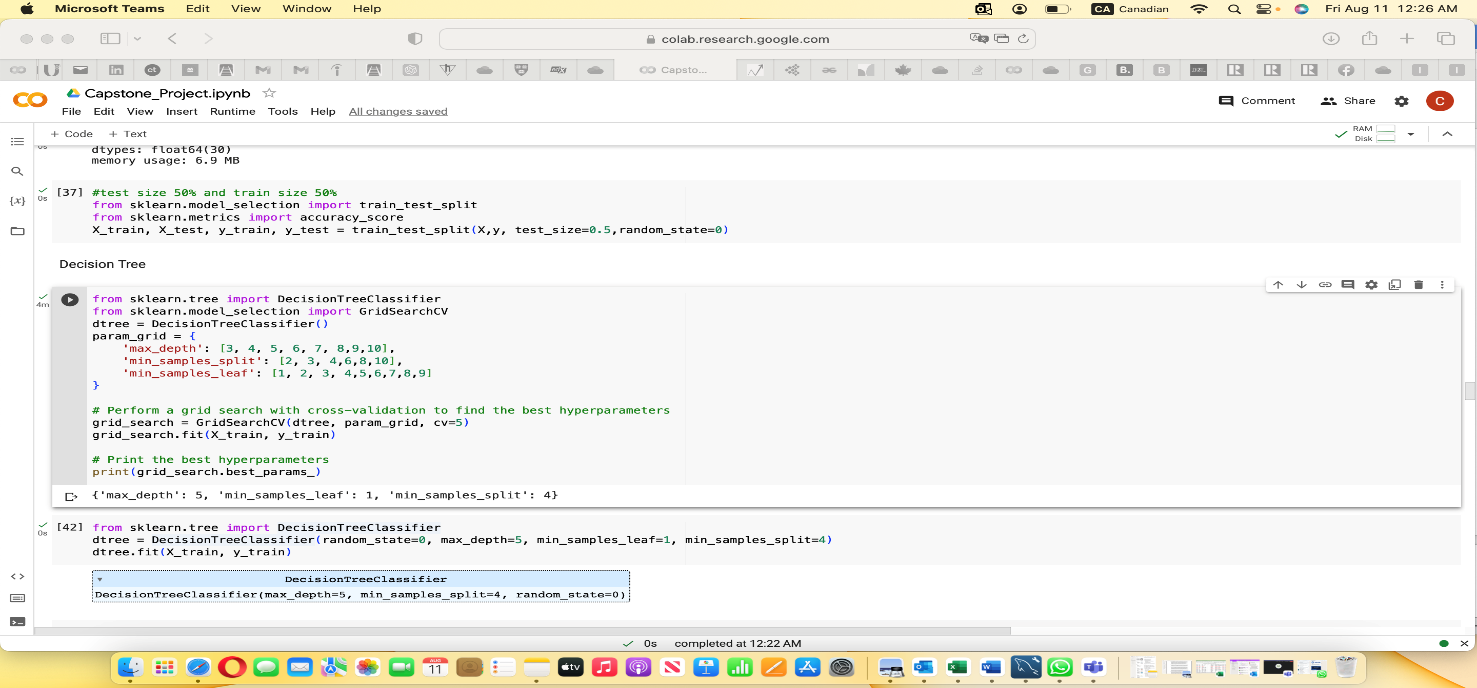
Min\_samples\_leaf is the minimum number of samples required to be at the leaf node.

Min\_samples\_split is the minimum number of samples required to split the internal node.

The random state is the model hyperparameter used to control the randomness of a machine-learning model.

Below is the screenshot of our best hyperparameters for building the decision tree.

Figure 8: **Decision Tree Hyperparameter Tuning**



After hyperparameter tuning, we trained and fit the model using the best parameters. We evaluated the model performance on the test data. Several metrics were used to evaluate how well the model performed, and below are the screenshots of some of the methods and their results.

Figure 9: **Decision Tree of the Train Data**

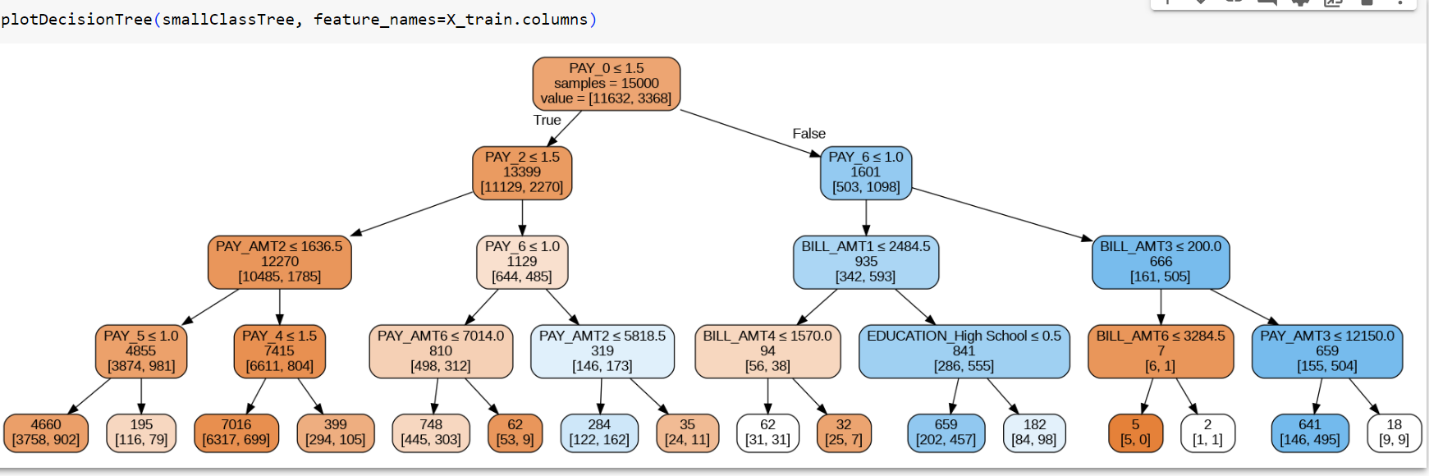


Figure 10: **Summary of Decision Tree Performance Metrics**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Model** | **Accuracy Score** | **Precision** | **F1 Score** | **Recall** | **ROC/AUC** |
| **Decision Tree** | **81.80%** | **81.8%** | **81.8%** | **81.8%** | **75.12%** |

