

1059  
B.E. (Computer Science and Engineering)  
Sixth Semester  
Elective – I  
CS-605C: Data Mining and Analysis

Time allowed: 3 Hours

Max. Marks: 50

**NOTE:** Attempt five questions in all, including Question No. 1 (Section-A) which is compulsory and selecting two questions each from Section B-C.

x-x-x

Section -A

- Q 1(a) List the main data mining functionalities. (10)
- (b) What is an interestingness measure?
- (c) What is the difference between dependent and independent data marts?
- (d) Why dimensional tables are called 'wide' and fact tables are called 'deep'?
- (e) Why do we prefer surrogate keys over the primary keys of the database?
- (f) In Association rule mining, if some constraint do not belong to any category, What it is called?
- (g) Which clustering approach overcome limitation of k-means algorithm?
- (h) If classifier say it is false class, but it is wrong then what is this case called?
- (i) How distance between ordinal variables is calculated?
- (j) List any two applications of web usage mining.

Section -B

- Q2 (a) Explain the multi-tiered architecture of data warehouse. What do we do in data staging area? (6)
- (b) Consider a cube with dimensions {year, brand, state}. To execute a query "to find the name of the item which was sold most in full region during last two quarters", What data warehouse operations need to be performed. (4)
- Q3 (a) What are normalization techniques? Consider the following marks of student in range 0 to 50. (5)  
12, 25, 47, 34, 48, 06, 28, 30, 20, 10, 40, 45, 35. Change the new range to 20 to 60.
- (b) Differentiate between OLAP architecture implementations. Describe when ROLAP is preferred over MOLAP. (5)
- Q 4 (a) Describe a method to find the relevance of attributes? Why discretization of attribute is necessary? (5)
- (b) What are aggregated warehouses? Explain the need of multiway aggregates with an example. (5)

Section -C

- Q5 Explain the Apriori algorithm in detail. Describe its time complexity. What are the different modifications that we can make to reduce its complexity. (10)
- Q6 (a) Describe the Naïve Bayesian Classification algorithm. How do we avoid zero probability problem? (6)
- (b) What are boosting approaches? How can we improve the accuracy of classifier using these? (4)
- Q7 (a) What are multimedia databases? Describe the different mining approaches used in multimedia databases. (6)
- (b) Describe the hierarchical clustering algorithm in detail. (4)

x-x-x