



MAULANA ABUL KALAM AZAD UNIVERSITY OF TECHNOLOGY, WEST BENGAL

Paper Code : MCAN-E305B Data Warehousing and Data Mining

UPID : 003889

Time Allotted : 3 Hours

Full Marks : 70

The Figures in the margin indicate full marks.

Candidate are required to give their answers in their own words as far as practicable

Group-A (Very Short Answer Type Question)

1. Answer *any ten* of the following :

[1 x 10 = 10]

- (I) Define OLAP.
- (II) What is KDD in data mining?
- (III) The partition algorithm in association uses how many scans of the database to discover all frequent sets?
- (IV) Which frequent pattern mining technique mines without candidate generation?
- (V) Define pattern evaluation.
- (VI) Which can be considered as the final output of the hierarchal type of clustering?
- (VII) What are Categorical variables?
- (VIII) _____ is an object-oriented, integrated, time-variant, and nonvolatile collection of data in support of management decisions.
- (IX) Define data cube.
- (X) ETL stands for _____.
- (XI) What is Gini index?
- (XII) What does OLTP stand for?

Group-B (Short Answer Type Question)

Answer *any three* of the following :

[5 x 3 = 15]

2. Define Frequent sets, confidence, support and association rule. [5]
3. Explain whether Association rule mining is a supervised or unsupervised type of learning. Justify with a suitable example. [5]
4. Explain the differences between "Explorative Data Mining" and "Predictive Data Mining" and give one example of each. [5]
5. A data set for analysis includes only one attribute X:
 $X = \{7, 12, 5, 8, 5, 9, 13, 12, 19, 7, 12, 12, 13, 3, 4, 5, 13, 8, 7, 6\}$ [5]
 - (a) What is the mean of the data set X?
 - (b) What is the median?
 - (c) Find the standard deviation for X.
6. Can you briefly describe the four stages of knowledge discovery(KDD)? Can you describe the multi-tiered data warehouse architecture? [5]

Group-C (Long Answer Type Question)

Answer *any three* of the following :

[15 x 3 = 45]

7. (a) What is the difference between OLTP and OLAP? [3]
 (b) What is ODS? What is VLDB? [4]
 (c) Discuss ROLAP, MOLAP, HOLAP and DOLAP in data warehousing technology. [8]
8. (a) What is time series data? [5]
 (b) What do you mean by trend analysis? [5]
 (c) Explain similarity search in time series analysis. [5]
9. (a) Define Apriori algorithm with all the steps involved? [5]
 (b) Explain how the efficiency of Apriori can improved, with a suitable example. [5]
 (c) Explain frequent item set with candidate generation and without candidate generation? [5]
10. (a) What do you mean by anomaly detection? [2]
 (b) What do you mean by data cleaning? [3]

(c) What are the different kind of regression used in the datamining? [10]

11. (a) What is ETL? Explain different steps of ETL. [4]

(b) Assume, you want to cluster 7 observations into 3 clusters using K-Means clustering algorithm. [6]

After first iteration clusters, C1, C2, C3 has following observations:

C1: {(2,2), (4,4), (6,6)},

C2: {(0,4), (4,0)},

C3: {(5,5), (9,9)};

What will be the cluster centroids if you want to proceed for second iteration?

(c) What is the Applications of DBMiner? What is a factless table? [5]

*** END OF PAPER ***