Multiple Choices :

1. Overfitting more likely occurs when there is a huge data amount to train.

   A) True   B) False

2. Which among the below is true regarding hypothesis testing?

   A) Defines the relationship between the data and the model.

   B) Provides the estimates regarding numerical characteristics of data.

   C) Describes how the data behaves in the model developed.

   D) Answer yes or No Questions about the data

3. What are the Importances of Dimensionality reduction.

   A) This reduces the storage space and time for model execution.

   B) Makes it easier for visualizing data when the dimensions are reduced.

   C) Avoids the curse of increased dimensionality.

   D) All

4. Logistic Regression Default Threshold value.

   A) 0.2 B) 0.5 C) 0.7 D) 0.3

5. Expand KPI _____

Questions :

1.What do you understand about Imbalanced Data?

2.Explain clearly about the Confusion matrix and Type1,Type2 Error ?

3.What is the difference between Random forest and decision tree ?

4. What is the difference between Random forest in Regression and classification ?

5.What are RMSE and MSE in a linear regression ?

6.What is the cross validation score ?

7 .What is the difference between correlation and covariance ?

8.What is the difference between Entropy and Gini ?

9.Why data cleaning is important and How do you clean the data ?

10 . Explain the DataScience Life Cycle ?

11 . Explain MAE

12. Explain R2

13.Explain Adjust R2

14.what is loss function

15.what is cost function