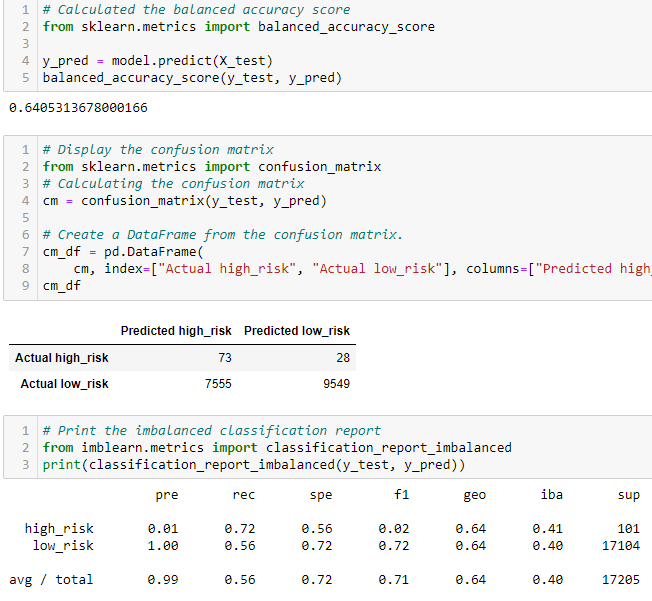
**Credit Risk Analysis**

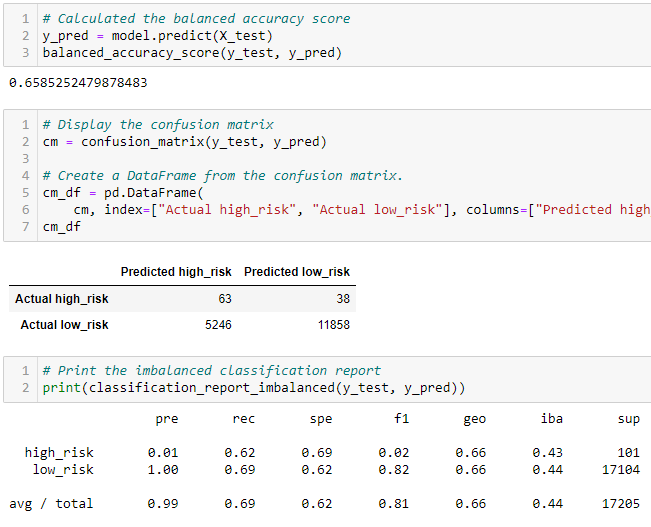
Using Supervised Machine Learning to Determine Credit Risk

**Results**

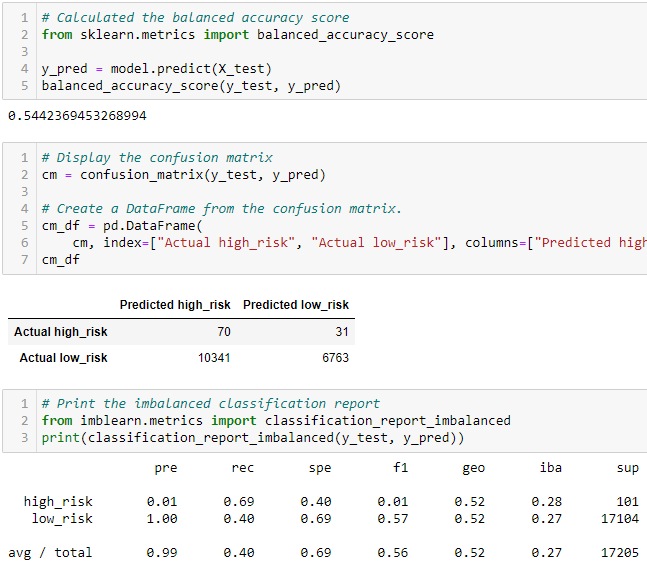
1. Naive Random Oversampling
   * Precision: High Risk = 0.01; Low Risk = 1.00
   * Recall: High Risk = .72; Low Risk = 0.56
   * Balanced Accuracy Score = 0.64
   * Accuracy score is low

[](https://user-images.githubusercontent.com/104606589/190297367-a45f5e66-6449-4db1-be94-6c5974900581.png)

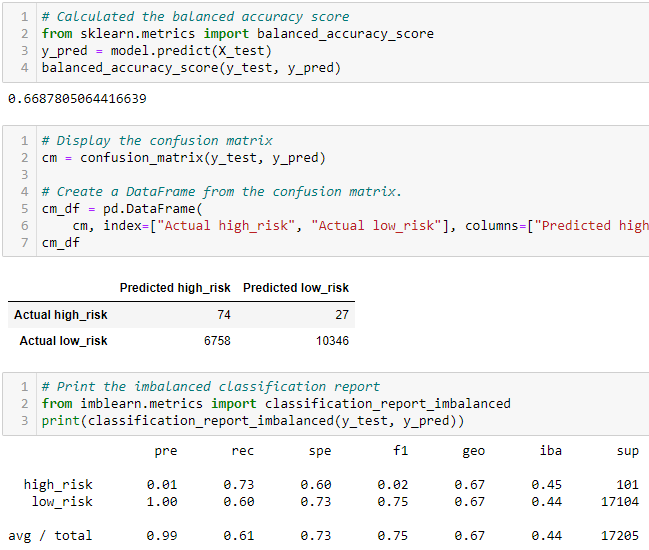
1. SMOTE Oversampling
   * Precision: High Risk = 0.01; Low Risk = 1.00
   * Recall: High Risk = .62; Low Risk = 0.69
   * Balanced Accuracy Score = 0.65
   * Accuracy score is low

