Based on the **GTU Paper Format** and standard examination trends for Diploma Engineering, Unit 3: **Data Mining Techniques** is the most critical part of the syllabus.

In a 70-mark paper, this unit covers **CO3** and typically accounts for **21–25 marks**. It is unique because it features **Question 2(c)** and **Question 3(c)**, which are often 7-mark **Application (A)** level questions involving numerical problems or algorithmic steps.

Unit 3: Predicted Question Bank (Data Mining Techniques)

1. Most Repeated / High-Probability Questions

These follow the (a) 3-mark, (b) 4-mark, and (c) 7-mark structure of your model paper.

[Short Answer Type - 03 Marks]

- **Define Support and Confidence** in Association Rule Mining.
- **What is a "Frequent Itemset"?** Give an example.
- **Define Classification.** How does it differ from Clustering?
- **What is Entropy?** Explain its role in building a Decision Tree.

[Descriptive Type - 04 Marks]

- **Explain the Apriori Property** (Downward Closure Property) with an example.
- **Compare Supervised vs. Unsupervised Learning.** List two algorithms for each.
- **Explain the concept of 'Pruning'** in Decision Trees. Why is it necessary?
- **Describe the K-Nearest Neighbor (KNN) algorithm** steps in brief.

[Long Answer / Diagram-Based - 07 Marks]

- **Explain the Apriori Algorithm** for finding frequent itemsets. Show the join and prune steps clearly.
- **Describe Decision Tree Induction** using the ID3 algorithm. Explain how **Information Gain** is used to select the root node.
- **Explain Bayesian Classification.** State and explain Bayes' Theorem with the formula for posterior probability.
- **Discuss K-Means Clustering.** Explain the algorithm with a neat diagram showing how centroids move.

---

2. Application & Logical Thinking Questions

*These questions target the **Cognitive Level: Apply (A)** mentioned in your paper format,*

separating average students from top scorers.

11. **Numerical Application (Apriori):** Given a database of 5 transactions (T1:{A,B,C}, T2:{A,C}, T3:{B,D}, T4:{A,B,C,D}, T5:{A,C}), find all frequent itemsets if the **Minimum Support is 2**. Show the L_1, C_2, L_2 tables.
12. **Case Study Reasoning:** A bank wants to group its customers into "Gold," "Silver," and "Bronze" categories based on their spending habits. Which technique should they use: **Classification or Clustering**? Justify your choice.
13. **Problem Interpretation:** In a Decision Tree, if an attribute has an Information Gain of **0**, what does it indicate about that attribute's ability to split the data? Should it be chosen as a splitting node?
14. **System-Level Design:** An e-commerce site wants to implement a "Frequently Bought Together" feature. Explain how **Association Rule Mining** can be used to generate these recommendations from thousands of daily transactions.
15. **Algorithm Selection:** You have a dataset with **labeled** historical data of "Email Spam" vs "Not Spam." Would you use **Naïve Bayes** or **K-Means** to build a filter? Explain why based on the type of learning (Supervised/Unsupervised).

Exam Success Strategy for Unit 3:

- **Master the Numericals:** In the GTU format, **Q.2(c)** and **Q.3(c)** are often 7-mark numericals on Apriori or Bayes. If you solve these correctly, you get a perfect score.
- **Comparison Tables:** Memorize the difference between **Classification vs. Clustering** and **Eager vs. Lazy Learners**.
- **Step-by-Step Algorithms:** When writing an algorithm (like K-Means), use a numbered list ($Step 1, Step 2 \dots$) and a small diagram. This makes your paper easy to grade.