With a one-vs-call classifier, we would build three separate binary classifiers on the dataset, one for each class. Here, the classifier attempts to differentiate between its concerned on one side, and every other class on the other, pitting one against the rest. Finally the class with the most votes is assigned to the datapoint.

In contrast, with an all vs all classifier, every class is pitted against every class. With combinatorial formulae, we can see that the total number of classifiers generated would then be n(n-1)/2, where n is the number of classes, 3 classifiers in this case, as have been built. Similarly, the class which gets the most votes after the datapoint has been run through all the classifiers is assigned.

Thus, one vs one would be more computationally expensive given the larger number of models to be built, but offers more insights when a direct comparison between classes is preferred over one with the entire data set.

Dealing with low amount of training data can often lead to overfitting, due to the potentially limited amount of variance among smaller amounts of data. To counter this, it is essential that we take into account only relevant features and drop other features which may lead to overfitting. We thus dropped the island attribute as it did not contribute as much to species classification.

It is also preferred to combine severa models to avoid high variance which may arise from using only one model. Using an all vs all classifier and then proceeding to poll all of them could help.

One can also avoid overfitting with smaller amount of data by using less compelx models such as a simple k-neighbours classifer which may produce lesser accuracy, but also lower variance as comparede to say random forests.

As part of feature engineering, we proceeded to normalize the data provided into intervals between 0 and 1, which we observed had a profound impact on the accuracy. For instance, the knn classifier returned an accuracy of around 0.75 when the data was

not normalized, as the model was skewed heavily towards features with a higher general magnitude, for in this case for instance the body mass was in the range os thousands, while the culmen length was largely lesser than a hundred.

Thus the data was normalized between the max and min of the attribute in question.

It was also observed that some null unfilled values were present within the tuples, and the relevant rows were thus dropped.

Furthermore, the features 'Sex' and 'Clutch Completion' were composed of string values. As part of feature engineering, since the attribute was binary, one value was assigned 1 and the other 0.

As part of feature selection, the Island and Clutch Completion were later removed as they were not relevant towards species classification, and the addition of too many attributes, especially when the magnitude of data was less, could have lead to overfitting.