1. A

2. A

3. B

4. A

5. A

6. C

7. B

8. B

9.

The formula for Gini Index= 1- Sigma (Pi2)

For Entropy = - Sigma (Pyi log2 (Pyi))

Therefore, Gini in this case = 0.48

And, Entropy = 0.292

10.

Decision trees are very prone to over-fitting. They are basically low-bias high-variance models is we let them grow till their maximum depth. The advantage of Random Forest over decision trees is it prevents over-fitting by ensembling weak learners and giving a low-bias low-variance model.

11.

We don’t always need to scale our data. For example, tree-based algorithms does not require scaling. Though, scaling is utmost important in case of distance-based algorithms like KNN, because if the data is on different scales, the distance calculated will not be a correct measure of near-ness or far-ness of two points, thus our model may fail.

Popular techniques of scaling are MinMaxScaler and StandardScaler.

12.

Optimization based algorithms try to find the optimal weights using gradient descent. Weights are nothing but numerical values corresponding to various features. Now, if the values are not scaled, there might be negligible change in some weights in gradient descent, while others may change rapidly. This might also cause our solution to diverge instead of converging at the minima. Scaling gets rid of all these problems for us and ensure that there is convergance at the local minima.

13.

No, In case of highly imbalanced data, accuracy score is not a good metric.

The problem lies in how accuracy is defined. It is the ratio of total correct classified points and total points in our dataset. Let’s understand it with help of an example. Suppose, out of 100 points, 95 points belong to class A and rest 5 belongs to class B. Now, if our model predicts every point to be of class A, the accuracy is still 95% (which is pretty great), but actually, the model is a dumb model because it mis-classified all the points in class B.

14.

F-score metric is designed to combine precision and recall into a single score.

The mathematical formula is,

2\*Precision\*Recall / (Precision + Recall)

15.

The fit function only fits the data. Transform function only transforms the data according to the pre-defined fit function. Whereas, fit\_transform first fits the data and simultaneously transforms the data in a single step.