**Project 2**

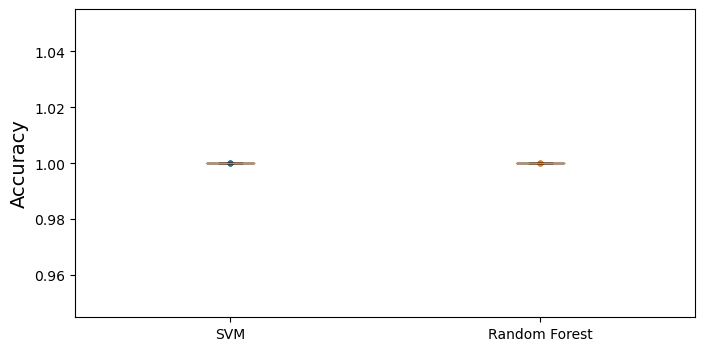
**Neural Networks/CSCI 441**

Mika Folstad, Jake Cronk, Devin Roelke

Interpretation:

At first glance, the models produced by this project appear to be exceptional, having literally perfect accuracy on the test data. However, this is too good to be true, and likely indicates a failure or flaw in the implementation of the models. It is possible that the model is overfitted, meaning that it performs very well on our test data but would struggle or underperform if deployed to a real-world environment. This could be a result of many things, the dataset could be too small for our implementation, the model could have too many trees or be too complex.

**Figure 1. Accuracy comparison of SVM and Random Forest Models.**



In figure 1, we see the comparison of accuracy results of our two models we tested (Support Vector Machine and Random Forest). Both returned the same results indicating perfect accuracy with zero spread. This could mean our data is well balanced, resulting in this similar performance.

Examining the feature importance values for the project reveals the source of the issue: the training data was not prepared properly, and contains the class column. After one-hot encoding, this column is divided into two columns that combined make up more than 0.78, or 78%, of the feature importance for the model. This is unsurprising, as those values together are one-to-one with the correct classification of each entry.

However, we can still learn from this output. If the class information is available to the model from the start, those should be the only two values of any importance. A combined importance of 0.78 could be considered relatively low in this context, meaning that the mdoel may not be very efficient at identifying important input variables.

Summary:

While this exercise did not go as planned, it was not a pointless effort. We demonstrated model training, evaluation, and tuning. While the incorrect data preparation did disrupt the validity of the model, this mistake was located through the use of feature importance.