**Random Forest Classifier**

First Attempt parameters

A computer screen shot of a program code

Description automatically generated

In the first attempt of the model training, I made min samples leaf fixed as 1 only, and below is the best model performance:

A screenshot of a computer program

Description automatically generated

A computer screen shot of a black screen

Description automatically generated

The accuracy of the best model looks good with 95% accuracy. However, when I tried to plot the learning curve:

A graph of a graph showing the number of training scores

Description automatically generated with medium confidence

It has a variance between Training and cross-validation. It probably is an overfitting model.

A screen shot of a computer

Description automatically generated

The tree plot will be very huge when we allow 1 min sample leaf:



It almost expends the result for each case, which means that if we try to use some brand-new data to perform the prediction, the accuracy might not be ideal.

Therefore, in the second attempt, we change as below parameters:

A computer screen with text and images

Description automatically generated

We do not allow min sample leaf is 1 from the second attempt, it must have at least 2.

Best params from the second attempt

A screenshot of a computer program

Description automatically generated

A screen shot of a computer

Description automatically generated

A graph of a graph showing the results of a training

Description automatically generated with medium confidence

A screen shot of a computer

Description automatically generated



The more the value of max\_depth, the more complex the tree will be. On increasing the max\_depth value, the training error will decrease but it can result in inaccurate predictions on the test data (overfitting). Hence, the correct max\_depth value is the one that results in the best-fit decision tree — neither underfits nor overfits the data.

From the third attempt, we tried to minimize max depth to 20 and refit the model.

A computer screen shot of a program code

Description automatically generated

A computer screen shot of a black screen

Description automatically generated

A screen shot of a computer

Description automatically generated

A computer screen with text

Description automatically generated



Extra Experiment: only allow max depth with 10 at most

A computer screen shot of a program code

Description automatically generated

A screenshot of a computer program

Description automatically generated

A screen shot of a computer

Description automatically generated

A computer screen with text and numbers

Description automatically generated

A graph showing the results of a training course

Description automatically generated

A screen shot of a computer program

Description automatically generated