**RandomUnderSampler & RandomOverSampler:**

When performing the RandomUnderSampler function from the “imblearn” library on the training set, the training set was reduced from 16,932 to only 4,208 rows. The resulting accuracy scores dropped significantly which is why we didn’t continue to use random undersampling.

**Table:** Accuracy scores with undersampled training data

| **Classifier** | **Accuracy Score** |
| --- | --- |
| RandomForest | 0.79 |
| AdaBoost | 0.52 |
| LogisticRegression | 0.72 |
| KNeighbors | 0.71 |
| SVC | 0.78 |

We also attempted to use random oversampling. Unfortunately, when applying the RandomOverSampling function to the training set, the kernel crashed after several hours and due to time issues we didn’t follow this strategy of balancing the training data any further.