# Deliverable 17 Written Analysis

# Overview of the analysis:

The purpose of this analysis is to apply data preparation, statistical reasoning and machine learning (ML) algorithms to solve a real-world challenge like credit card risk. Credit card risk can be classified as a good loan or a risky loan which is appropriate training, building and evaluating imbalanced-learn and scikit-learn libraries using resampling. The dataset from LendingClub will be oversampled using RandomOverSampler and SMOTE algorithm. As well as combinatorial approach of over- and undersampling using the SMOTEEN algorithm. Finally compare two new ML models to reduce bias leveraging BalancedRandomForestClassifier and EasyEnsembleClassifier algorithms to predict credit risk.

# Results:

### RandomOverSampler – Oversampling

\* Accuracy: 66.0%

\*![ RandomOverSample]( <https://github.com/mfGWU/Credit_Risk_Analysis/blob/main/img/RandomOverSampler.PNG> "RandomOverSample")

### SMOTE – Oversampling

\* Accuracy: 65%

\*![ SMOTE]( <https://github.com/mfGWU/Credit_Risk_Analysis/blob/main/img/SMOTE.PNG> "SMOTE ")

### ClusterCentroids – Under Sampling

\* Accuracy: 65%

\*![ ClusterCentroids]( <https://github.com/mfGWU/Credit_Risk_Analysis/blob/main/img/ClusterCentroids.PNG> " ClusterCentroids ")

### SMOTEENN - Combination (Over and Under)

\* Accuracy: 53%

\*![ SMOTEENN]( <https://github.com/mfGWU/Credit_Risk_Analysis/blob/main/img/SMOTEENN.PNG> " SMOTEENN ")

### RandomForestClassifier - Balanced Random Forest Classifier

\* Accuracy: 99.6%

\*![ RandomForestClassifier]( <https://github.com/mfGWU/Credit_Risk_Analysis/blob/main/img/RandomForestClassifier.PNG> " RandomForestClassifier ")

### AdaBoostClassifier - Easy Ensemble AdaBoost Classifier

\* Accuracy: 99.6%

\* ![ AdaBoostClassifier]( <https://github.com/mfGWU/Credit_Risk_Analysis/blob/main/img/AdaBoostClassifier.PNG> " AdaBoostClassifier ")

# Summary:

­

Looking at the results of the six algorithms the Classifiers were 99% accurate comparing them to the Oversampling, Undersampling and Combination algorithms. The Oversampling and Undersampling algorithms were 66% accurate while the Combination was 53% accurate. I am impress with accuracy of the Classifiers but can not recommend them for now as I need more data and understanding to support my recommendation and especially when the accuracy is 99%. The same for the Oversampling, Undersampling and Combination algorithms, I can not recommend them yet without further studies. With Oversampling and Undersampling average 66% accuracy, I am inclined to recommend them because their accuracy seen realistic.