Figure 11: **Fit Statistics for Decision Tree**

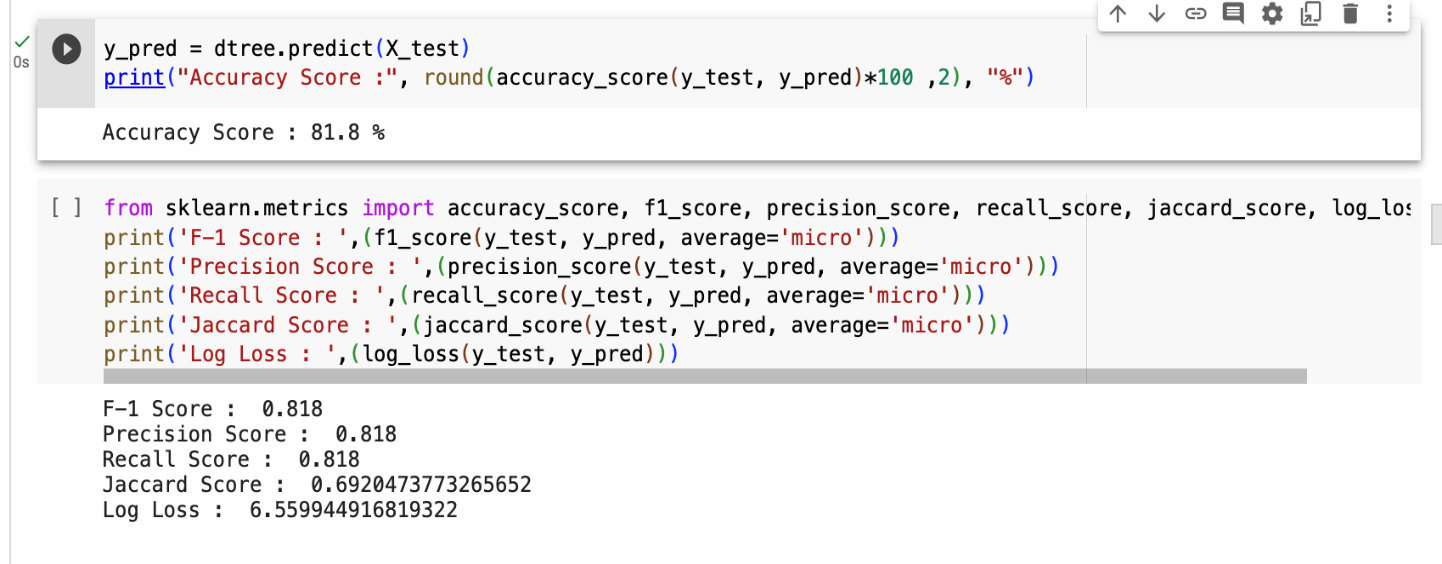
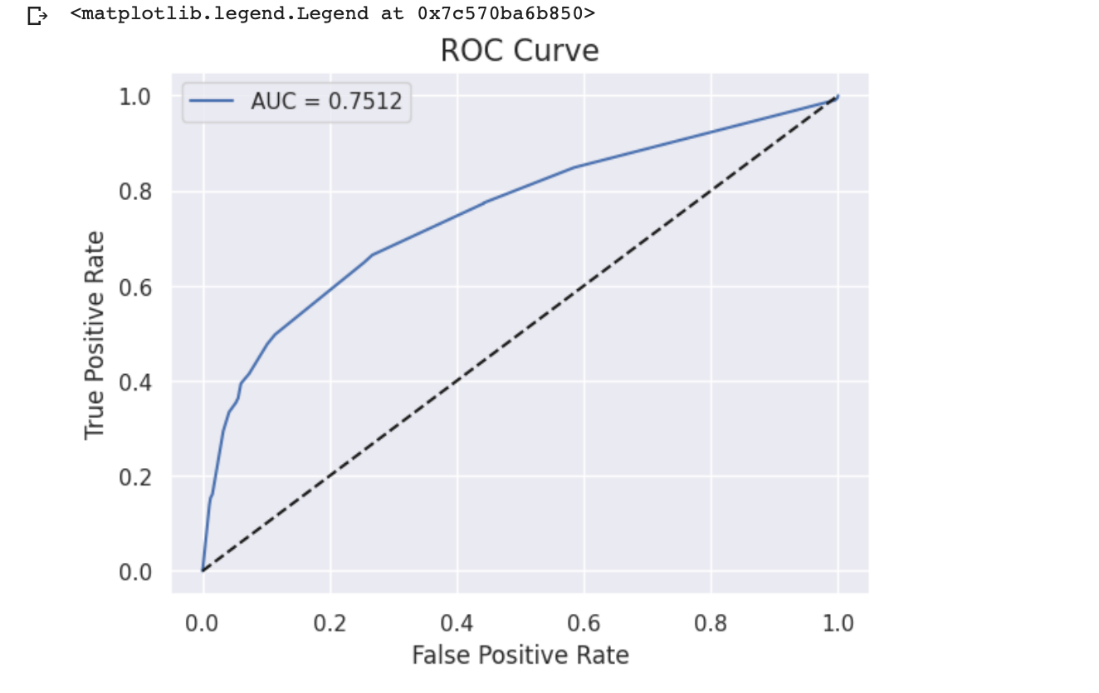


