# GUJARAT TECHNOLOGICAL UNIVERSITY

## BE 6 SEMESTER – OLD PAPER – S22 TO S26 – QUESTION BANK

**Subject Name & Code:** Data Mining - 3160714

---

## Chapter 1: Introduction to data mining (DM)

### Repeated Questions:

**1. Knowledge Discovery Process:**

- **Q.** Describe the steps involved in data mining when viewed as a process of knowledge discovery / Explain KDD process. [Diagram required – refer to the original question paper] **(SUMMER 2023, 7 Marks; WINTER 2023, 7 Marks; SUMMER 2024, 4 Marks; WINTER 2025, 7 Marks; SUMMER 2026, 7 Marks)**

**2. Data Mining Issues and Functionalities:**

- **Q.** Explain the major issues in data mining / Explain major requirements and challenges in data mining. **(WINTER 2022, 7 Marks; WINTER 2023, 7 Marks; SUMMER 2024, 7 Marks; WINTER 2025, 4 Marks; SUMMER 2026, 7 Marks)**
- **Q.** Explain the data mining functionalities / List out the data mining functionality. **(WINTER 2022, 4 Marks; WINTER 2025, 3 Marks)**
- **Q.** Compare descriptive and predictive data mining. **(SUMMER 2022, 4 Marks; WINTER 2022, 3 Marks)**

**3. Data Warehouse and OLAP:**

- **Q.** Differentiate between OLAP and OLTP. **(WINTER 2023, 4 Marks; SUMMER 2026, 4 Marks)**

### Other Important Questions:

**1. Data Warehouse:**

- **Q.** Define Schema. Explain the following schemas with suitable example. 1) Star 2) Snowflakes 3) Constellations **(SUMMER 2026, 3 Marks)**
- **Q.** A data warehouse is a subject-oriented, integrated, time-variant, and nonvolatile collection of data - Justify. **(SUMMER 2026, 4 Marks)**
- **Q.** Define Data Warehouse. State its features. **(WINTER 2023, 3 Marks)**
- **Q.** Define Data Cube. Explain any two operations on it. **(WINTER 2023, 3 Marks OR)**

**2. Data Mining Basics:**

- **Q.** Explain data mining application for fraud detection. **(SUMMER 2026, 3 Marks OR)**
- **Q.** Describe any three primitives for specifying a data mining task. **(SUMMER 2023, 3 Marks)**
- **Q.** Define data mining. Describe three challenges to data mining regarding data mining methodology and user interaction issues. **(SUMMER 2024, 3 Marks)**
- **Q.** What are the types of data? **(SUMMER 2022, 3 Marks)**
- **Q.** Draw and explain the data mining architecture. [Diagram required – refer to the original question paper] **(SUMMER 2022, 7 Marks)**

# Chapter 2: Data Pre-processing

## Repeated Questions:

**1. Data Cleaning:**

- **Q.** In real-world data, tuples with missing values for some attributes are a common occurrence. Describe various methods for handling this problem / What is missing values and noise? How to handle them? **(SUMMER 2023, 4 Marks; SUMMER 2024, 7 Marks; WINTER 2025, 3 Marks)**
- **Q.** Explain the smoothing techniques available to remove noise / Explain Binning method with the help of example. **(WINTER 2022, 4 Marks; WINTER 2023, 4 Marks; SUMMER 2024, 3 Marks; SUMMER 2026, 7 Marks OR)**
- **Q.** What is Data Cleaning? Describe various methods of Data Cleaning. **(SUMMER 2022, 7 Marks; WINTER 2022, 7 Marks)**

**2. Concept Hierarchy:**

- **Q.** What is concept hierarchy? List and explain types of Concept Hierarchy. **(SUMMER 2022, 4 Marks; WINTER 2022, 3 Marks; SUMMER 2026, 3 Marks)**

## Other Important Questions:

**1. Noise and Data Transformation:**

- **Q.** What is noise? Describe the possible reasons for noisy data. Explain the different techniques to remove the noise from data. **(SUMMER 2023, 7 Marks)**
- **Q.** Discuss normalization in detail. **(SUMMER 2024, 4 Marks)**
- **Q.** Define the following terms in Data Transformation: i. Smoothing ii. Normalization iii. Discretization **(WINTER 2023, 3 Marks)**

**2. Data Pre-processing Overview, Reduction and Integration:**

- **Q.** Why to preprocess the data in data Mining? **(WINTER 2023, 3 Marks)**
- **Q.** Explain data pre-processing functionalities. **(WINTER 2025, 7 Marks OR)**
- **Q.** Discuss dimensionality reduction techniques. **(WINTER 2025, 4 Marks)**
- **Q.** What is dimensionality reduction? **(SUMMER 2022, 3 Marks)**
- **Q.** Explain about the different Data Reduction techniques. **(WINTER 2022, 7 Marks OR)**
- **Q.** Discuss issues to be considered during data integration. **(SUMMER 2022, 7 Marks OR)**

**3. Normalization (Numerical):**

- **Q.** Explain in detail - min-max normalization method. Use this method to normalize the following group of data by setting min = 0 and max = 1. 200, 400, 600, 1000 **(SUMMER 2023, 7 Marks OR)**