[](https://user-images.githubusercontent.com/104606589/190297385-94e2cb07-b145-49e8-9ad3-3ef884e90c98.png)

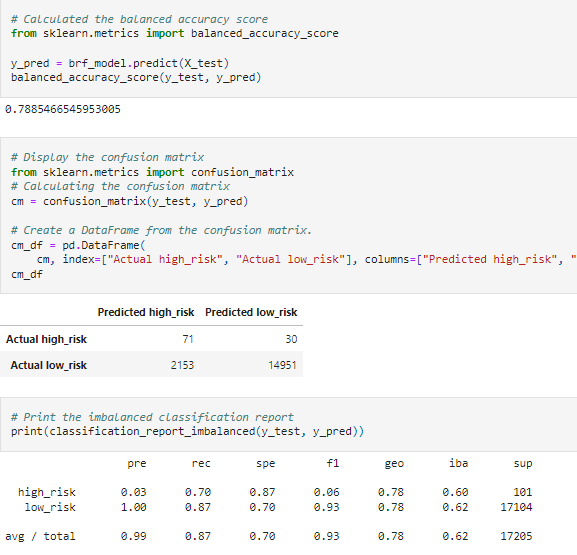
1. Cluster Centroids Under sampling
   * Precision: High Risk = 0.01; Low Risk = 1.00
   * Recall: High Risk = .69; Low Risk = 0.40
   * Balanced Accuracy Score = 0.54
   * Accuracy score is low

[](https://user-images.githubusercontent.com/104606589/190297415-498a9cf3-c182-4790-a5d8-230fb0c7746c.png)

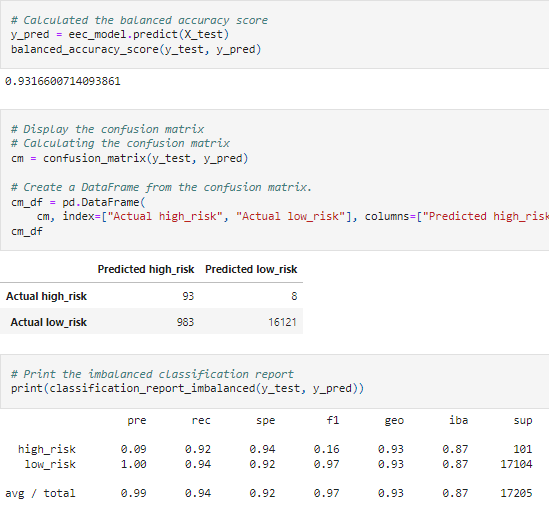
1. SMOTEENN (Combination of Over and Under sampling)
   * Precision: High Risk = 0.01; Low Risk = 1.00
   * Recall: High Risk = .73; Low Risk = 0.60
   * Balanced Accuracy Score = 0.66
   * Accuracy score is low

[](https://user-images.githubusercontent.com/104606589/190297469-d9008c73-acb0-4ee8-a5c4-53592330b3ce.png)

1. Balanced Random Forest Classifier
   * Precision: High Risk = 0.03; Low Risk = 1.00
   * Recall: High Risk = .70; Low Risk = 0.87
   * Balanced Accuracy Score = 0.78
   * Accuracy score is low

[](https://user-images.githubusercontent.com/104606589/189573035-8211d6e5-7b54-4e23-8537-ea2e4f72269a.png)

1. Easy Ensemble AdaBoost Classifier
   * Precision: High Risk = 0.09; Low Risk = 1.00
   * Recall: High Risk = .92; Low Risk = 0.94
   * Balanced Accuracy Score = 0.93
   * Accuracy score is low

[](https://user-images.githubusercontent.com/104606589/189573037-2f467176-c5d4-4802-ac14-6cc479b05606.png)

**Summary**

* At an accuracy of 0.54 Cluster Centroids Under sampling was the least successful algorithm in terms of accuracy.
* On the other side of the spectrum the Easy Ensemble method yielded the highest level of accuracy at 0.93.
* The Low-Risk precision didn't vary for any of the methods and sat a - 1.00.
* The best high risk precision score was with the Easy Ensemble method at 0.09.
* The recall scores for the Ensemble methods was better than the Oversampling/Under sampling/Combination algorithm.

**Conclusion**

When compared side by side for this population it would seem that the Easy Ensemble AdaBoost Classifier algorithm is the most effective strategy to determine credit risk. It had the highest level of accuracy and at 0.93. This is 15% higher than the next best strategy (Balanced Random Forest). The AdaBoost algorithm was able to predict high risk with a 0.92 level of accuracy and low risk with 0.94 level of accuracy. At scale this would save the company hundreds of thousands of dollars that the other models may not protect the company from. This level of accuracy would also greatly increase the level of customer satisfaction with far less rejections of qualified borrowers.