Credit Card Credit Score Prediction: Machine Learning Classification

Problem Statement and Project Goal:

The project goal of this credit card credit scoring machine learning classification exercise is to discover a combination of data attributes that can be calculated by a credit card issuing company to derive a Credit Card score that identifies good credit scores and bad credit scores that will minimize credit card losses and defaults by 20% and predict credit card usage level within a range of 15% to improve credit card profitability within 3 months of implementation.

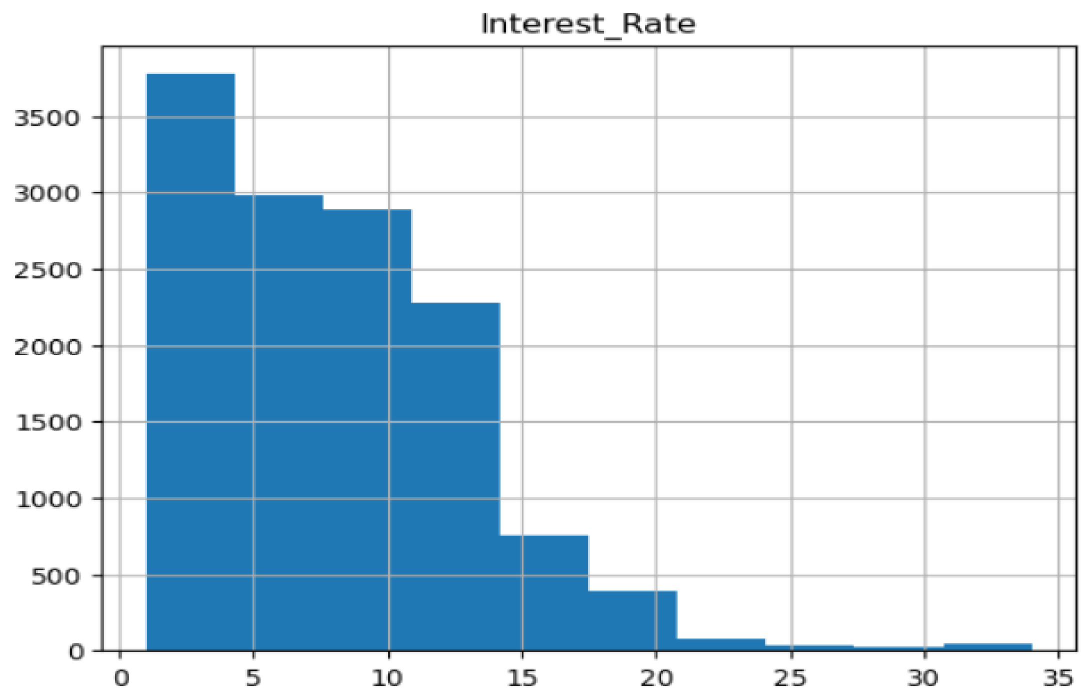
Data Wrangling and Preprocessing:

Since data is very rarely ready the first step is to perform data wrangling and preprocessing such as setting the missing values to their mean or median values for the respective columns as well as setting outliers to the mean. There were also some re-scaling of numerical data and the target Y credit scores had to be converted from text to numeric with values 1 = Poor, 2 = Standard, 3 = Good.

