MACHINE LEARNING

1.C) High R-squared value for train-set and Low R-squared value for test-set.
2.C) Decision trees are not easy to interpret
3.C) Random Forest
4.A) Accuracy
5.B) Model B
6.A) Ridge ,D) Lasso
7.B) Decision Tree,C) Random Forest
8.A) Pruning, C) Restricting the max depth of the tree
9.A) We initialize the probabilities of the distribution as 1/n, where n is the number of data-points
B) A tree in the ensemble focuses more on the data points on which the previous tree was not
performing well

10.
Adjusted R2 Penalizes You

The adjusted R2 will penalize you for adding independent variables that do not fit the model. In regression analysis, it can be tempting to add more variables to the data as you think of them. Some of those variables will be significant, but you can’t be sure that significance is just by chance. The adjusted R2 will compensate for this by that penalizing you for those extra variables.

11.
Lasso is a modification of linear regression, where the model is penalized for the sum of absolute values of the weights. Thus, the absolute values of weight will be (in general) reduced, and many will tend to be zeros. During training, the objective function become:

$$\frac{1}{2m}\sum_{i=1}^{m}(y - Xw)^2 + alpha\sum_{j=1}^{p}\left|w_j\right|$$

As you see, Lasso introduced a new hyperparameter, *alpha*, the coefficient to penalize weights.

Ridge takes a step further and penalizes the model for the sum of squared value of the weights. Thus, the weights not only tend to have smaller absolute values, but also really tend to penalize the extremes of the weights, resulting in a group of weights that are more evenly distributed. The objective function becomes:

$$\sum_{i=1}^{n}(y - Xw)^2 + alpha\sum_{j=1}^{p}w_j^{\,2}$$

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.

Detecting multicollinearity is important because while multicollinearity , it does reduce the statistical significance of the independent variables.

- VIF equal to 1 = variables are not correlated

- VIF between 1 and 5 = variables are moderately correlated

- VIF greater than 5 = variables are highly correlated

- VIF greater than10= problematic

The most suitable value is 1 to 5 while value 5 to 10 can be suitable as they need less attention value greater 10 are not so suitable

13.
If an algorithm uses gradient descent, then the difference in ranges of features will cause different step sizes for each feature. To ensure that the gradient descent moves smoothly towards the minima and that the steps for gradient descent are updated at the same rate for all the features, we scale the data before feeding it to the model.

14.

1. R Square/Adjusted R Square

2. Mean Square Error(MSE)/Root Mean Square Error(RMSE)

3. Mean Absolute Error(MAE)

15.

- Accuracy (all correct / all) = TP + TN / TP + TN + FP + FN
  =2501.2

- Precision (true positives / predicted positives) = TP / TP + FP
  =51

- Sensitivity aka Recall (true positives / all actual positives) = TP / TP + FN
  =251

- Specificity (true negatives / all actual negatives) =TN / TN + FP
  = 51

- Recall=TP/TP+FN
  =251