USN ☐☐☐☐☐☐☐☐☐☐

NEW SCHEME

Fifth Semester MCA Degree Examination, Dec. 06 / Jan. 07
Master of Computer Applications
Data Mining and Warehousing

Time: 3 hrs.]   [Max. Marks:100

Note: *Answer any FIVE full questions.*

1 a. What is a data warehouse and what are its four key features?   (05 Marks)
   b. Compare OLTP and OLAP systems.   (05 Marks)
   c. For a sale example, draw star-schema and snowflake schema and compare them.
                                                                           (10 Marks)

2 a. Describe a 3-tier data warehousing architecture.   (08 Marks)
   b. Explain On-Line Analytical Mining.   (06 Marks)
   c. What are the topics involved in data transformation? If the minimum and maximum
       values of an attribute cost is Rs.12,000 and Rs.96,000 respectively, what is the
       normalized value of Rs.75,000?   (06 Marks)

3 a. Describe data cube aggregation and dimension reduction methods.   (10 Marks)
   b. Define confidence and support with respect to transaction items. Also explain the
       syntax for specifying the kind of knowledge to be mined.   (06 Marks)
   c. Explain any four methods of presenting or visualizing the discovered patterns with
       an example.   (04 Marks)

4 a. Describe the generation of candidate itemsets and frequent itemsets where the
       minimum support count is 2 for the following transactional data: (08 Marks)

TID	List of item-IDs
T100	11, 12, 15
T200	12, 14
T300	12, 13
T400	11, 12, 14
T500	11, 13
T600	12, 13
T700	11, 13
T800	11, 12, 13, 15
T900	11, 12, 13

   b. Write apiori algorithm for finding frequent item sets and generate the association
       rules for the above example.   (12 Marks)

5   a.   Explain Bayes theorem with its formula. The probability of having a fever given that person has a cold is 0.45 and the probability of fever in the general population is 0.03. Determine the probability that the person has a cold given that he or she has a fever, if the prior probability of a person having a cold is 0.01. **(04 Marks)**

     b.   Describe the back propagation method in a neural network, with a block schematic and an algorithm. **(12 Marks)**

     c.   Differentiate between NNC and KNNC. **(04 Marks)**

6   a.   Describe linear regression method. **(06 Marks)**

     b.   Define Euclidean and Manhattan distance. What are the four requirements of a distance function? If $V_1 = (10, 5)$ and $V_2 = (8, 9)$, determine the Euclidean and Manhattan distance values. **(06 Marks)**

     c.   Describe K-medoids algorithm with a pseudocode. **(08 Marks)**

7   a.   Differentiate between agglomerative and divisive hierarchical clustering techniques. **(02 Marks)**

     b.   Describe BIRCH clustering method. **(07 Marks)**

     c.   Explain decision tree classifier. **(07 Marks)**

     d.   Discuss any two methods to detect outliers based on the distance. **(04 Marks)**

8   a.   Describe Chameleon and DBSCAN clustering techniques. **(10 Marks)**

     b.   What are the five DNA analysis topics for which the data mining techniques are used? **(05 Marks)**

     c.   How data mining techniques can be applied to telecommunication industry? **(05 Marks)**

*****