Module 17: Credit\_Risk\_Analysis

Resampling Models & Classifiers to Predict Potential Credit Risk

OVERVIEW

The purpose of this module is to run an analysis to learn about the   
process in which “Machine Models” asses credit card risk.

Deliverable 1  
RESULTS for Naïve Random OVERSAMPLING:Calendar

Description automatically generated with medium confidence

Calculated the balanced accuracy score: 0.5048920291935648

High-Risk Performance:  
 precision is low (0.33)

recall is very low (0.01)

Low-Risk Performance:  
 precision is very high (0.99)

recall is perfect (1.00)

Weighted Average F1 Score: 0.99

Deliverable 2  
RESULTS for SMOTE:  
Calendar

Description automatically generated

Calculated the balanced accuracy score: 0.6583599806425919

High-Risk Performance:  
 precision is very low (0.01)

recall is good (0.63)

Low-Risk Performance:  
 precision is very high (1.00)

recall is good (0.68)

Weighted Average F1 Score: 0.81

RESULTS for Cluster Centroid:A screenshot of a computer

Description automatically generated with medium confidence

Calculated the balanced accuracy score: 0.6583599806425919

High-Risk Performance:  
 precision is very low (0.01)

recall is above average (0.69)

Low-Risk Performance:  
 precision is perfect (1.00)

recall is low (0.40)

Weighted Average F1 Score: 0.56

RESULTS for SMOTEEN:A picture containing text, monitor, meter, outdoor

Description automatically generated

Calculated the balanced accuracy score:   
 0.5447046721744204

High-Risk Performance:  
 precision is very low (0.01)

recall is good (0.72)

Low-Risk Performance:  
 precision is perfect (1.00)

recall is average (0.57)

Weighted Average F1 Score: 0.72

RESULTS for Balanced Random Forest Classifier:A screenshot of a computer

Description automatically generated with low confidence

Calculated the balanced accuracy score:   
 0.7885466545953005  
High-Risk Performance:  
 precision is very low (0.03)

recall is very good (0.70)

Low-Risk Performance:  
 precision is perfect (1.00)

recall is very good (0.87)

Weighted Average F1 Score: 0.93

RESULTS for Easy Ensemble AdaBoostClassifier:A picture containing text, meter, monitor, black

Description automatically generated

Calculated the balanced accuracy score:   
 0.9316600714093861

High-Risk Performance:  
 precision is very low (0.09)

recall is very good (0.92)

Low-Risk Performance:  
 precision is perfect (1.00)

recall is very good (0.94)

Weighted Average F1 Score: 0.97

Summary

I learned that F1 score is a tool used as a single summary statistic of both precision and sensitivity(recall). This makes the F1 score a powerful measurement to check model performance. Upon comparing the “Weighted Average” F1 scores the best forming model is the Naïve Random Over-sampler with an average F1 score of 0.99. While the worst performing is Cluster Centroid with an average F1 score of 0.56.

Based on performance in this scenario the Naïve Random Over-sampler model is greatly recommended for assessing credit risk.