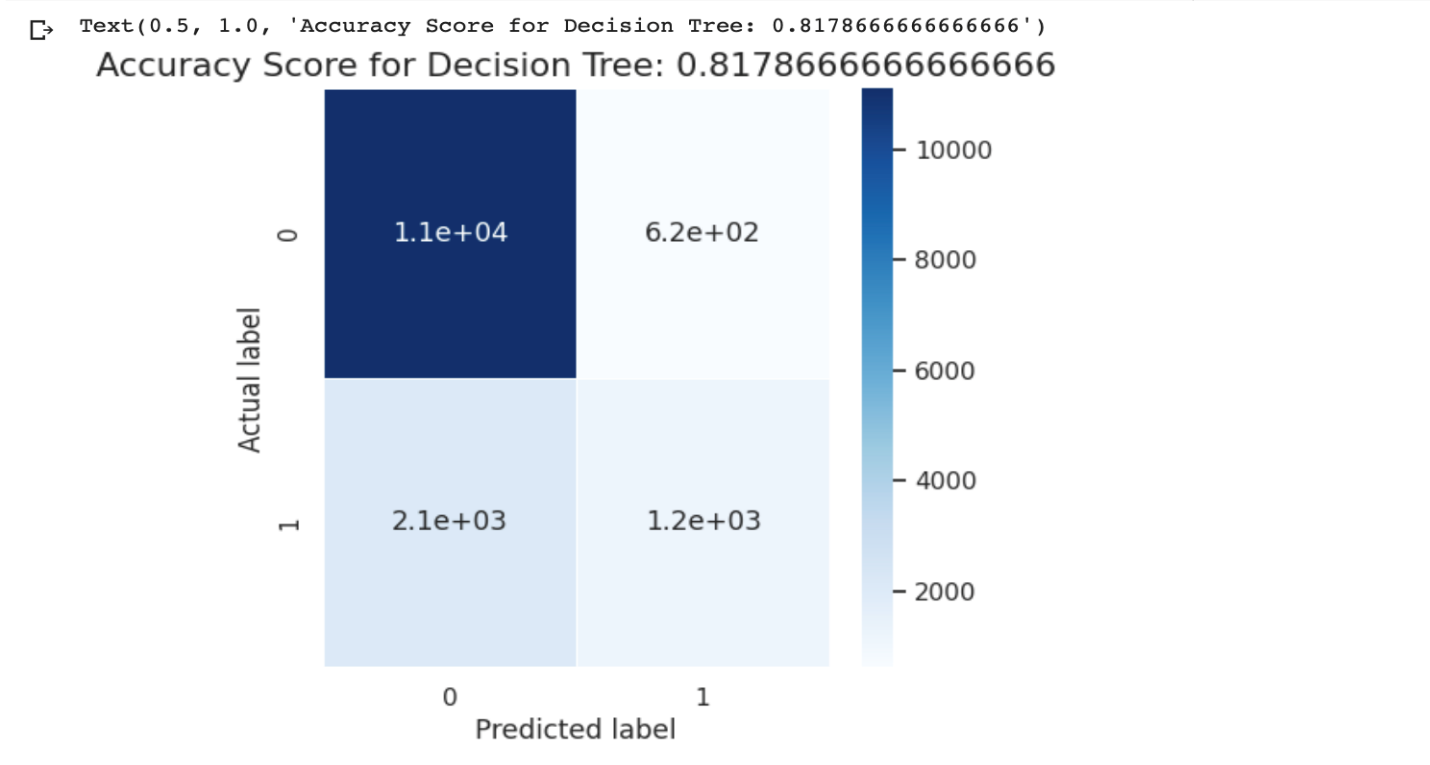
Figure 12: **ROC Curve for Decision Tree**



**Fit statistics for our decision tree**

**The accuracy score** is defined as the overall correctness of the model predictions, a pivotal touchstone of its efficacy. The decision tree model has an accuracy score of **81.8%.**

Figure 13: **Confusion Metric for Decision Tree**



**Precision Score is** a magnifying glass that probes into the model’s ability to deliver accurate positive predictions. The precision score for the decision tree is **81.8%**

**The Jaccard score** is a statistical measure used to quantify the similarity and dissimilarity between two sets. It is particularly useful for comparing the overlap between sets or groups when dealing with binary or categorical data. The Jaccard score provides a value between 0 and 1, where a higher value indicates greater similarity and a lower value indicates greater dissimilarity. Our Jaccard score is **0.69**.

**The recall score**, also known as sensitivity or true positive rate, is a fundamental metric used in classification and machine learning to evaluate the performance of a model, particularly in tasks involving binary or multiclass classification. It measures the proportion of actual positive instances that were correctly identified by the model. A high recall score indicates that the model is effective at capturing a significant portion of the positive instances. On the other hand, a lower recall score suggests that the model might be missing a considerable number of relevant instances. Our Recall score is 81.8%.

**FI Score** is a machine learning evaluation metric that measures the accuracy of a model. It is the weighted average of precision and recall values and reaches its best at 1 and worst at 0. It is measured in percentages. The F1 score combines the precision and recall scores of the model. The F1 score of the decision tree model is **81.8%.**

**Receiver Operator Characteristics (ROC) or Area Under Curve (AUC)**. The ROC curve is the plot of true positive and false positive rates. True positive means the predicted positive that is truly positive. False positive means the predicted that negative that is truly negative. ROC for the best model is **75%.**

## Random Forest

We used the following hyperparameters to train our model for Random Forest.

* Max\_depth = 10
* Max\_features = sqrt
* N\_estimators = 500
* Random\_state = 0

**Model metrics for Random Forest.**

**The accuracy score** can be defined as the overall correctness of the model predictions, a pivotal touchstone of its efficacy. Our model has an accuracy score of **82.44%**

**Precision Score:** is a magnifying glass that probes into the model’s ability to deliver accurate positive predictions. Our precision score is **82.44%**

**The Jaccard score** is a statistical measure used to quantify the similarity and dissimilarity between two sets. It is particularly useful for comparing the overlap between sets or groups when dealing with binary or categorical data. The Jaccard score provides a value between 0 and 1, where a higher value indicates greater similarity and a lower value indicates greater dissimilarity. Our Jaccard score is **0.70**

**The recall score**, also known as sensitivity or true positive rate, is a fundamental metric used in classification and machine learning to evaluate the performance of a model, particularly in tasks involving binary or multiclass classification. It measures the proportion of actual positive instances that were correctly identified by the model. A high recall score indicates that the model is effective at capturing a significant portion of the positive instances. On the other hand, a lower recall score suggests that the model might be missing a considerable number of relevant instances. Our Recall score is **82.44%**.

