MACHINE LEARNING – WORKSHEET 4

1) A

2) A

3) A

4) A

5) B

6) C

7) B

8) B

9) Entropy = -[4/10\*log(4/10) + 6/10\*log(6/10)

=0.288

Gini Index=1-[0.16+ 0.36]

= 1-0.52

=0.48

10)Random forests are a strong modeling technique and much more robust than a single decision tree. They aggregate many decision trees to limit overfitting as well as error due to bias and therefore yield useful results.

11)In many machine learning algorithms, Feature scaling is used to bring all features in the same standing, we need to do scaling so that one significant number doesn’t impact the model just because of their large magnitude.

Feature scaling in machine learning is one of the most critical steps during the pre-processing of data before creating a machine learning model. Scaling can make a difference between a weak machine learning model and a better one.

a) Standardization

b) Min-Max Scaling

13) NO , Accuracy is not a good metric to measure the performance of the model in case of a highly imbalanced dataset for a classification problem, as the data will be trained keeping the majority class that is high in number , and this will lead to overfitting .

14)In statistical analysis of binary classification, the F1 score (also F-score or F-measure) is a measure of a test's accuracy. It is calculated from the precision and recall of the test, where the precision is the number of correctly identified positive results divided by the number of all positive results, including those not identified correctly, and the recall is the number of correctly identified positive results divided by the number of all samples that should have been identified as positive.

15)The fit() function calculates the values of these parameters. The transform function applies the values of the parameters on the actual data and gives the normalized value. The fit\_transform() function performs both in the same step. Note that the same value is got whether we perform in 2 steps or in a single step.