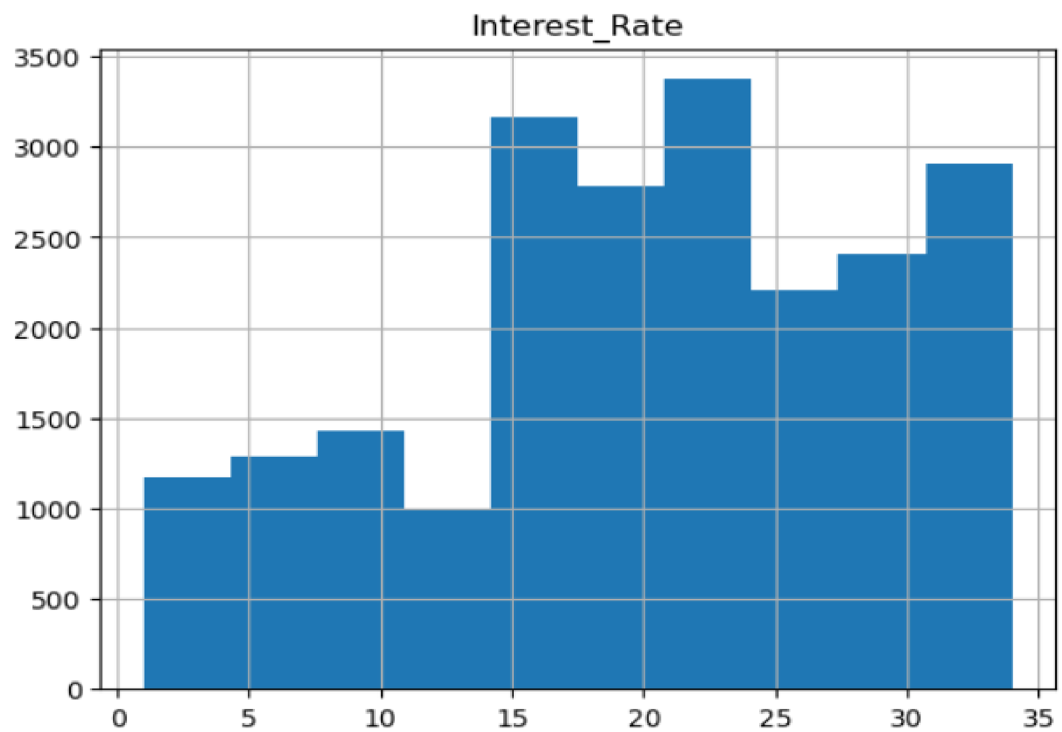
Exploratory Data Analysis:

The obvious first step was to disaggregate explanatory numerical column values along Standard, Good and Poor credit score values by looking at their means, medians, and standard deviations and look to see which columns had noticeable differences and which columns were normally distributed versus those that were not. Interestingly enough, some columns such as Credit_Utilization_Ratio did not show noticeable differences when looking at the three levels of credit scores.

The interest rate charged for the array of customers with **'Good'** credit:



The interest rate charged for the array of customers with **'Poor'** credit:



As the old saying goes, "A picture is worth a thousand words" and these two histograms showed a striking difference between good and poor credit scores interest rate charged which inspired the detectives to drill down into the data from new angles.

The Credit_History_Age variable scores was inline with expectations when split out by Good, Poor, and Standard credit scores where accounts with more months in their history tended to reflect Good credit scores and a lower age of credit history was linked to a Poor credit score more often:

	Credit_History_Age				
	count	mean	std	min	max
Credit_Score					
Good	6258	283.311122	75.934139	6.0	404.0
Poor	12066	164.990220	83.915847	2.0	404.0
Standard	20566	220.135904	100.119190	2.0	404.0

Loan Type was originally storing multiple values per cell and it was broken out into new 9 features: Loan_Type_Total_Count, Auto_Loan_Count, Credit_Builder_Count, Debt_Consolidation_Count, Home_Equity_Loan_Count, Mortgage_Count, Not_Specified_Count, Payday_Count, Personal_Loan_Count that stores the count of these other columns.

Another early step was to test all the columns on whether or not they were normally distributed and to also look at their respective histograms. Most of the data columns were not normally distributed:

