KUMARAGURU
college of technology
character is life

Register Number:……………………..

MCA DEGREE EXAMINATIONS: MAY 2015

(Regulation 2009)

Fifth Semester

MASTER OF COMPUTER APPLICATIONS

MCA528: Data Warehousing & Data Mining

Time: Three Hours   **Maximum Marks: 100**

Answer all the Questions:-

PART A (10 x 2 = 20 Marks)

1. What are the major challenges of mining huge amounts of data?

2. Why is outlier detection and analysis an interesting data mining task?

3. What is fact constellation schema?

4. Why concept hierarchies are useful in data mining?

5. Suppose the minimum and maximum values of salaries for the attribute *income* are Rs.11,000 & Rs.98,000 respectively. Use min-max normalization to map *income* of Rs.43,567 in the range [0.0 , 1.0].

6. What is the need for OLAP?

7. Given two objects represented by the tuples (23, 2, 44, 11) and (20, 0, 22, 8). Compute the Minkowski distance the two objects using q=3.

8. Differentiate between training and test data.

9. What does precision mean in text retrieval?

10. State any four applications of data mining.

PART B (5 x 16 = 80 Marks)

11. a) i)   Explain the steps in data mining.   8

       ii) Data mining is a multidisciplinary field. Justify.   8

(OR)

   b) i)   With a neat diagram explain the KDD process. Explain the significance of each step of the KDD process.   8

       ii) Explain the functionalities of data mining with examples.   8

12. a) i) Enumerate the three classes of schema that are popularly used for modeling 8
data warehouses.

ii) Suppose a data warehouse consist of three dimensions: *time, doctor and* 8
patient, and two measures, *count* and *charge*, where *charge* is the fee that the
doctor charges a patient for the visit.
   a) Starting with the base cuboid, [day, doctor, patient], what specific
   OLAP operations should be performed in order to list the total fee
   collected by each doctor in 2013.
   b) What OLAP operation should be performed to get the fee collected by
   Dr.Ramu and Dr.Somu during September 2013?

(OR)

b) i) Explain about data warehouse implementation with a neat diagram. 10

ii) Explain about MOLAP servers in detail. 6

13. a) i) Explain the various data preprocessing techniques in detail. 16

(OR)

b) i) Explain the apriori algorithm. Given a sample transactional database X:
(min – support = 2)

TID	Items_bought
T101	A,B,C,D
T102	A,C,D,E
T103	A,D,E,B
T104	A,B,D,E
T105	A,C,E

12

Use the threshold values support = 25% and confidence = 60%. Find all the
frequent itemsets and strong association rules using apriori algorithm.

ii) Write the drawbacks of apriori algorithm and explain how FP growth algorithm 4
is used to overcome this.

14. a) i) Write the algorithm for decision tree induction. Construct a decision tree and generate classification rules for the given data.   16

Eye	Hair	Height	Oriental (Class Label)
Black	Black	Short	Yes
Black	White	Tall	Yes
Black	White	Short	Yes
Black	Black	Tall	Yes
Brown	Black	Tall	Yes
Brown	White	Short	Yes
Blue	Gold	Tall	No
Blue	Gold	Short	No
Blue	White	Tall	No
Blue	Black	Short	No
Brown	Gold	Short	No

(OR)

b) i) What is clustering? Write K-means cluster analysis.   8
Given the samples X1 = {1,0}, X2 = 0,1}, X3 = {2,1} and X4 = {3,3}. Assume that the initial centriods are X1 and X3 and k = 2. Use K-means algorithm to show the final 2 clusters. The distance function is Euclidean distance.

ii) Explain the categories of major clustering techniques.   8

15. a) i) Explain about spatial data warehouse construction with an example.   8

ii) Discuss about the social impacts of data mining.   8

(OR)

b) i) Explain about web mining with suitable examples.   8

ii) Explain about any two applications in data mining.   8

****