MACHINE LEARNING

1. D
2. A
3. A
4. D
5. A
6. B
7. B
8. B
9.
   GINI INDEX = 1 [(40/100)^2 + (60/100)^2] = 0.48
   ENTROPY = 0.97095

10.
    A decision tree is built on an entire dataset, using all the features of interest, whereas a random forest randomly selects observations and specific features to build multiple decision trees from and then averages the results. This means a random forest is a collection of many decisions trees used as weak learner to build up a strong learner. Random Forsest Algorithms is Powerful and highly accurate. We do not need to to normalize data for random forest, it will take care of the data. These Tress in the Random Forest runs parallel reducing the time complexity of our algorithm.
11.
    In Machine learning dataset we have various feature which vary of variour scale. For instance in a housing dataset, there could be a feature No of bedrooms with values like 2,3,4 etc… ad also have a feature area wih values like 2500,700,15000,'etc which is on a way different scale. Thus we would scale all these features to contraints theres values between a particulas range. The two techniques used for scaling the dataset are ;
    Standariation:
    In this algorithm the data is scaled on the scales between [–1,1]. The idea behind StandardScaler is that it will transform your data such that its distribution will have a mean value 0 and standard deviation of 1.
    Normalization:
    In this algorithm the data is scaled on the scales between [-3,3]. Normalization is a scaling technique in which values are shifted and rescaled so that they end up ranging between 0 and 1. It is also known as Min-Max scaling

12. The data must be scaled for gradient descent as If we don't scale the data, the level curves would be narrower and taller which means it would take longer time to converge, thus increasing time as well as space complexity since the increase of parameters.

13.

No, In case of a highly imbalanced dataset for a classification problem, accuracy is not a good metric to measure the performance of the model. Accuracy is metric to be used where the classes in the target variable are almost equal or class to each other. In case of imbalanced dataset we face the accuracy paradox. Consider an instance of fraud detection, we have 1000 records. Out of them only 10 of them will be fraud cases. In such cases , our machine learning model will be biased and predicted values of will be mostly TN . Very less cases will will be TP (i.e., actually fraud).And if in such cases if will use accuracy we will get biased results. In such cases will be get very high accuracy results which is false

14.

The F-score, also called the F1-score, is a measure of a model's accuracy on a dataset. It is used to evaluate binary classification systems, which classify examples into 'positive' or 'negative'.

The F-score is a way of combining the precision and recall of the model, and it is defined as the harmonic mean of the model's precision and recall.
The formula for the standard F1-score is the harmonic mean of the precision and recall. A perfect model has an F-score of 1.

F1-score = 2 / [1/recall * 1/precision]

F1-SCORE = TP / [TP + ½(FP + FN)]

15.

The fit() function is used for learing the data. In this step the model learns our data and calculated the weights and biases for our model. In the transform() function the learned model apply the results to an unknown dataset on transform them as per the weights they calculated while learning from the fit() function. The fit_transform() function is used to do both of the above steps in a single step . This is learn from the data , and the transform the same data all in one step.