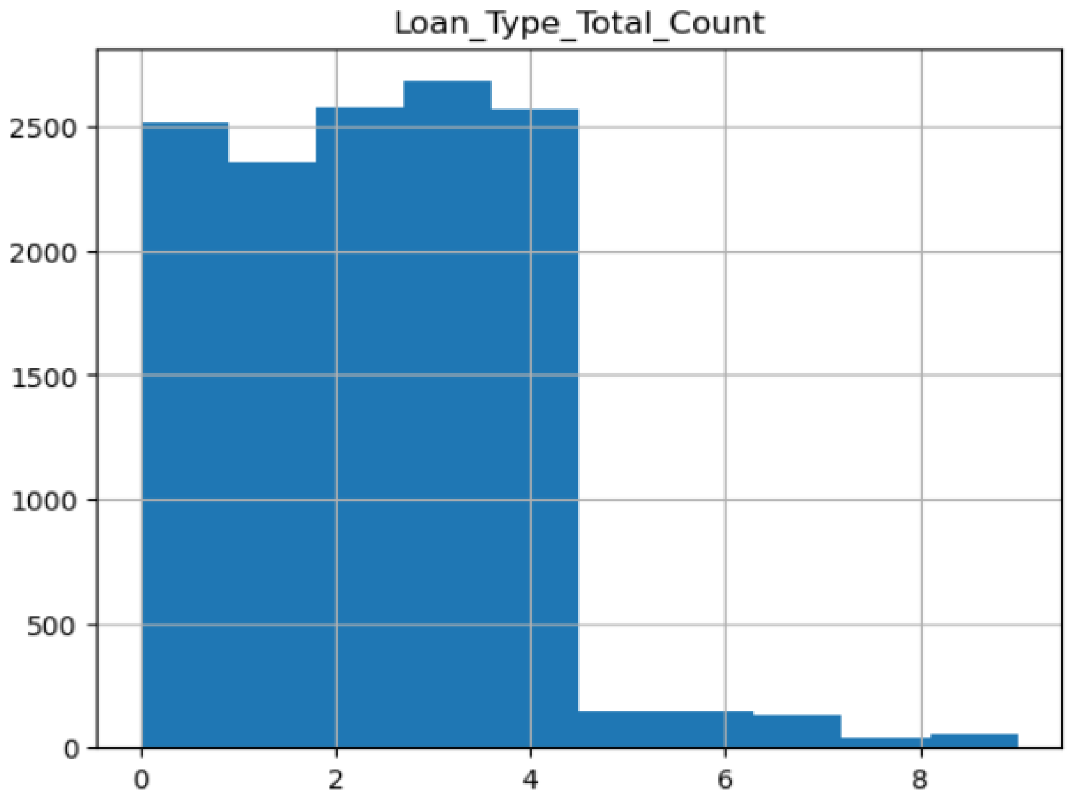
```
List of Data Columns Being Normally Distributed

Unnamed: 0 is not normally distributed
Age is not normally distributed
Annual_Income is not normally distributed
Monthly_Inhand_Salary is not normally distributed
Num_Bank_Accounts is not normally distributed
Num_Credit_Card is not normally distributed
Interest_Rate is not normally distributed
Num_of_Loan is not normally distributed
Num_of_Delayed_Payment is not normally distributed
Delay_from_due_date is not normally distributed
Changed_Credit_Limit is not normally distributed
Num_Credit_Inquiries is not normally distributed
Credit_History_Age is not normally distributed
Loan_Type_Total_Count is not normally distributed
Auto_Loan_Count is not normally distributed
Credit_Builder_Count is not normally distributed
Debt_Consolidation_Count is not normally distributed
Home_Equity_Loan_Count is not normally distributed
Mortgage_Count is not normally distributed
Not_Specified_Count is not normally distributed
Payday_Count is not normally distributed
Personal_Loan_Count is not normally distributed
Outstanding_Debt is not normally distributed
Credit_Utilization_Ratio is not normally distributed
Total_EMI_per_month is not normally distributed
Amount_invested_monthly is not normally distributed
Is_Low_Spent is not normally distributed
Is_High_Spent is not normally distributed
Is_Small_PMT is not normally distributed
Is_Medium_PMT is not normally distributed
Is_Large_PMT is not normally distributed
Monthly_Balance is not normally distributed
Credit_Score_Number is not normally distributed
Mon_Bal_By_Mon_Salary is not normally distributed
```