**FI Score** is a machine learning evaluation metric that measures the accuracy of a model. It is the weighted average of precision and recall values and reaches its best at 1 and worst at 0. It is measured in percentages. The F1 score combines the precision and recall scores of the model. The F1 score of our best model, Random Forest, is **82.44%.**

**Receiver Operator Characteristics (ROC) or Area Under Curve (AUC)**. The ROC curve is the plot of true positive and false positive rates. True positive means the predicted positive that is truly positive. False positive means the predicted that negative that is truly negative. Our ROC for the best model is **78%**.

Figure 14: **Fit Statistics for Random Forest**

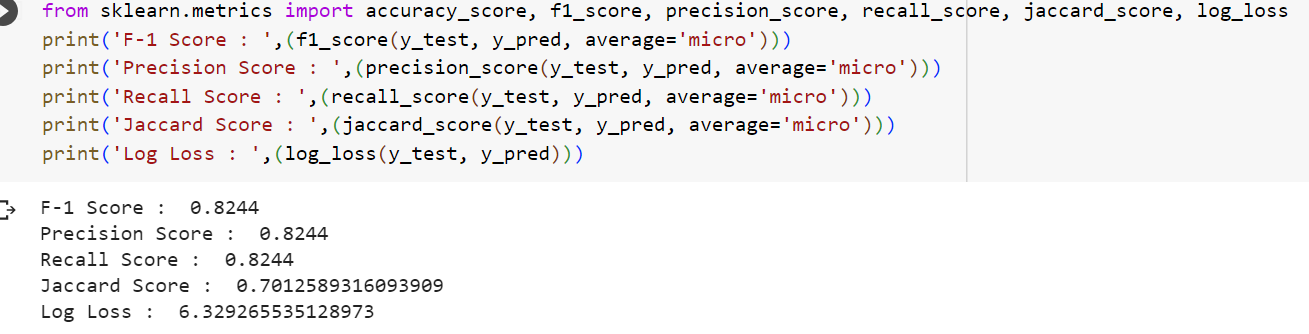


Figure 15: **ROC Curve for Random Forest**

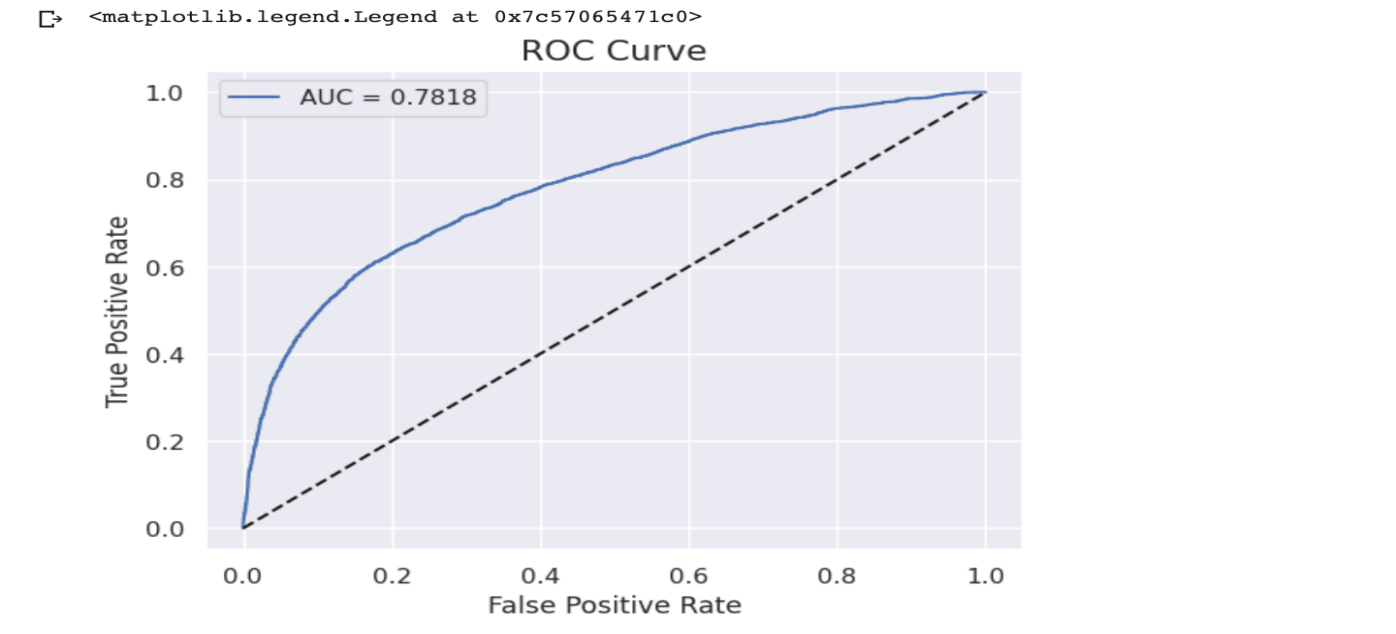


Figure 16: **Confusion Metrics for Random Forest**

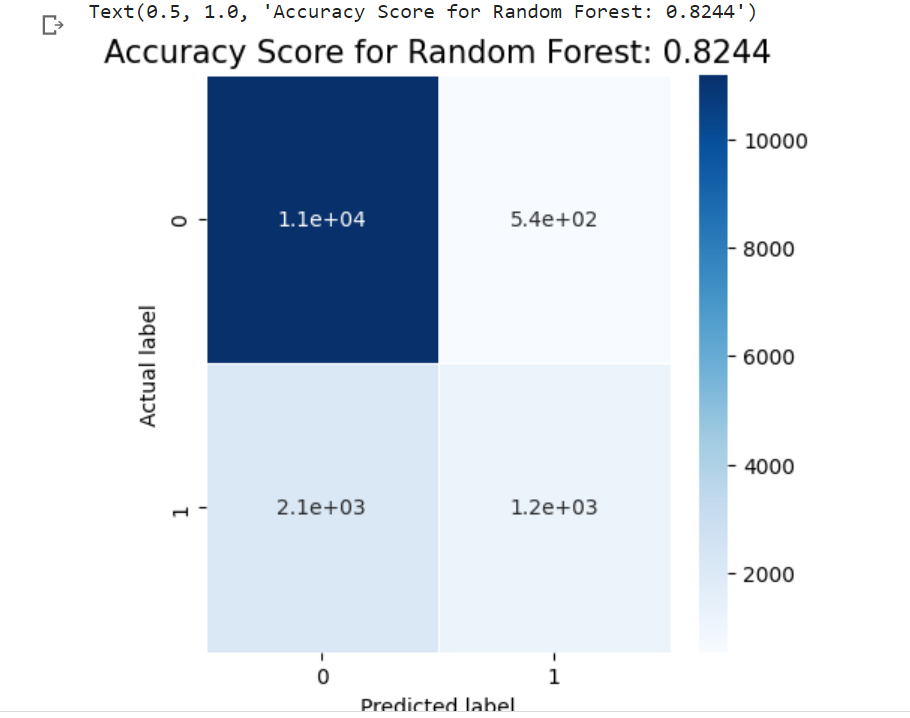
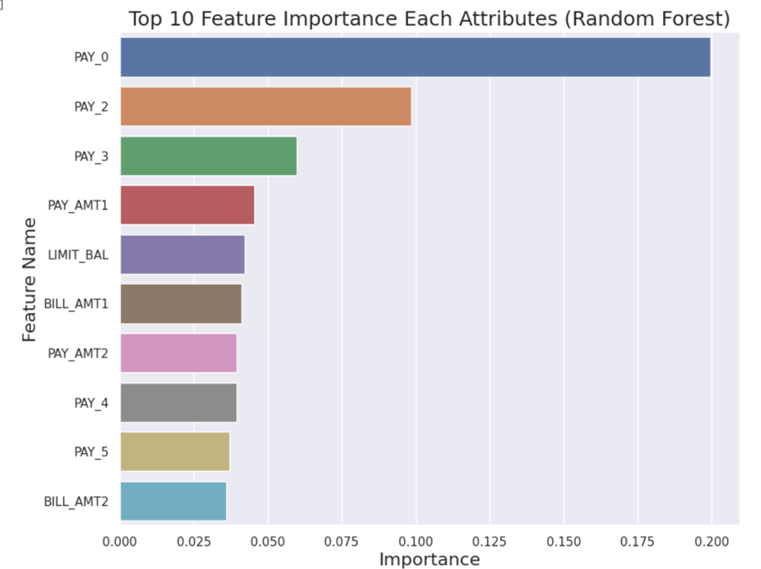


Figure 17:**Summary of Random Forest Performance Metrics**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Model** | **Accuracy Score** | **Precision** | **F1 Score** | **Recall** | **ROC/AUC** |
| **Random Forest** | **82.44%** | **82.44%** | **82.44%** | **82.44%** | **78.18%** |

## Model Sensitivity to Key Drivers



Model sensitivity to key drivers refers to how much a model's predictions or outcomes change in response to variations or changes in the values of its key input variables or features. In other words, it measures how sensitive the model's output is to changes in the most influential factors that drive its predictions.

In our model, we identified the key drivers of the model from the features of importance. The outcome of the model is sensitive to these features and any change in the features will change the model’s outcome such as the performance metrics.

# Logistic Regression

To build our logistic regression, we imported the LogisticRegression from Sklearn.linear\_model and imported GridSearchCV from SKlearn.model\_selection. The parameters used are penalty: l1, solver: Liblinear, CV:5, and Random\_state=0. Below are the performance statistics for our logistic regression

