

GUJARAT TECHNOLOGICAL UNIVERSITY
MCA INTEGRATED– SEMESTER VI - EXAMINATION –WINTER-2022

Subject Code: 2668603**Date: 15/12/2022****Subject Name: Data Mining****Time: 10:30 AM TO 01:00 PM****Total Marks: 70****Instructions:**

1. Attempt all questions.
2. Make Suitable assumptions wherever necessary.
3. Figures to the right indicate full marks.

- Q.1** (a) What is a Data Warehouse? Explain Differences between operational database systems and Data Warehouses. **07**
- (b) Discuss following Data Warehousing Modeling: **07**
1. Stars.
 2. Snowflakes.

- Q.2** (a) What is Data Mining; Why Data Mining? What kind of data can be mined? What Kind of Patterns Can be Mined? **07**
- (b) Explain Knowledge Discovery (KDD) Process. **07**

OR

- (b) What is Data Pre-processing? Discuss Data Cleaning and Data Integration techniques. **07**

- Q.3** (a) What Is Frequent Pattern Analysis? Basic Concepts of Frequent Patterns and Association Rules **07**
- (b) Prepare following Apriori Algorithm **07**

Database TDB Supmin = 2

Tid	Items
10	A, C, D
20	B, C, E
30	A, B, C, E
40	B, E

OR

- Q.3** (a) Explain Scalable Methods for Mining Frequent Patterns. **07**
- (b) Discuss Challenges of Frequent Pattern Mining. **07**

- Q.4** (a) What is Classification & Prediction? Discuss Classification A Two-Step Process. **07**
- (b) Discuss Decision Tree Induction. Also discuss Attribute Selection: Information Gain using Example. **07**

OR

- Q.4** (a) Why Bayesian Classification? Discuss Naïve Bayesian Classifier using Training dataset. **07**
- (b) Discuss following Techniques: **07**
1. Bagging.
 2. Boosting.

- Q.5** (a) What is Cluster Analysis? Discuss Requirements of Clustering in Data Mining. **07**
- (b) Discuss K-Means Clustering Method. **07**

OR

- Q.5** (a) What are outliers? Types of outliers, Challenges of outlier Detection. **07**
- (b) Short note on: **BIRCH** **07**
