# Credit Risk Analysis Report

Overview of the analysis:

The purpose of this analysis is to utilize different methods and models to effectively improve our success metric for a supervised learning, binary classification problem in the field of credit risk to determine whether credit card applicant is a high risk for defaulting on their payments. The success metric that should be considered for this analysis should be mainly recall as we want to minimize the chance of a false negative for the credit card company.

For the methodology portion of the analysis, we will utilize different sampling techniques:

1. Naïve random oversampling
2. SMOTE
3. Undersampling – ClustterCentroids
4. Combination of Over and Undersampling also called: SMOTEENN

Then we will evaluate a number of different machine learning models to evaluate performance:

1. BalancedRandomForestClassifier
2. EasyEnsembleAdaBoostClassifier

## Results:

Here are the full results of the analysis:

|  |  |  |  |
| --- | --- | --- | --- |
| **Method / Model** | **Balanced Accuracy** | **Avg Precision** | **Avg Recall** |
| Naïve random oversampling | 0.590268965 | 0.99 | 0.58 |
| SMOTE | 0.6172722 | 0.99 | 0.66 |
| Undersampling – ClustterCentroids | 0.500727204 | 0.99 | 0.47 |
| Combination of Over and Undersampling also called: SMOTEENN | 0.609723179 | 0.99 | 0.58 |
| BalancedRandomForestClassifier | 0.733858491 | 0.99 | 0.96 |
| EasyEnsembleAdaBoostClassifier | 0.855191752 | 0.99 | 0.92 |

Based on the metrics of average recall, we see a clear winner with BalancedRandomForestClassifier with a 4% lift on average recall and therefore we should go with this model

## Recommendations / Summary

We will utilize the model that gave us the best metric of recall which in this case is BalancedRandomForestClassifier