**Overview**

Using the credit card credit dataset from LendingClub, a peer-to-peer lending services company, I’ll oversample the data using the RandomOverSampler and SMOTE algorithms, and undersample the data using the ClusterCentroids algorithm. Then, I’ll use a combinatorial approach of over- and undersampling using the SMOTEENN algorithm. Next, I’ll compare two new machine learning models that reduce bias, BalancedRandomForestClassifier and EasyEnsembleClassifier, to predict credit risk. Once I’m done, I’ll evaluate the performance of these models and make a written recommendation on whether they should be used to predict credit risk.

Below is a list of tasks which will be completed:

* Use Resampling Models to Predict Credit Risk.
* Use the SMOTEENN Algorithm to Predict Credit Risk.
* Use Ensemble Classifiers to Predict Credit Risk.

**Summary of different Machine Learning Algarythims**

Result from RandomOverSampler

* Accuracy = 66.20%
* Precision = 99%

Recall = 60%  
 pre rec spe f1 geo iba sup

high\_risk 0.01 0.67 0.42 0.01 0.53 0.29 101

low\_risk 1.00 0.42 0.67 0.59 0.53 0.27 17104

avg / total 0.99 0.42 0.67 0.58 0.53 0.27 17205

Result from SMOTE

* Accuracy = 64%
* Precision = 99%

Recall = 70%

high\_risk 0.01 0.67 0.42 0.01 0.53 0.29 101

low\_risk 1.00 0.42 0.67 0.59 0.53 0.27 17104

avg / total 0.99 0.42 0.67 0.58 0.53 0.27 17205

Result from RandomUnderSampler

* Accuracy = 70%
* Precision = 99%
* Recall = 58%

pre rec spe f1 geo iba sup

high\_risk 0.01 0.67 0.42 0.01 0.53 0.29 101

low\_risk 1.00 0.42 0.67 0.59 0.53 0.27 17104

avg / total 0.99 0.42 0.67 0.58 0.53 0.27 17205

Result from SMOTEENN

* Accuracy = 67%
* Precision = 99%
* Recall = 58%

pre rec spe f1 geo iba sup

high\_risk 0.01 0.67 0.42 0.01 0.53 0.29 101

low\_risk 1.00 0.42 0.67 0.59 0.53 0.27 17104

avg / total 0.99 0.42 0.67 0.58 0.53 0.27 17205

Result from BalancedRandomForestClassifier

* Accuracy = 70%
* Precision = 99%
* Recall = 58%

pre rec spe f1 geo iba sup

high\_risk 0.01 0.67 0.42 0.01 0.53 0.29 101

low\_risk 1.00 0.42 0.67 0.59 0.53 0.27 17104

avg / total 0.99 0.42 0.67 0.58 0.53 0.27 17205

Result from EasyEnsembleClassifier

* Accuracy = 93.16%
* Precision = 99%
* Recall = 94%

**Summary**

After reviewing and checking 6 different machine learning methods the best one suited to the data is the EasyEnsembleClassifier module from imblearn.ensemble. This method had the highest accuracy score of 93.16%.

**Recommendations**

There could be other, more efficient machine learning modules yet discovered which can better suit the needs of the project. Determining the parameters will ultimately yield if this module will work. Knowing the official accuracy range (96% to 99%) that is required to meet the needs of the department or group can help focus on 2 or 3 different machine learning methods which could be tested against real data for speed and accuracy.