**Model Metrics for Logistic Regression**

**The accuracy score**: **81.51%**

**Precision Score:** 81.51%

**The Jaccard score**: 0.68

**The recall score: 81.51%.**

**FI Score**: 81.51%.

**Receiver Operator Characteristics (ROC) or Area Under Curve (AUC):** 78%.

Figure 18: **Logistic Regression Performance Statistics**

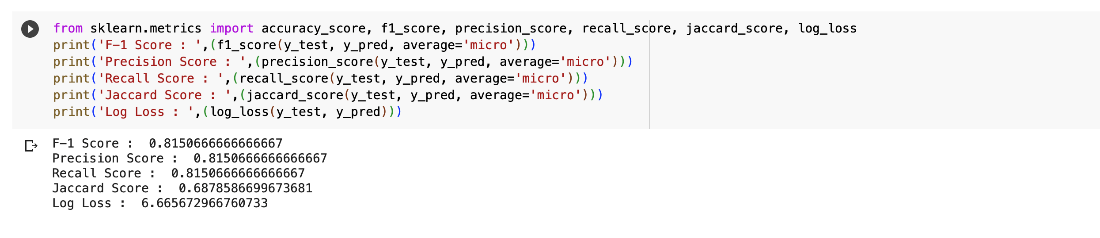
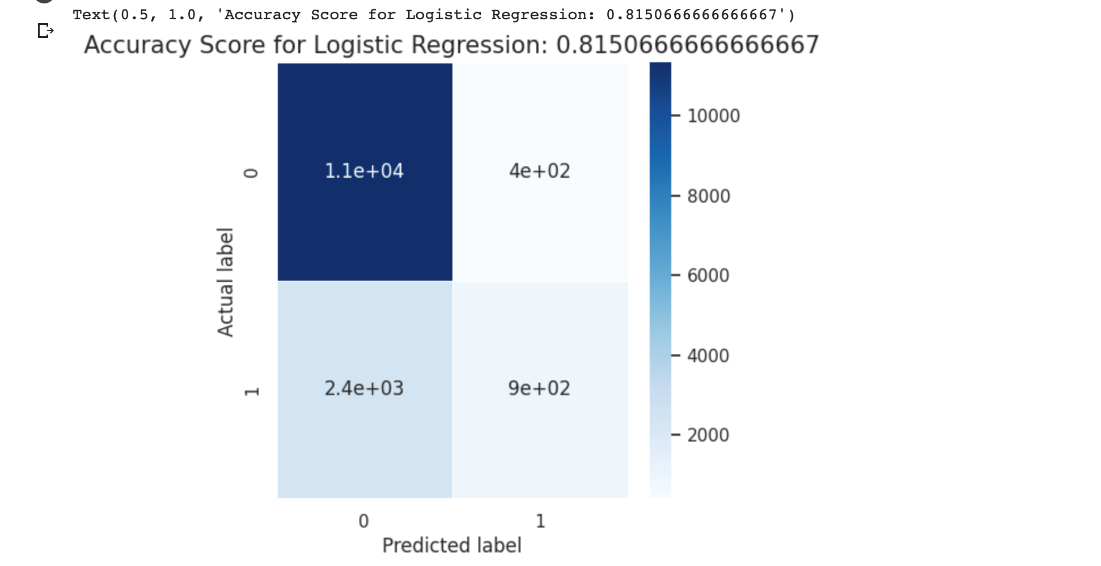


Figure 19: **Logistic Regression Confusion Metrics**



# Neural Networks.

For Neural Networks, our hyperparameters are MLPClassifier as follows.

* Activation = logistic
* Hidden\_layer\_size = 3
* Random\_state = 0
* Solver = lbfgs

