Total No. of Printed Pages:2

SUBJECT CODE NO:- P-19
FACULTY OF ENGINEERING AND TECHNOLOGY
B.E.(CSE/ IT) Examination May/June 2017
Data Warehousing & Data Mining (CSE/IT)
(Revised)

[Time: Three Hours] **[Max.Marks:80]**

Please check whether you have got the right question paper.

N.B (i) Q.1 and Q.6 are compulsory. Solve any two questions from the remaining in each section.
(ii) Assume suitable data if necessary and state it clearly.

Section A

Q.1 a) What is Fact Table? 03
b) What is the structure of Dimension Table? 03
c) What is the difference between OLTP and OLAP? 04

Q.2 a) What is Multi-Dimensional Modeling? What is the use of Snowflake schema? 08
b) Describe the functions of various components in a typical Multi-tiered Data Warehouse architecture with the block diagram 07

Q.3 a) How is a Data Mining System integrated (coupled) with DW/DB system? 08
b) Explore and explain the use of data mining is in Web Search Engine? 07

Q.4 a) What are the methods of measuring data dissimilarity between objects of mixed types? 07
b) What is data dissimilarity? Two objects are represented by the tuples (22, 2, 10,6) and (20,0, 12, 8): 08
i) Compute the Euclidean distance between the two objects.
ii) Compute the Manhattan distance between the two objects.
iii) Compute the Minkowski distance between the two objects, using q=3

Q.5 a) What are the statistical parameters to measure central tendency of the data? 07
b) What are the major issues in data mining? 08

Section B

Q.6 a) What is the role of Confusion Matrix for Classifiers? 03
b) Define the Terms – Entropy and information Gain 03
c) What is the use of k-medoids algorithm? 04

Q.7 a) Give the working principle of Rule based Classifier-Using IF-THEN Rules. How do we calculate the accuracy and coverage of this classifier 08
b) What is the method of extracting rules from Decision Tree? 07

Q.8 a) Using Apriori algorithm find frequent itemsets for database given below. Use support =2. Generate association rules using confidence = 80% 10

TID	**Items_bought**
T10	{K,E,Y, M, O, N}
T20	{K,I,E, O, O, C}
T30	{K, E, Y, N, O, D}
T40	{C,K, E, M, A}
T50	{I, C, Y, M.K}

b) What is the general approach of classification using two phases- i)Learning ii) Testing 05

2017

Q.9	a) Cluster following points in three clusters. Take initially A1, B1 and C1 as Centre points Use k-means algorithm and show the final clusters formed (Use Euclidean distance.) A1(2,10), A2(2,5), A3(8,4), B1(5,8), B2(7,5), B3(6,4), C1(1,2), C2 (4,9).	10
	b) What do you understand by – Prediction, Classification, Clustering, Regression	05
Q.10	a) What are the steps for a successful BI implementation?	07
	b) What is the process of intelligence creation and use and BI Governance	08