#Written Report on Credit Risk Analysis

1. Overview of the analysis

The purpose of this analysis is to apply six different unbalanced machine learning models to the same dataset and compare the results obtained. The dataset used contains the credit applications from different individuals in the file named 2019Q1.csv , obtained from a personal credit company called LendingClub.

The csv file contains 115,676 individuals with various loan status, the first starter code provided in the challenge is for doing oversampling, underdamping and combination algorithm. The second starter code provide in the challenge for doing Ensemble learning to reduce bias in the model.

Following types of Algorithms and corresponding names of Algorithms used in the challenge

1. Oversampling Algorithms- Naïve Random Oversampling and SMOTE Oversampling.
2. Under sampling Algorithms- Cluster Centroid
3. Combination of over and under sampling-SMOTEENN
4. Ensemble Learner- Balanced Random Forest Classifier and Easy Ensemble AdaBoost classifier

The starter code provided had codes which cleaned data e.g remove columns with null values, remove rows where the loan status is “issued” and convert remaining loan status to ‘Low Risk” and “High risk category”.

All models are evaluated based on the accuracy score to determine which is the best model to use for predicting credit risk.

1. Results

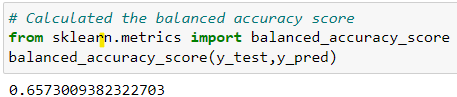
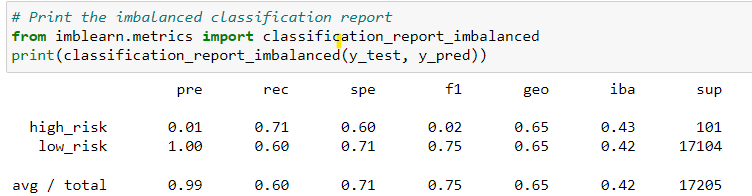
After running various Algorithm, below accuracy, precision and recall scores for high and low risk category generated.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Name of Algorithm** | **Accuracy** | **High Risk** | |  | **Low Risk** | |
| **Precision** | **Recall** |  | **Precision** | **Recall** |
| Naive Random Oversampling | 65.73% | 1.00% | 71.00% |  | 100.00% | 60.00% |
| SMOTE Oversampling | 65.73% | 1.00% | 71.00% |  | 100.00% | 60.00% |
| Cluster Centroids | 65.73% | 1.00% | 71.00% |  | 100.00% | 60.00% |
| SMOTEENN | 65.73% | 1.00% | 71.00% |  | 100.00% | 60.00% |
| Balanced Random Forest Classifier | 67.82% | 95.00% | 36.00% |  | 100.00% | 100.00% |
| Easy Ensemble AdaBoost Classifier | 67.82% | 95.00% | 36.00% |  | 100.00% | 100.00% |

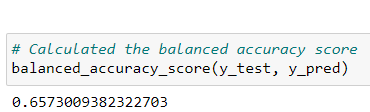
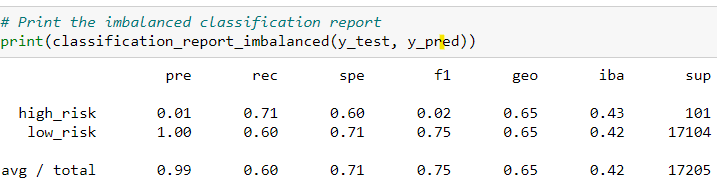
Based on the accuracy , precision and recall scores, it is recommended to use Balanced Random Forest Classifier or Easy Ensemble Ada Boost Classifier model for predicting the credit score.

**Screen shots of the model output**

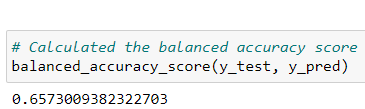
Naive Random Oversampling

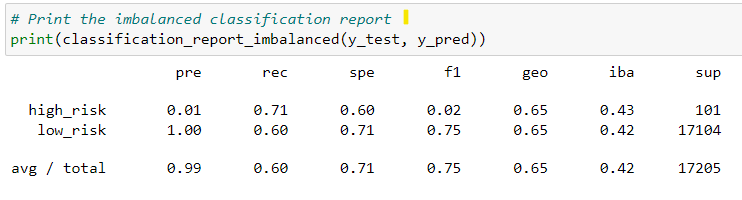
 

SMOTE Oversampling

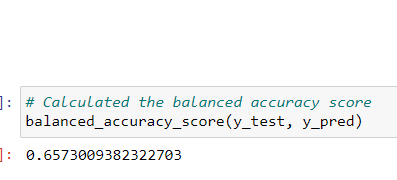
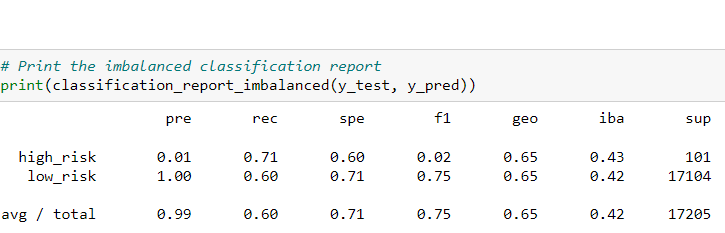
 

Cluster Centroids

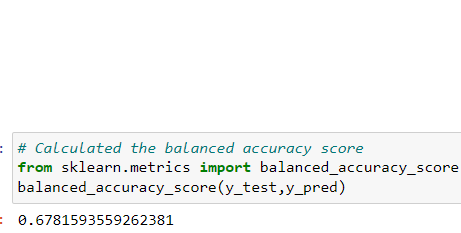
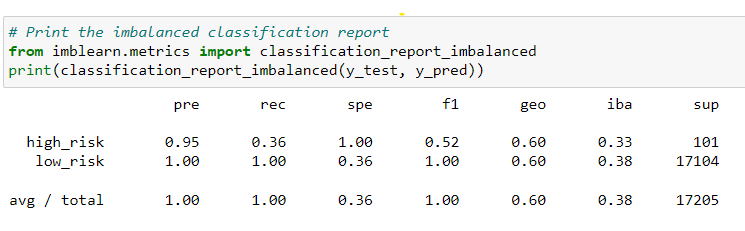




SMOTEENN

Balanced Random Forest Classifier

Easy Ensemble AdaBoost Classifier

