



Printed Pages : 3

MCA – 404(5)

(Following Paper ID and Roll No. to be filled in your Answer Book)

**PAPER ID : 1479**

Roll No.

**M. C. A.**

(SEM. IV) EXAMINATION, 2006-07

**DATA MINING & WAREHOUSING**

Time : 3 Hours]

[Total Marks : 100

Note : Attempt *all* questions.

**1** Attempt any **two** : **2×10=20**

- (a) (i) Define the data warehouse
- (ii) How is a data warehouse different from a database? How are they similar?
- (b) (i) What are the different phases of the knowledge discovery from data base?
- (ii) What data mining attempts to facilitate?
- (c) What are schemas. Discuss various schemas used in data warehouse.

**2** Attempt any **two** : **2×10=20**

- (a) What is OLAP? Explain various OLAP operations used on a data cube.
- (b) Apply the Apriori algorithm on the following data set and find the maximum frequent item set. Set having support **20%** or more are frequent item set.  
Show two association rules that have a confidence **70%** or greater.

V-1479]

1

[Contd...

<i>Trans ID</i>	<i>Items Purchased Sold</i>
10	I1, I2, I4
20	I1, I7
30	I7, I5
40	I1, I2, I5
50	I6, I4
60	I6
70	I6, I7
80	I1, I2, I3, I4
90	I3, I5
00	I1, I2

- (c) Define the Association rule mining. How market basket analysis forms the association rules? Discuss basic concepts.

3 Attempt any **four** : **4×5=20**

- (a) What are classification rules and how are decision trees related to them ?
- (b) What is data classification? How it is differ than prediction ?
- (c) Describe the **ID3** algorithm of the decision tree construction. Why is it unsuitable for data mining applications?
- (d) How hypothesis testing and refinement task can be done in datamining using gentic algorithm.
- (e) Describe neural networks techniques for data mining. What are the main difficulties in using these techniques ?
- (f) What is Bayesian classification ? How it classifies the Input data ?

V-1479]

2

[Contd...

- 4** Attempt any **two** parts : **2×10=20**
- (a) How does clustering differ from classification?
  - (b) What is supervised and unsupervised learning ? Why clustering is known as unsupervised learning ?
  - (c) Describe the genetic algorithms as data mining techniques. What are the main difficulties in using these techniques ?
- 5** Attempt any **two** parts : **2×10=20**
- (a) What is backpropagation Neural Network topology? How it is used in classification?
  - (b) Compare hierarchical clustering and non hierarchical clustering algorithm. Explain the advantages and disadvantages over each other.
  - (c) Write differences between the Nearest Neighbour Data Mining Techniques and Clustering.
-