MACHINE LEARNING WORKSHEET-7

1. A
2. A
3. A
4. A
5. C
6. B
7. B
8. B

9. Entropy of dataset = -0.4 log(0.4) – 0.6 log (0.6) = -0.4(-1.32) – 0.6 (-0.73)=0.52 +0.43= 0.96
   Gini index= 1-[0.6*0.6 + 0.4 *0.4] = 0.48

10. Decision Trees are prone to overfitting whereas Random Forest removes this problem since it uses a voting mechanism to decide for the output hence combination of several decision trees reduces the overfitting problem.

    If you wanted to put all your decision making in one decision tree, it can get very complicated for you to not only create it but also to maintain it.Random forests on the other hand normalize overfitting risk by analysing a random sub-sample of your data.

11. Most of the times, your dataset will contain features highly varying in magnitudes, units and range. But since, most of the machine learning algorithms use Eucledian distance between two data points in their computations, this is a problem. Hence scaling the data on one scale is good before feeding it to the model to ensure model receives same kind of data within a scale.

12. Since gradient descent is focused on convergence at global maxima therefore if we have value with larger range(large magnitude) the process for gradient descent will happen very slowly increasing the time .

13. If dataset is imbalanced there are huge chances that model get trained on more percentage of one class or get validated on more percentage of one class and hence accuracy will be very high whereas model might be very poor

14. The F measure (F1 score or F score) is a measure of a test's accuracy and is defined as the weighted harmonic mean of the precision and recall of the test.

    F-Measure = (2 * Precision * Recall) / (Precision + Recall)

15. Fit() method first fit the data to the object and transform() is used to transform the data in accordance with the object and return transformed values. Fit_transform collectively does this task but if fit_transform is used then the object cannot be utilized to transform any other data whereas if fit and transform are used separately then the object ca be used to transform any other data in accordance with the original data fitted to the object.