**ICE – 7**

1. **According to you, why do overfitting and underfitting occur, and how resolve them? What is the difference between them?**

The goal of Machine Learning model is to predict future problems with the help of the relations between the features of the past data we have. When we don’t have the required data and have less observation or less feature, model learns from only the given data which may be not accurate. This is called underfitting. When we have more data, but we select feature which are not required by the model or which are too specific about the data and apply the model to more accuracy, it might perform good on the train data, but on the test data, there will be less accuracy. This is called Overfitting.

We can utilize more complicated models, as well as more features and less noise, as well as more training data, to alleviate the over- and under-fitting model issues.

Our model's first inaccurate forecasts are also referred to as "Under-Fitting," and when that model does not produce accurate predictions to the data, it is referred to as "Over-Fitting."

1. **What kind of pattern did you analyze in the Train and Test score while running the code of overfitting?**

The trained model gave us the best result for the data, but when we tested it, we saw a poorer result than our training model. In this case, the situation might be characterized as over-fitted because the forecasted results are so much worse than the trained result. When there are positive findings in both the train and forecasted data, we can state that the model is flawless.

1. **What is cross-validation, and what did you analyze in a different type of validation that you performed?**

In order to determine the complimentary subset of the data we are evaluating the machine learning models on, many machine learning models will be trained on a subset of the input data. This process is known as cross validation.

With the use of cross validation, we are able to spot overfitting by observing differences in output accuracy caused by each validation iterator's unique setting to the data, which results in changes to the training and output prediction accuracy.

The validation curve is a graphical tool that may be used for a single hyper-parameter to quantify its influence, where it primarily indicates how much our model over-fits and how much it under-fits with both the train and forecast values.

1. **Explain the analysis from generated ROC and validation curve and what they represent?**

The ROC curve, which graphs the True positive rate and False positive rate with the data, will demonstrate the graphical performance of a classification model at all classification levels. Better performance is indicated by curves on the ROC Curve that are closer to the top-left corner.