Zach, Evan, Logan, Serena

CMPT300 Group Project

To choose which outliers to remove we calculate three standard-deviations out from the mean and remove any data points that are beyond 3 standard deviations in the dataset. There were approximately 1,800 outliers removed from our dataset. The remaining points that are outside of the boxes in the boxplot are still valuable data points that will be considered in our analysis. Using a decision tree classifier, we determined the best feature of our dataset. The best feature is x-edge with a value of .16. After training the model on 16,000 values, the accuracy of our test set of approximately 2,000 values is 87.6%.