USN | | | | | | | | | | |

07MCA542

Fifth Semester MCA Degree Examination, May/June 2010
Data Mining

Time: 3 hrs. Max. Marks:100

Note: ***Answer any FIVE full questions.***

1 a. What is data mining? Explain the data mining tasks and different types of attributes. **(10 Marks)**

b. "Data mining is an integral part of knowledge discovery in databases"- Justify. Also explain motivating challenges. **(10 Marks)**

2 a. Define data pre – processing. Mention the approaches of data pre – processing and briefly explain them. **(10 Marks)**

b. i) Calculate SMC and J for the binary vectors
x = (1, 0, 0, 0, 0, 0, 0, 0, 0, 0) and y = (0, 0, 0, 0, 0, 0, 1, 0, 0, 1).
ii) Calculate cosine similarity for the data objects
x = (3, 2, 0, 5, 0, 0, 0, 2, 0, 0) and y = (1, 0, 0, 0, 0, 0, 0, 1, 0, 2). **(10 Marks)**

3 a. What is a decision tree? Give the Hunt's algorithm for a decision tree induction and illustrate it using a training set. **(10 Marks)**

b. i) Explain rule based classifier ii) Calculate the different impurity measures for the node N using the data given below. **(10 Marks)**

Node N	Count
Class = 0	1
Class = 1	5

4 a. Discuss frequent itemset generation by means of Apriori principle. **(10 Marks)**

b. i) Construct an FP – tree for the following transaction data set.

TID	Items
1	{a, b}
2	{b, c, d}
3	{a, c, d, e}
4	{a, d, e}
5	{a, b, c}
6	{a, b, c, d}
7	{a}
8	{a, b, c}
9	{a, b, d}
10	{b, c, e}

ii) Explain the criteria for merging sequences, with an example. **(10 Marks)**

5 a. What is cluster analysis? What are the applications of cluster analysis? **(10 Marks)**

b. Describe the density based clustering with the algorithm. **(10 Marks)**

6 a. Explain spatial data mining. **(10 Marks)**

b. Briefly explain audio and video data mining and web usage mining. **(10 Marks)**

7 a. Briefly explain data mining applications. **(10 Marks)**

b. List the different features considered while choosing a data mining system and explain them. **(10 Marks)**

8 Write short notes on :

a. The origins of data mining b. Association rule.

c. Graphs and subgraphs d. Different types of clusters. **(20 Marks)**

*****