No. of Printed Pages : **2**

Sl. No.

BE6-R4 : DATA WAREHOUSE AND DATA MINING

NOTE :

1. **Answer question 1 and any FOUR from questions 2 to 7.**
2. **Parts of the same question should be answered together and in the same sequence.**

Time: 3 Hours Total Marks: 100

1. (a) Explain steps of Data Warehouse design process.

(b) How are multiple facts handled ? What are the schemas used for handling a single fact and for using multiple facts ?

(c) Are all strong association rules interesting ? Justify with example.

(d) How are concept hierarchies useful in data mining tasks ? Describe briefly.

(e) Discuss issues to consider during data integration.

(f) What is the difference between supervised learning and unsupervised learning ? Which of the two learning methods is also termed as "Learning by Examples" ?

(g) What are the various dimensions to access the quality of data ? Explain any two with an appropriate example. **(7x4)**

2. (a) What is the difference between Online Transaction Processing (OLTP) and Online Analytical Processing (OLAP) ? Explain with example. Also group the following under OLTP and OLAP : Analysis, upto date, consolidated, lots of scan, high performance, high flexibility

(b) Explain the three tier architecture of data warehouse. **(10+8)**

3. (a) What is Neural Network ? Briefly describe multilayer-feed forward neural network.

(b) What are the various steps involved in data transformation for making data suitable for mining process ? Explain data transformation using an example.

(c) Write algorithm for K-Nearest Neighbor. Why is it called a Lazy Classifier ? **(6+6+6)**

4. (a) Explain the Join step and the prune step as used in Apriori Algorithm.

(b) What are Multidimensional Association Rule ? Explain in brief. **(10+8)**

5. (a) Define : Support, Confidence. A following database has five transactions. Let min support = 60% and min confidence = 80%

TID	Items bought
T100	{M, O, N, K, E, Y}
T200	{D, O, N, K, E, Y}
T300	{M, A, K, E}
T400	{M, U, C, K, Y}
T500	{C, O, O, K, I, E}

Find all frequent item sets using Apriori algorithm.

(b) What is the connection between computations of Minkwoski, Euclidean and Manhanttan distance ? Compute Euclidean distance between each pair of the points given below and show in form of a distance matrix

A(2,5), B(3,2), C(7,2), D(6,2),E(1,1)

(c) What is the importance of correlation analysis ? Also give an example to show that "Correlation does not imply causality". **(6+6+6)**

6. (a) How crossover and mutation are performed in genetic algorithms ? Explain them with example.

(b) Write a short note on : "data mining packages and data warehousing packages"

(c) Explain : Web Content Mining and Web Usage Mining. **(6+6+6)**

7. (a) What is cluster analysis ? What are typical applications of clustering ? What are the typical requirements of clustering ?

(b) Describe why fuzzy set approach is needed in rule bases system and also explain the fuzzy set approach in detail.

(c) Explain : Multimedia databases, Time Series databases **(6+6+6)**

- o 0 o -