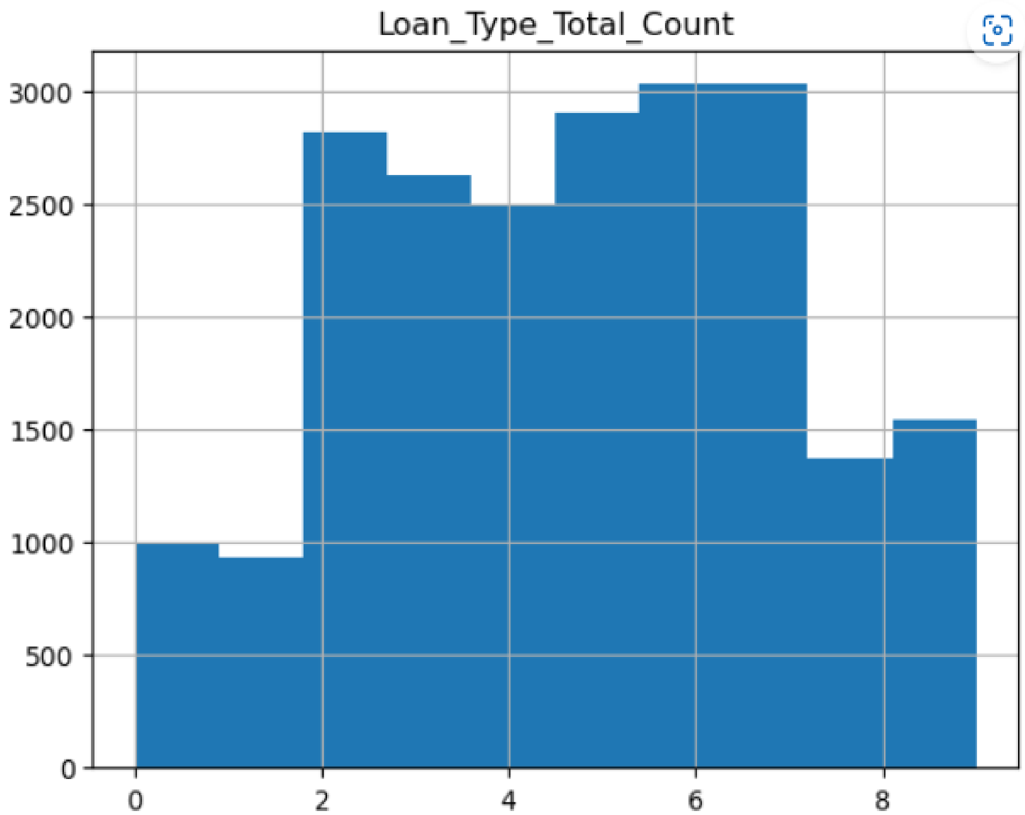
As expected, there was a noticeable difference between good and poor credit scores when it came to assessing the number of types of loans:

Credit Card Credit Score Prediction: Machine Learning Classification

Good credit score:



Poor credit score:



The table below shows that the Poor credit scores were almost 50% higher than Good credit scores for Debt Consolidation loans.

	Debt_Consolidation_Count				
	count	mean	std	min	max
Credit_Score					
Good	6258	0.290348	0.518699	0.0	3.0
Poor	12066	0.544091	0.738220	0.0	4.0
Standard	20566	0.409705	0.644558	0.0	5.0

The table below shows that there is a difference between Good, Standard and Poor for the number of Credit Card inquiries. The industry standard for the total count of what is acceptable is 6 and the Good and Standard stayed lower than that threshold and the Poor grade did not.

Number of Credit Card Inquiries

Credit_Score	count	mean	std	min	max
Good	6258	3.669863	3.545385	0.0	22.01
Poor	12066	8.602178	3.951250	0.0	22.01
Standard	20566	5.843453	4.220495	0.0	45.00

The table below shows the Number of Loans among Standard, Good and Poor Credit Scores. Not surprisingly, a lower number of loans outstanding correspond to a higher credit score:

Number of Loans

Credit_Score	count	mean	std	min	max
Good	6258	2.687317	1.415550	1.0	9.0
Poor	12066	4.801035	2.322812	1.0	9.0
Standard	20566	3.684034	2.166557	1.0	9.0

The table below shows the count of Payday loans broken out by Good, Standard, and Poor credit scores and the Good credit scores have lower Payday loan counts. These are short term high interest rate loans that have a shorter duration and are generally not considered to be good loans due to their high interest rate.

