

**GUJARAT TECHNOLOGICAL UNIVERSITY**  
**MCA INTEGRATED– SEMESTER VI- EXAMINATION –SUMMER-2024**

**Subject Code: 2668603**

**Date: 16/05/2024**

**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.
4. Use of simple calculators and non-programmable scientific calculators are permitted.

- Q.1** (a) Describe the steps involved in data mining when viewed as a process of knowledge discovery. **07**
- (b) What is a Data Warehouse? Give Differences between operational database systems and Data Warehouses. **07**
- Q.2** (a) What is Data Pre-processing? Discuss Data Cleaning and Data Integration techniques. **07**
- (b) Draw and explain the Data warehouse architecture. **07**
- OR**
- (b) Explain Stars, Snowflakes, and Fact Constellations Schemas. **07**
- Q.3** (a) Explain Market Basket Analysis in detail. **07**
- (b) Explain Information Gain, Gain Ratio, and Gini Index **07**
- OR**
- Q.3** (a) What Is Frequent Pattern Analysis? Basic Concepts of Frequent Patterns and Association Rules **07**
- (b) Explain Classification and Regression for Predictive Analysis. **07**
- Q.4** (a) Explain Apriori Algorithm with an example. **07**
- (b) Discuss the following Techniques: 1. Bagging. 2. Boosting. **07**
- OR**
- Q.4** (a) Explain Tree pruning? Explain scalability and decision tree induction. **07**
- (b) Why Bayesian Classification? Discuss Naïve Bayesian Classifier using Training dataset. **07**
- Q.5** (a) What are outliers? Types of outliers, Challenges of outlier Detection. **07**
- (b) What is Cluster Analysis? Explain Requirements for Cluster Analysis. **07**
- OR**
- Q.5** (a) Write a short note on BIRCH **07**
- (b) Enlist the steps of K-Mean clustering algorithm. Explain it with a suitable example. **07**

\*\*\*\*\*