Credit Risk Analysis

Purpose:

The purpose of this analysis is to predict credit risk. We are using a credit card dataset from LendingClub, a peer-to-peer lending service to complete our analysis. We are tasked with using several different statistical models to evaluate the data and try to figure out which model performs the best at predicting credit risk.

* Results:

**Balanced accuracy scores:**

Over 0.9 - Very good

Between 0.7 and 0.9 - Good

Between 0.6 and 0.7 - OK

Below 0.6 - Poor

1 RandomOverSampler : 0.6293939430565123

2 SMOTE : 0.6277008271188627

3 RandomUnderSampler : 0.5902034962189427

4 SMOTEENN : 0.6411460410698961

5 Random Forrest : 0.9945945945945946

6 GradientBoostingClassifier: 0.9945945945945946

**Precision/Recall scores:**

1 RandomOverSampler :

[Credit\_Risk\_Analysis/RandomOversample Prec\_recall.png at main · russellctaylor/Credit\_Risk\_Analysis (github.com)](https://github.com/russellctaylor/Credit_Risk_Analysis/blob/main/RandomOversample%20Prec_recall.png)

2 SMOTE :

[Credit\_Risk\_Analysis/SMOTE Prec\_recall.png at main · russellctaylor/Credit\_Risk\_Analysis (github.com)](https://github.com/russellctaylor/Credit_Risk_Analysis/blob/main/SMOTE%20Prec_recall.png)

3 RandomUnderSampler :

[Credit\_Risk\_Analysis/RandomUndersampler Prec\_recall.png at main · russellctaylor/Credit\_Risk\_Analysis (github.com)](https://github.com/russellctaylor/Credit_Risk_Analysis/blob/main/RandomUndersampler%20Prec_recall.png)

4 SMOTEENN :

[Credit\_Risk\_Analysis/SMOTEENN Prec\_recall.png at main · russellctaylor/Credit\_Risk\_Analysis (github.com)](https://github.com/russellctaylor/Credit_Risk_Analysis/blob/main/SMOTEENN%20Prec_recall.png)

5 Random Forrest :

[Credit\_Risk\_Analysis/RandomForrest Prec\_recall.png at main · russellctaylor/Credit\_Risk\_Analysis (github.com)](https://github.com/russellctaylor/Credit_Risk_Analysis/blob/main/RandomForrest%20Prec_recall.png)

6 GradientBoostingClassifier:

[Credit\_Risk\_Analysis/GradientBoostingClassifier Prec\_recall.png at main · russellctaylor/Credit\_Risk\_Analysis (github.com)](https://github.com/russellctaylor/Credit_Risk_Analysis/blob/main/GradientBoostingClassifier%20Prec_recall.png)

* Summary:
  + In summary based on the results of the different models it seems the Random Forrest model preformed the best. It had the greatest accuracy and also had a high precision for predicting credit risk. I would recommend using the Random Forrest model for predictive analytics on this data set.