Payday Loan Count

Credit_Score	count	mean	std	min	max
Good	10716	0.313550	0.550101	0.0	4.0
Poor	20742	0.576994	0.767969	0.0	5.0
Standard	35166	0.431212	0.664706	0.0	5.0

Seeking a broader perspective, the detective constructed a correlation matrix to capture the underlying relationships or associations between all the numeric explanatory variables. At this phase of the investigation the correlation matrix showed medium positive correlations for a few variables which were expected: A higher credit history age score corresponded with a lower interest rate; a higher outstanding debt ratio corresponded to a higher number of loans outstanding; a higher outstanding debt amount had a positive association with a higher interest rate being charged. between balances and credit limits for poor credit scores, and weaker correlations for standard and good scores.

	Unnamed: 0	Age	Annual_Income	Monthly_Inhand_Salary	Num_Bank_Accounts	Num_Credit_Card	Interest_Rate	Num_of_Loan
Unnamed: 0	1.000000	0.000405	0.006401	-0.003847	0.014488	-0.006666	0.010366	-0.008729
Age	0.000405	1.000000	0.006537	0.095729	-0.192077	-0.003723	-0.215249	-0.191602
Annual_Income	0.006401	0.006537	1.000000	0.027789	-0.005451	0.007503	-0.009489	-0.001336
Monthly_Inhand_Salary	-0.003847	0.095729	0.027789	1.000000	-0.288211	-0.005635	-0.303768	-0.235116
Num_Bank_Accounts	0.014488	-0.192077	-0.005451	-0.288211	1.000000	0.007750	0.576491	0.449068
Num_Credit_Card	-0.006666	-0.003723	0.007503	-0.005635	0.007750	1.000000	0.004394	0.010778
Interest_Rate	0.010366	-0.215249	-0.009489	-0.303768	0.576491	0.004394	1.000000	0.510806
Num_of_Loan	-0.008729	-0.191602	-0.001336	-0.235116	0.449068	0.010778	0.510806	1.000000
Num_of_Delayed_Payment	0.008072	0.001314	-0.001934	0.003084	0.010509	0.003272	0.014502	0.008523
Delay_from_due_date	0.005247	-0.171818	-0.005524	-0.249938	0.558084	0.009130	0.580832	0.469891
Changed_Credit_Limit	0.018339	-0.156744	-0.004437	-0.181023	0.322721	-0.001463	0.354320	0.347899
Num_Credit_Inquiries	0.002124	-0.217436	-0.007357	-0.253903	0.458855	0.001706	0.552440	0.456566
Credit_History_Age	-0.007152	0.238613	-0.001524	0.284942	-0.489981	-0.009330	-0.574338	-0.568035
Loan_Type_Total_Count	-0.006945	-0.202439	-0.002407	-0.248202	0.471124	0.010388	0.535389	0.950572
Auto_Loan_Count	0.009191	-0.070118	0.003709	-0.085548	0.170079	-0.001125	0.196295	0.345173
Credit_Builder_Count	-0.010746	-0.073139	0.006257	-0.103675	0.188852	-0.007321	0.206149	0.357549
Debt_Consolidation_Count	0.002048	-0.076587	-0.012468	-0.094067	0.167054	0.007093	0.189647	0.333629
Home_Equity_Loan_Count	0.012180	-0.067123	-0.002148	-0.087147	0.171889	0.000966	0.205498	0.352543
Mortgage_Count	0.002796	-0.064835	-0.001070	-0.084210	0.178248	0.010692	0.187712	0.346904
Not_Specified_Count	0.008077	-0.085214	0.004402	-0.080772	0.157751	0.006849	0.185284	0.345078
Payday_Count	-0.022653	-0.080561	-0.012272	-0.094941	0.178232	0.007605	0.206840	0.365832
Personal_Loan_Count	-0.005393	-0.081884	0.012090	-0.100797	0.186451	0.005421	0.199091	0.351732
Outstanding_Debt	0.007193	-0.202455	0.000942	-0.277525	0.515518	0.006741	0.622886	0.610713
Credit_Utilization_Ratio	0.000669	0.023071	0.006058	0.149691	-0.059697	-0.001655	-0.063865	-0.073228
Total_EMI_per_month	0.009388	0.004962	-0.008809	0.008100	0.003176	0.001032	0.000090	0.000085

One of the counterintuitive variables was Monthly_Balance, which showed lower values for Poor credit scores versus Good credit scores; perhaps this was not a monthly credit card balance or was mis-labeled but it seemed out of the ordinary.

