Meerut Institute of Engineering & Technology, Meerut
First Sessional Examination, 2020-21

Course:	**M. Tech.**	**Branch:**	**CSE**
Subject:	**Data warehouse and Data Mining**	**Subject Code:**	**MTCS024**
M.M.	**60**	**Time:**	**2.00 hrs**

Section A

Q.1 Attempt any TWO from the following. **(2X 4 = 8)**

Q. No.	Question
a.	Define Data Mining. What are the various functionalities of data mining?
b.	Explain the concept of multidimensional data model?
c.	Describe the various components of data warehouse in detail?

Q.2 Attempt any TWO from the following. **(2X 4 = 8)**

Q. No.	Question
a.	What do you understand by concept hierarchy? Explain with an example.
b.	Discuss why data warehouse is maintained separated from database?
c.	Why we need to preprocess data? Explain various techniques to handle missing data

Section B

Q.3 Attempt any TWO from the following. **(2 X 6 = 12)**

Q. No.	Question
a.	Define OLAP and also discuss the various OLAP operation of multidimensional data?
b.	A travel company wishes to determine if the type of vacation purchased in its market area is independent of income level of purchasers. A random survey of purchasers gave the following results: Vacation Type / Income Level: High: Domestic 50, Foreign 25; Medium: Domestic 120, Foreign 30; Low: Domestic 65, Foreign 10 At the 0.05 level of significance, can it be concluded that vacation preference and income level are statistically independent? Critical value : 5.991 at 0.05 level of significance 9.210 at 0.01 level of significance
c.	Define the term Knowledge Discovery and Databases (KDD). Explain the phases of KDD process with diagram?

Q.4 Attempt any TWO from the following. **(2 X 6 = 12)**

Q. No.	Question
a.	Diagrammatically illustrate and discuss the architecture of ROLAP, MOLAP AND HOLAP.
b.	A database has five transactions. Let *min sup* = 3 and *min con f* = 60%. Find the association rule and frequent itemset. *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}
c.	"A data warehouse can be modeled by either a star schema or a snowflake schema" With relevant examples discuss the two types of schema?

Section C

Q.5 Attempt any ONE from the following. (1 X 10 = 10)

Q. No.	Question
a.	Why to preprocess data? What are the major tasks to preprocess the data?
b.	Explain the concept of dimensionality reduction using PCA.

Q.6 Attempt any ONE from the following. (1 X 10 = 10)

Q. No.	Question
a.	Enumerate the steps involved in mapping the data warehouse to a multiprocessor architecture?
b.	Design three tier architecture of data warehouse. Also describe the three tier architecture in detail?