**Model Fit Statistics for Neural Networks**

**The accuracy score**: **77.88%**

**Precision Score:** **77.88%**

**The Jaccard score: 0.64**

**The recall score: 77.88%.**

**FI Score**: **77.88%.**

**Receiver Operator Characteristic (ROC) or Area Under Curve (AUC):** 56.67%.

Figure 20:**Performance Statistics for Neural Networks**

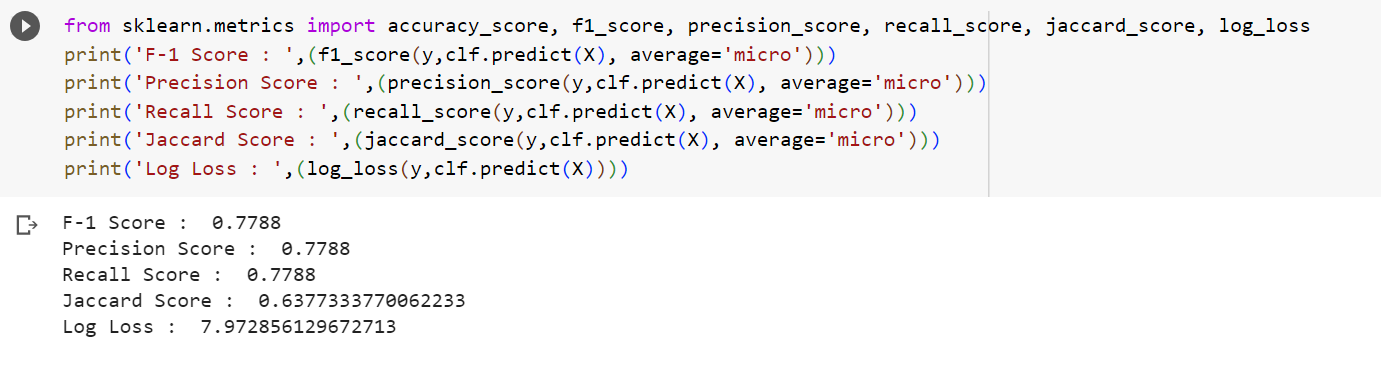


Figure 21: **Confusion Metrics for Neural Networks**

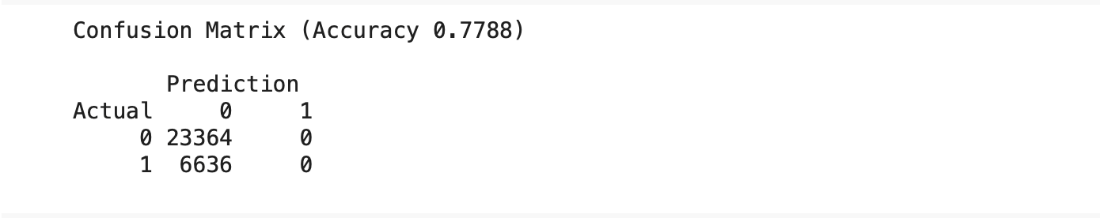
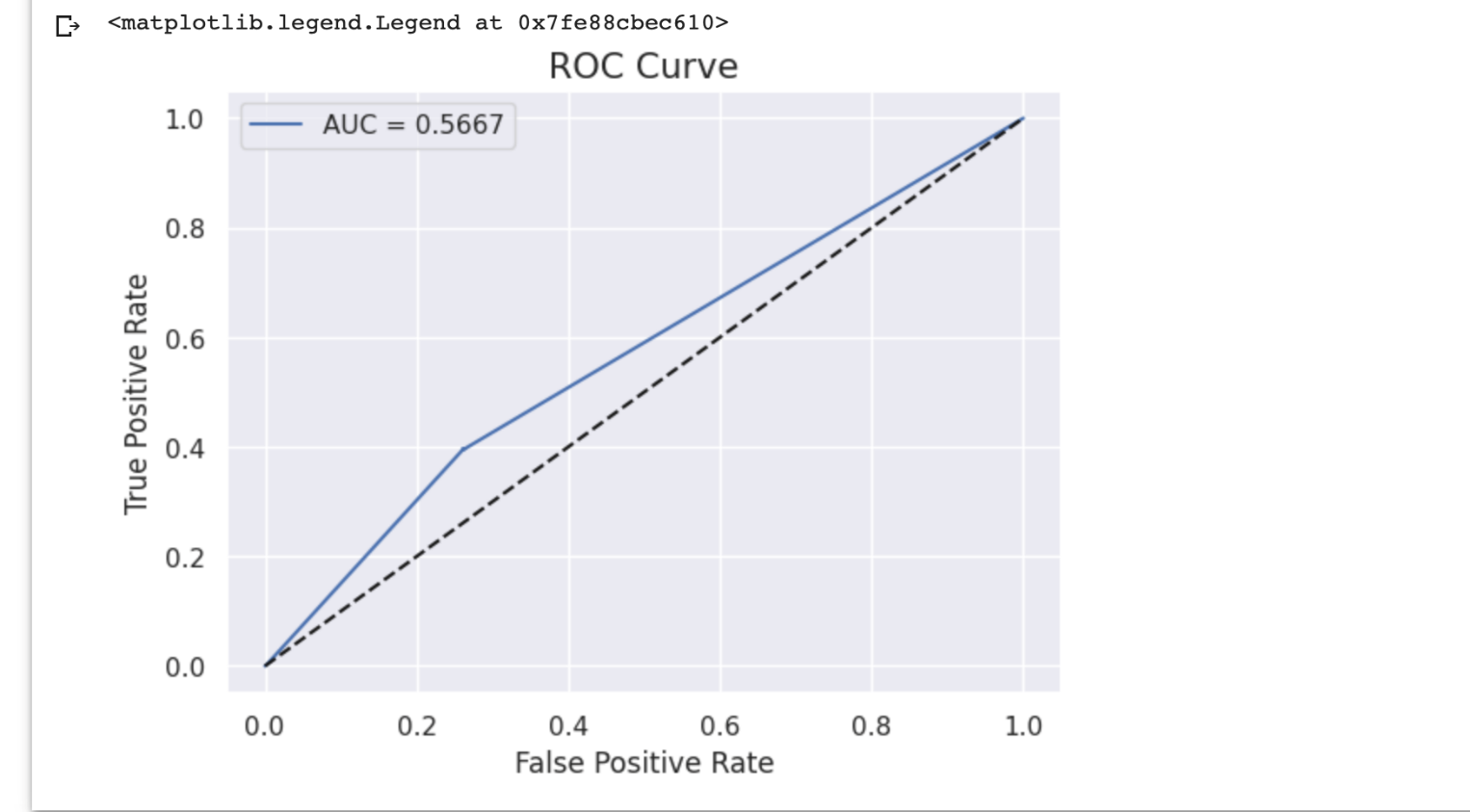


Figure :**ROC/AUC for Neural Networks**



# Model Comparison

Figure 23: **Models’ Primary Metrics Comparison**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Model** | **Accuracy Score** | **Precision** | **F1 Score** | **Recall** | **ROC/AUC** |
| **Random Forest** | **82.37%** | **82.37%** | **82.37%** | **82.37%** | **78.18%** |
| **Decision Tree** | **81.80%** | **81.8%** | **81.8%** | **81.8%** | **75.12%** |
| **Logistic Regression** | **81.51%** | **81.51** | **81.51%** | **81.51** | **72.82%** |
| **Neural Networks** | **77.88%** | **77.88%** | **77.88%** | **77.88%** | **56.67%** |

Although we computed many measurement metrics for all the models, our primary metrics are Accuracy Score and Receiver Operator Characteristics/Area Under Curve (ROC/AUD). We chose the accuracy because it measures the overall of the models while the ROC/AUC is best for binary classification since our project is to classify the customers into defaulters and non-defaulters. Below are the metrics used to compare the performance of each of the models. From the table above, our best model is the Random Forest which has the highest accuracy of 82.37% and ROC of 78.18%.

**Initial Model Fit Statistics**

These initial model fit statistics are used to assess how well the model captures patterns and relationships in the training data. However, they are not always sufficient to determine the model's performance on new, unseen data. It's important to also perform validation using a separate validation or test dataset to ensure that the model generalizes well to new data. Additionally, refining the model through iterative processes of feature