Monthly Balance

Credit_Score	count	mean	std	min	max
Good	6258	445.180757	239.693291	0.01	1443.569276
Poor	12066	330.090150	160.206879	0.01	1468.313963
Standard	20566	385.821443	199.595777	0.01	1552.946094

Another surprising metric was Credit_Utilization_Ratio, which showed virtually no difference for Poor credit scores versus Good credit scores;

Credit_Utilization_Ratio

Credit_Score	count	mean	std	min	max
Good	6258	32.582331	5.130783	20.880082	48.176599
Poor	12066	31.980051	5.043068	20.985606	46.230683
Standard	20566	32.212623	5.052935	20.830946	49.564519

Perhaps interesting is the observation in the data that there was not a huge dollar difference amount when viewing the Annual Income variable across Good, Poor and Standard credit scores:

Annual Income

Credit_Score	count	mean	std	min	max
Good	6258	178298.359822	1.409499e+06	7261.91	24198062.0
Poor	12066	160571.225888	1.417956e+06	7005.93	23912939.0
Standard	20566	184948.757235	1.467858e+06	7006.52	24160009.0

Similarly, the Total Equated Monthly Installment amount was not different among the Poor, Standard, and Good Credit scores; this can perhaps be accounted for by the different interest being charged to the three groups. EMI is a fixed payment made by a borrower to a lender at a specified date at each calendar month.

Total Equated Monthly Installment (EMI) Per Month

	count	mean	std	min	max
Credit_Score					
Good	13233	1476.562484	8450.021653	0.0	82236.0
Poor	21730	1376.620418	8199.564539	0.0	81971.0
Standard	40115	1373.291245	8220.851644	0.0	82193.0

The conclusion of the exploratory data analysis phase has provided valuable insights into the underlying patterns and relationships within the credit card classification dataset. While our analysis revealed very noticeable differences across several columns for the Poor, Standard, and Good credit score categories such as in Debt Consolidation and Payday loans, Credit History Age, Number of Credit Inquiries, and the Interest Rate among other variables. It is also noteworthy that Credit Utilization, Annual Income, Total EMI Charged, and Monthly_Balance demonstrated no significant differences among these groups. These observations highlighted the need for further investigation to understand the roles these variables play in credit score determination and potential refinement of the credit scoring model. Moreover, our correlation matrix analysis unveiled medium positive associations among a few columns, indicating that these variables may share a degree of underlying interconnectedness that could be leveraged in the development of predictive models for credit score estimation. As we move forward, it is essential to delve deeper into these relationships, incorporating advanced statistical techniques and machine learning algorithms to enhance our understanding of credit score dynamics to provide an optimally tailored product to our customers at a more profitable level for our credit card company.

Modeling Phase: Six Machine Learning Classifier Models

Modeling Phase: For the modeling phase of the credit card credit score classification project, I evaluated six different classifier models to run against the train and test data credit card files: a Decision Tree Classifier, a Logistic Regression classifier, Gaussian Naive Bayes Classifier, KNeighborsClassifier, XGBoost Classifier, and a Random Forest classifier. The Python code was run in separate Jupyter files for each type of model.

The Problem: The problem being measured here is to predict credit scores (target Y variable with values 1 (Poor), 2 (Standard), 3 (Good)) when compared with a range of numerical X variable data columns. The challenge is to make predictions of Credit Score and correctly identify both good and poor credit scores from the training data using the different classification models and evaluate their Accuracy, Recall, Precision scores and F1 Scores for all six models:

Model Type	Accuracy	Recall	Precision	F1 Score
Random Forest Classifier	0.77	0.77	0.77	0.77
Decision Tree Classifier	0.68	0.68	0.68	0.68
K-Neighbors Classifier	0.67	0.67	0.67	0.65
XG Boost Classifier	0.74	0.74	0.74	0.74
Gaussian Naive Bayes Classifier	0.55	0.55	0.56	0.55
Logistic Regression Classifier	0.56	0.56	0.46	0.5

I chose the F1-score as the best metric to use when evaluating the different machine learning classifier models reviewing credit card credit scores. The F1-score is a measure of a model's accuracy that takes into account both precision and recall. Precision is the fraction of predicted positives that are actually positive, and recall is the fraction of actual positives that are predicted positive. A high F1-score indicates that the model is both accurate and precise.

