# GUJARAT TECHNOLOGICAL UNIVERSITY
## MCA INTEGRATED – SEMESTER VI- EXAMINATION –WINTER-2024

**Subject Code: 2668603**                                      **Date: 25/11/2024**
**Subject Name: Data Mining**
**Time:02:30 PM TO 05:00 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)** | Define following terms :<br>1. Data warehouse<br>2. Starnet Query model<br>3. BIRCH | **07** |
| | **(b)** | Differences between operational database systems and Data Warehouses. | **07** |
| | | | |
| **Q.2** | **(a)** | Explain the basic concepts of classification. | **07** |
| | **(b)** | What are some typical OLAP (Online Analytical Processing) operations performed in a data warehouse? | **07** |
| | | **OR** | |
| | **(b)** | What is the process of extraction, transformation, and loading (ETL) in a data warehouse? | **07** |
| | | | |
| **Q.3** | **(a)** | What is data mining, and why is it important in data analysis? What types of data can be mined in data mining processes? | **07** |
| | **(b)** | Describe decision tree and How does tree pruning contribute to improving decision tree models? | **07** |
| | | **OR** | |
| **Q.3** | **(a)** | What is decision tree induction and What are attribute selection measures, and why are they important in decision tree induction? | **07** |
| | **(b)** | Short Note on : Rule-based Classification. | **07** |
| | | | |
| **Q.4** | **(a)** | Describe the BIRCH (Balanced Iterative Reducing and Clustering using Hierarchies) multiphase hierarchical clustering approach with clustering feature trees. | **07** |
| | **(b)** | Explain density-based methods such as DBSCAN (Density-Based Spatial Clustering of Applications with Noise) and OPTICS (Ordering Points To Identify the Clustering Structure) in outlier analysis. | **07** |
| | | **OR** | |
| **Q.4** | **(a)** | How does the Apriori algorithm work in mining frequent itemsets? | **07** |
| | **(b)** | Define market-basket analysis and explain its significance in data mining. Discuss the basic concepts of frequent itemsets and closed itemsets in the context of market-basket analysis. | **07** |
| | | | |
| **Q.5** | **(a)** | Explain Bayes' Theorem and its application in Naïve Bayesian classification. | **07** |
| | **(b)** | Describe the bootstrap technique in model evaluation. | **07** |
| | | **OR** | |
| **Q.5** | **(a)** | What is decision tree induction, and how is it used in classification? | **07** |
| | **(b)** | What is cluster analysis, and what are its requirements? | **07** |

*************