

Duration: 3 hours

[Max Marks:80]

- N.B. : (1) Question No 1 is Compulsory.
(2) Attempt any three questions out of the remaining five.
(3) All questions carry equal marks.
(4) Assume suitable data, if required and state it clearly.

Q.1

[20]

- (a) What is the difference between classification and prediction? Give examples of each.
- (b) What is decision support system
- (c) How to improve efficiency of Apriori algorithm?
- (d) What is K-Means clustering?

Q.2

[10]

- (a) Suppose that a data warehouse consists of the three dimensions time, doctor, and patient, and the two measures count and charge, where charge is the fee that a doctor charges a patient for a visit. Draw a star schema diagram for the above data warehouse. Starting with the base cuboid [day, doctor, patient], what specific OLAP operations should be performed to list the total fee collected by each doctor in 2004?
- (b) Given a data set of 5 elements: [10, 20, 30, 40, 50]. Apply Min-Max Normalization (Range [0, 1]) and give normalized value for each input data element. Also apply z-score normalization and give the normalized value for input 30.

Q.3

[10]

- (a) Explain the following OLAP operations with examples: Roll-up, Drill-down, Slice, Dice, Pivot
- (b) What is attribute selection in decision trees? Explain Information Gain and Gini Index with formulas.

Q.4

[10]

- (a) What are ensemble methods? Why are they useful in classification tasks?
- (b) What is BIRCH (Balanced Iterative Reducing and Clustering using Hierarchies) algorithm?

[10]

- Q.5 (a) Consider the transaction details as given below. Apply Apriori algorithm with minimum support of 60% and confidence of 75%. Find all the frequent itemsets and all the association rules. [10]

| Transaction ID | Items Bought |
|----------------|---------------------|
| T1 | Milk, Bread |
| T2 | Milk, Diaper |
| T3 | Milk, Bread, Diaper |
| T4 | Bread, Diaper |
| T5 | Milk, Bread |

- (b) Describe any two outlier detection methods. [10]

- Q.6 [20]

- (a) Given scores of 10 students: [45, 50, 55, 58, 60, 65, 70, 72, 75, 80] Create a histogram with 3 bins. And Show bin ranges and frequency counts in graphical form.
- (b) Given a confusion matrix:

| | Predicted Yes | Predicted No |
|------------|---------------|--------------|
| Actual Yes | 50 | 10 |
| Actual No | 5 | 35 |

Calculate: Accuracy and Precision

- (c) What are the key differences between Business Intelligence and traditional reporting systems?
- (d) Write a short note on simple linear regression