In the context of credit card classification, a high F1-score indicates that the model is able to correctly identify both good and poor credit scores. This is important because it allows lenders to make informed decisions about who to extend credit to.

Here is a table of the different metrics and what they measure:

Metric	**Definition**
Accuracy	The fraction of predictions that are correct.
Precision	The fraction of predicted positives that are actually positive.
Recall	The fraction of actual positives that are predicted positive.
F1-score	The harmonic mean of precision and recall.

Credit Card Credit Score Prediction: Machine Learning Classification

Finding the optimal combination of parameters: hyperparameter tuning

In machine learning we have two types of parameters: those that are learned from the training data, for example, the weights in logistic regression, and the parameters of a learning algorithm that are tuned separately and are called hyperparameters. For example, the regularization parameter in logistic regression or the depth parameter of a decision tree. The optimization technique called "grid search" is used to find the optimal combination of hyperparameter values.

Hyperparameter Tuning: The other classifier models underwent hyperparameter tuning and the Random Forest hyperparameter will be examined here since the Random Forest classifier scored the best at an accuracy of .77 when using n_estimator at 20; setting it at larger n_estimator caused the Python compiler to run for a very long time. The cross validation function returned these parameters as being optimal:

'criterion': 'gini', 'max_depth': 7, 'min_samples_leaf': 5, 'min_samples_split': 2, 'n_estimators': 20

These are the parameters used in the Grid Search Cross Validation phase when tuning the Hyperparameters:

n_estimators: This is the number of trees in the random forest. A higher number of trees will generally lead to a better model, but it will also take longer to train the model.

max_depth: This is the maximum depth of each tree in the random forest. A higher maximum depth will generally lead to a better model, but it can also lead to overfitting.

min_samples_split: This is the minimum number of samples required to split a node in a tree. A higher minimum number of samples will generally lead to a more robust model, but it can also lead to a less accurate model.

min_samples_leaf: This is the minimum number of samples required to be in a leaf node. A higher minimum number of samples will generally lead to a more robust model, but it can also lead to a less accurate model.

max_features: This is the maximum number of features considered when splitting a node in a tree. A higher maximum number of features will generally lead to a more accurate model, but it can also lead to overfitting.

criterion: the criterion parameter in the decision tree classification model is used to measure the quality of a split. The most common criteria are Gini impurity and entropy.

Gini impurity: This is a measure of how pure a node is. A node is pure if all of the instances in the node belong to the same class. The Gini impurity of a node is calculated by summing the Gini impurities of its children nodes.

Entropy: This is a measure of how uncertain a node is. A node is uncertain if the instances in the node are evenly distributed between the different classes. The entropy of a node is calculated by summing the entropies of its children nodes.

Credit Card Credit Score Prediction: Machine Learning Classification

Predictions and Findings:

In concluding the modeling phase, the Random Forest Classifier greatly improved on the lowest ranking classifier, the Logistic Regression Classifier model: the Random Forest Classifier posted an Accuracy score of .77 versus .56 for the Logistic Regression Classifier, which corresponds to a 37.5% increase in credit scoring ability. After hyperparameter tuning, the performance of the Random Forest Classifier improved further, solidifying its position as the most suitable model for credit score prediction in our study. It may be possible to increase the number of n_estimators, i.e. the number of trees in the Random Forest. In the context of credit card classification, a high F1-score was chosen as the most important metric since it indicates that the model is able to correctly identify both good and poor credit scores. This is important because it allows lenders to make informed decisions about who to extend credit to while maximizing profit and minimizing risk.