

# Solution

## Solution

Here is an explanation of the finer points of this milestone.

In Step 1, we have chosen 0.2 for the `test_size` parameter but the student can choose any other reasonable fraction up to 0.3. Also, we have intentionally chosen `stratify=y` as the parameter for the `train_test_split` method otherwise, we might run the risk of the train and the test data having different distributions.

In Step 2, knowing the baseline performance of a naive 'classifier' that predicts as belonging to the majority class is a critical step to do before undertaking any further predictive modeling. This is accomplished by a simple plotting of the distributions. We will also use the fraction of the minority class to tell the Novelty Detection algorithms how many outlier samples they would expect to find. We will encounter algorithms later where letting it know that it will expect to see 4x the actual number of minority samples, gives better average precision.

In Step 3, removing the outliers from the training data might sound counterintuitive to those of you who have done supervised machine learning techniques in the past. You might be wondering, "How will the algorithm learn to recognize the outliers if they don't even exist in the training data?!" This is because the one-class SVM is actually an example of a Novelty Detection algorithm that learns the distribution of the majority class better when there are no outliers. Hence the need to have the training data containing only the majority class.

For Step 9, we observe that one-class SVM with the RBF kernel has the best Average Precision and Recall while being reasonable with the no. of False

Positives relative to the two variants of the polynomial kernel.