engineering, hyperparameter tuning, and evaluation is crucial to achieving optimal performance. Initial model fit statistics, therefore, are the summary metrics and information obtained after training a statistical or machine learning model using a given dataset. Fit statistics is the starting point for evaluating the model performance. They include accuracy score, precision score, F1 score, Recall score, Jaccard score, Log loss, Receiver Operator Characteristics (ROC), or Area Under Curve (AUC). Our best model based on the Fit Statistics is Random Forest. Random Forest is a versatile machine-learning algorithm that leverages an ensemble of multiple decision trees to generate predictions. An important feature of the Random Forest algorithm is its ability to handle continuous and categorical variables. It can handle complex datasets and mitigate against overfitting thereby making it a very useful tool for machine learning prediction.

Risk Tiering (e.g., no action, report, refit, rebuild)

Risk tiering is a process used in risk management to categorize and group entities, assets, or activities based on their level of risk exposure. It involves assigning different risk levels or tiers to different elements to prioritize and allocate resources, interventions, and mitigation strategies more effectively. In our case, risk tiering is the ability of financial institutions or credit lenders to be able to assess and properly categorize credit card customers into buckets of risk. It is imperative to state that our model is built on historical data of customers available to us. There is a possibility of model risk associated with the prediction because the other information about customers is not available in the dataset. For instance, the dataset does not contain information about customers’ deposit balance history, no information about past credit ratings, and, also, no information about customers’ dependents. Due to the unavailability of such information, we categorize the model risk as low. However, users of the model need to develop a risk model using risk scorecards to group customers based on the model prediction of default and not default.

