The notebook starts by splitting the dataset into test and train using a seeded random number generator

Next the notebook trains a KNN Classifier from the Sklearn package and plots the predictions in matplotlib.

The plots show the predicted categories via colours and are plotted against two parameters on the axis. Incorrect predictions are plotted with an x

For KNNs the hyperparameters are the distance metric and the k number of neighbours. The notebook explores

The notebook highlights that is is important to experiment with the hyperparameters to fit the dataset in use and optimize them

Next the notebook trains a decision tree using the BaggingClassifier and experiments with number of tree, again a hyperparameter

Finally the notebook discusses the train and test split. The notebook ran on a 50:50 split but 80:20 would be better as this gives us more training data.

The split used will be conditional on the amount of data available