10. What are the advantages of Random Forests over Decision Tree?

Random forest algorithm avoids and prevents overfitting by using multiple trees. The results are not accurate. This gives accurate and precise results. Decision trees require low computation, thus reducing time to implement and carrying low accuracy.

11. What is the need of scaling all numerical features in a dataset? Name any two techniques used for scaling.

Scaling is required to rescale the data and it's used when we want features to be compared on the same scale for our algorithm. And, when all features are in the same scale, it also helps algorithms to understand the relative relationship better.

12. Write down some advantages which scaling provides in optimization using gradient descent algorithm.

- We can use fixed learning rate during training without worrying about learning rate decay.
- It has straight trajectory towards the minimum and it is guaranteed to converge in theory to the global minimum if the loss function is convex and to a local minimum if the loss function is not convex.

13. In case of a highly imbalanced dataset for a classification problem, is accuracy a good metric to measure the performance of the model. If not, why?

No. Data Imbalance can range from small to huge differences in the number of instances of the classes. Small data imbalances such as 4:1, 10:1, etc., won't harm your model much, but as the data imbalance starts to increase to 1000:1 and 5000: it can create problems for your machine learning model.

14. What is “f-score" metric? Write its mathematical formula.

An F-score is the harmonic mean of Precision and Recall values of a system, and it answers to the following formula: 2 x [(Precision x Recall) / (Precision + Recall)]. Criticism around the use of F-score values to determine the quality of a predictive system are based on the fact that a moderately high F-score can be the result of an imbalance between Precision and Recall, ergo not telling the whole story. On the other hand, most systems, past a high level of quality, face a challenge when trying to improve one of the two indicators (Precision, Recall) without negative effects on the other. Critical (risk) applications that value the retrieving of information more than its precision (producing a large number of false positives, but virtually guaranteeing that all the true positives are found) can sometimes adopt a different scoring system called F2 measure, where Recall has a higher weight. The opposite (higher weight on Precision) is achieved by using the F0.5 measure.

15. What is the difference between fit(), transform() and fit_transform()?

The fit(data) method is used to compute the mean and std dev for a given feature to be used further for scaling. The transform(data) method is used to perform scaling using mean and std dev calculated using the . fit() method. The fit_transform() method does both fits and transform.