CS - 8203
B.E. VIII Semester
Examination, June 2014
Data Mining and Knowledge Discovery
Unit - I

1 (a) Explain the concept that metadata is like a newe center. Describe how the concept applies to the data warehouse environment.
(b) Explain the snowflake schema with the help of examples.

Or

2 (a) Discuss system development life cycle of a data warehouse. What factors should bepconsidered while designing a data warehouse.
(b) Explain how process model and data model can apply to the architecture environment? Why process model is not suitable for data Warehouse.

Unit II

3 In data Warehouse technology, a multiple dimensional view can be implemented by a ROLAP, or by MOLAP or by HOLAP.
a) Briefly describe each implementation technique.
b) For each technique explain how each of the following functions may be implemented:
i) Roll-up ii) Drill down iii) Thermental updating
Which implementation technique do you prefer and why?

Or

4 a) Discuss typical OLAP operations in brief. 7
b) Why most data warehouse system support index structures? Discuss methods to index OLAP data, 7

Unit – III

5 a) What is data mining? Discuss its various strategies. Also list out its applications. 7
b) Briefly suggest some important guidelines for successful data mining? 7

Or

6. a) What do you mean by data reduction? What are the strategies of the data reduction? 7
b) Briefly explain: 7
i) Data cleaning
ii) Data integration.

Unit - IV

7 a) What is the 'Apriori property'? How is it used by the ARIORI algorithm? What are the drawbacks of the Apriori algorithm? 7
b) Give a brief note on mining frequent patterns without candidate generation? 7

Or

8. Given are the following eight transactions on items {A,B,C,D,E}:

tid	Items
1	{A, B}
2	{A;B,C}
3	{B, C, D}
4	{B, C}
5	{A,B, C,D}
6	{B ,D}
7	{B, E}
8	{B,D, E}

Use the apriori algorithm to compute all frequent item sets,and their support, with minimum support as 3 clearly indicate the steps ofalgorithm. Give all generation of closed frequent item sets and their closure. 14

Unit- V

9 a) Discuss Naive Bayesian classification. Why is it called as "naive". 7
b) What is Hierarchical method of clustering? Differentiate
Agglornerative and Divisive Hierarchical clustering? 7

Or

10 Write short notes on: 14
i) Decision Trees Rules ii) Tree Induction Algorithm iii) Overfitting and pruning iv) Split algorithms based on Gini Index.