MACHINE LEARNING-6

Answer of Following Questions-

Ans 1-B) Low R-squared value for train-set and High R-squared value for test-set.

Ans 2- B) Decision trees are highly prone to overfitting.

Ans 3- C) Random Forest

Ans 4-C) Precision

Ans 5-B) Model B

Ans 6-A) Ridge , D) Lasso

Ans 7-C) Random Forest, B) Decision Tree

Ans 8-A) Pruning

Ans 9-D) None of the above

Ans 10-The adjusted R-squared compensates for the addition of variables and only increases if the new predictor enhances the model above what would be obtained by probability. Conversely, it will decrease when a predictor improves the model less than what is predicted by chance

Ans 11-A regression model that uses L1 regularization technique is called Lasso Regression and model which uses L2 is called Ridge Regression. The key difference between these two is the penalty term. Ridge regression adds “squared magnitude” of coefficient as penalty term to the loss function.

Ans 12-A variance inflation factor (VIF) is a measure of the amount of multicollinearity in regression analysis. Multicollinearity exists when there is a correlation between multiple independent variables in a multiple regression model. As a rule of thumb, a VIF of three or below is not a cause for concern. As VIF increases, the less reliable your regression results are going to be

Ans 13-Scaling the target value is a good idea in regression modelling; scaling of the data makes it easy for a model to learn and understand the problem. Scaling of the data comes under the set of steps of data pre-processing when we are performing machine learning algorithms in the data set

Ans 14-There are three error metrics that are commonly used for evaluating and reporting the performance of a regression model; they are: Mean Squared Error (MSE). Root Mean Squared Error (RMSE). Mean Absolute Error (MAE)

Ans 15-sensitivity=0.96

specificity=0.8

precision=0.95

recall= 0.8

accuracy=0.88