Besides these, the possibility of variable drift could also pose a risk to the prediction of this model. If there are changes to the factors used in this model, they affect the accuracy of the predicted model when deployed. Below is the categorization of the risks associated with our model.

**Low-Risk Tier:** Our report covers the initial stages of model development, including data preprocessing, variables monitoring, and data quality checks. These steps are fundamental and aim to ensure the reliability and accuracy of the data, which is essential for building a robust predictive model. The absence of missing values and the utilization of techniques such as Cap and Floor for handling outliers indicates a thorough approach to data preparation.

**The Moderate Risk Tier:** We discuss the implementation of Variable Drift Monitoring Tolerance, which indicates an awareness of potential changes in the statistical properties of variables over time. This demonstrates a proactive approach to ensuring model accuracy and stability post-deployment. The recognition of variables’ sensitivity to changing customer statuses and the need for ongoing monitoring aligns with responsible model maintenance.

**Moderate to Higher Risk Tier:** Our Initial Model Fit Statistics and the evaluation of model performance metrics introduces more complexity and technical detail. Our report acknowledges key metrics such as accuracy score, precision score, Jaccard score, Recall score, and F1 score. While these metrics provide valuable insights into model performance, their correct interpretation and effective utilization require a deeper understanding of machine learning concepts and statistics.

**High-Risk Tier:** The utilization of Random Forest algorithms for the final model, along with the interpretation of metrics like ROC and AUC, enters the realm of advanced machine learning. Random Forest is a powerful technique, but its optimal implementation demands a strong grasp of ensemble methods, feature importance, and hyperparameter tuning. Similarly, ROC and AUC involve intricate analysis of model discrimination and classification performance.

# **Model Assumptions and Limitations**

* The model assumed that the variables used for training the model are independent and not correlated. There is a class correlation between some classes of variables where some variables in the same class tend to have a correlation coefficient above our range. We did not remove such variables because they were correlated with themselves and not correlated with other variables in the dataset. The variables affected are Pay\_2 to Pay\_5 and Bill\_Amt\_2 to Bill\_Amt5. Removing these variables will affect the performance of the model.
* Our training data is representative of the entire population or the problem domain.

**The limitations of the model are**

* Black Box Model. Our best model, Random Forest is a black box model because its internal is difficult to interpret especially when dealing with a large dataset.
* Hyperparameter Tuning. Our best model, Random Forest is less sensitive to hyperparameter tuning, and finding the right combination of hyperparameters for optimal performance of the model can be challenging. It also takes a lot of time to run using a set of hyperparameters that can improve its performance.

# **Recommendations**

1. From the outcome of the model, the categorical features are risk factors that the credit analyst must pay attention to female gender, university students, ages between 24 to 29, limit balance of $50,000, and single marital status because these were the classes in the model that have the highest rate of default.
2. The data analyst should reduce the skewness of the four skewed variables to further improve the model performance.
3. We also recommend that more advanced analytical techniques be used to get a better model performance.
4. By implementing proactive measures, such as analytics in credit risk assessment, borrower education, and early intervention programs, financial institutions can reduce the incidence of credit card defaults. This leads to lower default rates, indicating improved credit quality and reduced financial distress for cardholders.
5. Addressing credit card defaults involves providing support and assistance to cardholders facing financial challenges. By offering financial education, debt counseling, and tailored repayment options, financial institutions can help individuals regain control of their finances and improve their overall financial well-being.
6. Addressing credit card defaults requires financial institutions to review and enhance their lending practices. This includes implementing robust credit risk assessment processes, ensuring appropriate credit limits, and promoting responsible borrowing behavior. These practices contribute to a healthier credit ecosystem and reduce the risk of customers obtaining credit beyond their means.
7. Effective default management strategies, such as responsive customer service, clear communication, and fair treatment of cardholders in financial distress, enhance the overall customer experience. By addressing credit card defaults promptly and providing supportive services, financial institutions can build stronger relationships with their customers.
8. Addressing credit card defaults necessitates robust risk management practices. Financial institutions that implement comprehensive credit risk management frameworks, including effective monitoring systems, risk assessment models, and stress testing, strengthen their ability to identify, measure, and mitigate credit risk. This contributes to the overall quality of their risk management practices.
9. By demonstrating a commitment to addressing credit card defaults and assisting cardholders in financial distress, financial institutions can build trust and enhance their reputation. Trustworthy and reputable institutions are more likely to attract new customers, retain existing ones, and establish long-term relationships with stakeholders.

# **Conclusions**

In conclusion, we believe that the insights drawn from this project will assist banks and other lending institutions in their credit risk management. We encourage credit analysts to make use of the predictive analytics model to manage existing credit card customers to minimize credit risk and also apply the model in granting credit approval to new customers to proactively mitigate the risk of credit card default.

# **References**

<https://www.fico.com/blogs/canada-bankcard-industry-benchmarking-trends-q4-2022-update>

Government of Canada, S. C. (2023b, March 16). *The traditional credit card debt hangover following the holidays, or something more ominous?* <https://www.statcan.gc.ca/o1/en/plus/3222-traditional-credit-card-debt-hangover-following-holidays-or-something-more-ominous>

Q1 2023 credit industry insights | transunion canada. (n.d.-e). <https://www.transunion.ca/lp/iir>

<https://www.statcan.gc.ca/o1/en/plus/3222-traditional-credit-card-debt-hangover-following-holidays-or-something-more-ominous>

*Press releases: About us: Equifax canada*. Increased Demand for Credit Continues and Delinquency Rates Rise as Some Consumers Struggle with Affordability. (n.d.-a). <https://www.consumer.equifax.ca/about-equifax/press-releases/-/blogs/increased-demand-for-credit-continues-and-delinquency-rates-rise-as-some-consumers-struggle-with-affordability//>

Topre, G. (2022, March 2). *Credit card defaulter prediction*. Kaggle. <https://www.kaggle.com/datasets/gauravtopre/credit-card-defaulter-prediction>