

SRI KRISHNA INSTITUTE OF TECHNOLOGY

Bangalore

Department of Computer Science & Engineering

Subject: Data Mining and Data Warehousing

Sub Code: 18CS641

Assignment -4

1. What is classification. Explain the general approach for solving a classification problem with an example.
2. How decision trees are used for classification. Explain decision tree induction algorithm for classification.
3. Write Hunts algorithm and illustrate it's working.
4. Explain the Methods for Expressing Attribute Test Conditions.
5. Explain various measures for selecting the best split with an example.
6. Explain the importance of evaluation criterion for classification methods.
7. Explain the characteristics of decision tree Induction.
8. Explain Model Over fitting. What are the reasons for overfitting? How to address overfitting problems
9. Explain how to estimate generalization errors.
10. Give the difference between rule-based ordering and class-based ordering scheme.
11. Discuss the different measures for selecting the best split. Also compute three impurity measures for binary classification problems for the following three nodes:

Node number	Count if class-0	Count of class-1
N1	0	6
N2	1	5
N3	3	3

12. Consider the training set that contain 100 +ve and 400 -ve examples for each of the following candidate rule. Determine which is the best and worst candidate according to

1) Rule accuracy 2) Foil information gain 3) Laplace

R1: A → + (covers 4 +ve and 1 -ve examples)

R2: B → + (covers 30 +ve and 10 -ve examples)

R3: C → + (covers 100 +ve and 90 -ve examples)