2005 ANDHRA UNIVERSITY M.C.A COMPUTER

MCA 3.1.5

DATA WARE HOUSING AND DATA MINING

Elective III

Time: 3 Hrs.
Max. Marks: 100

First Question is Compulsory

Answer any four from the remaining

Answer all parts of any Question at one place.

1. Briefly discuss.
a. Correlation analysis for handling redundancy.
b. Discretization
c. Advantages of ROLAP and MOLAP
d. Ice-berg query.
e. Constraint –based rule mining
f. Scalability of an algorithm
g. Cross table reporting
h. Slicing operations
i. Reasons for data partitioning
j. Components of five-number summary

2. a) What is data mining? Briefly describe the components of a data mining system.
b) What kinds of patterns can be identified in a data mining system?

3. a) Write the differences between operational database and data warehouse.
b) Briefly describe 3-tier Data warehouse architecture

4. a) Write different approaches to data transformation.
b) Propose an algorithm in pseudo-code for automatic generation of a concept hierarchy for categorical data based on the number of distinct values of attributes in the given schema.

5. a. Discuss the essential features of a typical data mining query language like DMQL.
b. Consider association Rule below, which was mined from the student database at Big- University: Major(X,"science") status(X,"undergrad").
Suppose that the number of students at the unive rsity (that is, the number of task-relevant data tuples) is 5000, that 56% of undergraduates at the university major in science, that 64% of the students are

registered in programs leading to undergraduate degrees, and that 70% of the students are majoring in science.

a. Compute the confidence and support of above rule.
b. Consider Rule below:
Major(X,"biology") status(X,"undergrad"). [17%,80%]
Suppose that 30% of science students are majoring in bioogy. Would you consider Rule 2 to be novel with respect to Rule 1? Explain.

6. a. Discuss why attribute relevance analysis is needed and how it can be performed.
b. Outline a data cube-based incremental algorithm for mining analytical class comparisons.

7. Write the A priori algorithm for discovering frequent item sets for mining single-dimensional Boolean Association Rule and discuss various approaches to improve its efficiency.

8. a. Discuss the back propagation algorithm for neural network-based classification of data.
b. What are the different categories of clustering methods?