3

USN □□□□□□□□□□

08IS401

for ISE only

PES Institute of Technology, Bangalore

(Autonomous Institute under VTU, Belgaum)

DECEMBER 2011 SEMESTER END EXAMINATION (SEE) B. E. 7th SEMESTER ISE

08IS401 - DATA WAREHOUSING AND DATA MINING

Time: 3 Hrs | Answer All Questions | Max Marks: 100

1. a) Describe by means of a diagram the process of Knowledge Discovery of Databases (KDD) — 8

b) Indicate any two major issues in Data Mining — 2

c) Suppose a hospital tested the age and body fat for 18 randomly selected adults with the following result:

age	23	23	27	27	39	41	47	49	50
%fat	9.5	26.5	7.8	17.8	31.4	25.9	27.4	27.2	31.2

age	52	54	54	56	57	58	58	60	61
%fat	34.6	42.5	28.8	33.4	30.2	34.1	32.9	41.2	35.7

(i) Calculate the mean, median, and standard deviation of age and %fat — 5

(ii) Draw the box plots for age and %fat — 5

2. a) Compare OnLine Transaction Processing (OLTP) and OnLine Analytical Processing (OLAP). — 5

b) Describe briefly Relational OLAP (ROLAP) Server, Multi Dimensional OLAP (MOLAP) Server, Hybrid OLAP (HOLAP) Server and Specialized SQL Servers. — 5

c) Suppose the base cuboid has three dimensions, A,B,C with the following number of cells :
|A| = 10,000,00 |b| = 100 and |C| = 1000. Suppose that each dimension is evenly partitioned into 10 portions for chunking

(i) Assuming each dimension has only one level, draw the complete lattice of the cube — 3

(ii) If each cube cell stores one measure with 4 bytes, what is the total size of the computed cube if the cube is dense? — 3

(iii) State the order for computing the chunks in the cube that requires the least amount of space, and compute the total amount of main memory space for computing the 2-D planes. — 4

3. a) What is Support and Confidence ? Give an example. — 4

b) A database of transactions in a book mart is as follows: Let min-sup = 25%

Trans_ID	Items
101	Book, Pen, Eraser
102	Pen, Pencil
103	Notebook, Book, Pen, Eraser
104	Book, Pen
105	Book, Notebook, Eraser

Using the abbreviations B for Book, P for Pen, E for Eraser, PN for Pencil and N for Notebook, find all frequent itemsets using Apriori algorithm Construct FP Tree, conditional pattern base and conditional FP Tree. — 10

c) Describe briefly the three strategies for defining minimum threshold levels at multiple level of abstraction, for mining multi –level association rules. — 6

USN □□□□□□□□□□

08IS401

for ISE only

<table>
<tr><td>4.</td><td>a)</td><td>Compare Classification and Prediction Methods in terms of accuracy, speed, robustness, scalability and interpretability</td><td>5</td></tr>
<tr><td></td><td>b)</td><td>Describe Bayesian Belief Networks by means of a figure.</td><td>5</td></tr>
<tr><td></td><td>c)</td><td>Describe by means of a figure, Support Vectors and Maximum Margin in the context of Support Vector Machines (SVM)?</td><td>5</td></tr>
<tr><td></td><td>d)</td><td>Give the table for X (years of experience) and Y (corresponding salary of engineers in Rs.thousands) predict the salary of engineers with 12 years of experience using linear regression technique
<table>
<tr><td>1</td><td>20</td></tr>
<tr><td>3</td><td>36</td></tr>
<tr><td>6</td><td>43</td></tr>
<tr><td>8</td><td>57</td></tr>
</table></td><td>5</td></tr>
<tr><td></td><td></td><td></td><td></td></tr>
<tr><td>5.</td><td>a)</td><td>Describe k-Means algorithm by means of a figure.</td><td>6</td></tr>
<tr><td></td><td>b)</td><td>A relational table where patients are described by binary attributes is given below:
<table>
<tr><td>Name</td><td>gender</td><td>fever</td><td>cough</td><td>test-1</td><td>test-2</td><td>test-3</td><td>test-4</td></tr>
<tr><td>Aleem</td><td>M</td><td>Y</td><td>Y</td><td>P</td><td>N</td><td>N</td><td>N</td></tr>
<tr><td>Ayan</td><td>F</td><td>Y</td><td>N</td><td>P</td><td>N</td><td>P</td><td>N</td></tr>
<tr><td>Alex</td><td>M</td><td>Y</td><td>Y</td><td>N</td><td>N</td><td>N</td><td>N</td></tr>
</table>
Compute the distance between the each pair of the three patients Asha, Ayan and Abdul</td><td>3</td></tr>
<tr><td></td><td>c)</td><td>Given two objects represented by tuples (22, 1, 42, 10) and (20, 0, 36, 8) compute the following
i. Euclidean distance between the two objects.
ii. Manhattan distance between the two objects.
iii Minkowski distance between the two objects, using the power q=3.</td><td>5</td></tr>
<tr><td></td><td>d)</td><td>Describe briefly any one data mining application</td><td>6</td></tr>
</table>