

1. Two-step process of classification are _____ and _____
2. Attribute selection measures are also known as _____
3. Write the formula for accuracy
4. The information content in a training data set D is given by $\text{Info}(D) = \text{_____}$
5. Information Gain $\text{Gain}(A)$ by splitting on attribute A, for a training data set D is given by,
 $\text{Gain}(A) = \text{_____}$
6. Split Information value by splitting on attribute A, for a training data set D is given by,
 $\text{Splitinfo}_A(D) = \text{_____}$
7. Gain ratio on attribute A, for a training data set D is given by,
 $\text{GainRatio}(A) = \text{_____}$
 The splitting criterion Gini index which measures the impurity of training data set D is given
 by, $\text{Gini}(D) = \text{_____}$
8. What is the bias or shortcoming of Information gain criterion for attribute splitting?
9. What is the bias or shortcoming of Gain ratio criterion for attribute splitting?
10. What is prepruning in the context of decision tree construction?
11. What is postpruning of Decision Trees?
12. What is AVC in the context of decision tree induction from large data set?
13. Which decision tree construction algorithm uses AVC –set?
14. BOAT algorithm is based on the concept of B _____
15. Write Bayes theorem
16. Laplacian correction is done to avoid _____
17. Accuracy of a machine learning model, $\text{Accuracy} = \text{_____}$
18. Error-rate or misclassification rate = _____
19. Sensitivity of a machine learning model, $\text{Sensitivity} = \text{_____}$
20. Specificity of a machine learning model, $\text{Specificity} = \text{_____}$
21. Precision of a machine learning model, $\text{Precision} = \text{_____}$
22. F-1 score of a machine learning model, $F-1 = \text{_____}$
23. $F\beta$ - score of a machine learning model, $F-\beta = \text{_____}$
24. In ROC curves, ROC stands for _____
25. Name three ensemble methods.
26. What is oversampling w.r.t training dataset?
27. What is undersampling w.r.t training dataset?