# Chapter 3: Concept Description, Mining Frequent Patterns, Associations and Correlations

## Repeated Questions:

**1. Apriori Algorithm and Frequent Itemsets:**

- **Q.** Explain frequent item-set mining algorithm with suitable example / Explain the steps of the "Apriori Algorithm" for mining frequent itemsets. **(WINTER 2022, 7 Marks OR; SUMMER 2024, 4 Marks OR; WINTER 2025, 7 Marks; SUMMER 2026, 7 Marks)**
- **Q.** What are the techniques to improve the efficiency of Apriori algorithm? **(WINTER 2022, 3 Marks; SUMMER 2023, 4 Marks; SUMMER 2024, 4 Marks; SUMMER 2026, 7 Marks OR)**
- **Q.** Write short notes on Maximal Frequent Item Set & Closed Frequent Item Set. **(SUMMER 2022, 7 Marks OR; WINTER 2025, 3 Marks OR)**

**2. Association Rules:**

- **Q.** Explain measures for finding rule interestingness (support, confidence) with example / Explain following terms related to Association Rule Mining: Itemset, Support Count, support, and Association rule. **(WINTER 2023, 7 Marks; WINTER 2025, 4 Marks; SUMMER 2026, 4 Marks)**
- **Q.** How are association rules mined from large databases? **(SUMMER 2022, 4 Marks; SUMMER 2024, 7 Marks)**
- **Q.** Give a short example to show that items in a strong association rule are not always interesting. **(SUMMER 2022, 3 Marks OR; SUMMER 2023, 3 Marks; SUMMER 2024, 7 Marks OR)**

**3. Association Rules (Numerical):**

- **Q.** Find the frequent itemsets and generate association rules. Assume that minimum support threshold (s = 33.33%) and minimum confidence threshold (c = 60%). [Diagram required – refer to the original question paper] **(WINTER 2022, 7 Marks; SUMMER 2023, 7 Marks OR; WINTER 2023, 7 Marks OR)**

## Other Important Questions:

**1. Association Rule Concepts:**

- **Q.** What is market basket analysis? Precisely explain the meaning of the following association rule: computer → antivirus_software [support = 60%, confidence = 60%] **(SUMMER 2023, 3 Marks)**
- **Q.** What is an Itemset? What is a Frequent Itemset? **(WINTER 2022, 4 Marks)**
- **Q.** What is meant by association rule? **(SUMMER 2022, 3 Marks)**
- **Q.** Give the difference between Boolean association rule and quantitative association rule. **(SUMMER 2024, 3 Marks)**
- **Q.** Describe the different classifications of Association rule mining. **(WINTER 2022, 3 Marks OR)**
- **Q.** Explain the various criteria for the classification of frequent pattern mining. **(SUMMER 2022, 7 Marks)**
- **Q.** What is meant by multidimensional association rules? **(SUMMER 2022, 4 Marks OR)**
- **Q.** How Meta rules are useful in constraint based association mining. **(SUMMER 2024, 3 Marks OR)**
- **Q.** What is meant by Reduced Minimum Support? **(WINTER 2022, 4 Marks OR)**
- **Q.** Write a note on incremental Association Rule Mining. **(SUMMER 2026, 4 Marks OR)**
- **Q.** Compare association and classification. Briefly explain associative classification with suitable example. **(SUMMER 2026, 7 Marks)**

**2. FP-Growth and Support/Confidence (Numerical):**

- **Q.** Discuss how frequent itemsets can be generated using FP-Growth algorithm with the help of the following transactions. Let minimum support threshold is 2. [Diagram required – refer to the original question paper] **(SUMMER 2023, 7 Marks)**
- **Q.** For given example find support & confidence for {Milk, Chocolate} ⇒ Pepsi, {Milk, Pepsi} → {Chocolate}, {Chocolate, Pepsi} → {Milk}. [Diagram required – refer to the original question paper] **(WINTER 2023, 7 Marks OR)**

## Chapter 4: Classification and Prediction

### Repeated Questions:

**1. Classification Basics:**

- **Q.** Differentiate between Classification and Prediction / Explain classification Vs. Prediction. **(SUMMER 2022, 3 Marks OR; WINTER 2023, 4 Marks; WINTER 2025, 3 Marks OR)**
- **Q.** Differentiate between supervised learning and unsupervised learning. **(SUMMER 2022, 4 Marks OR; WINTER 2022, 4 Marks OR; WINTER 2023, 3 Marks OR; SUMMER 2024, 3 Marks)**

**2. Decision Tree:**

- **Q.** Explain Decision Tree Classification algorithm with the help of example / Explain decision tree algorithm with suitable example. **(WINTER 2022, 7 Marks OR; SUMMER 2023, 4 Marks OR; WINTER 2023, 7 Marks; SUMMER 2024, 7 Marks; WINTER 2025, 7 Marks)**

**3. Regression:**

- **Q.** Differentiate linear and logistic regression / Briefly explain Linear and Non-linear regression. **(WINTER 2023, 4 Marks OR; SUMMER 2024, 3 Marks OR; WINTER 2025, 4 Marks; SUMMER 2026, 3 Marks)**

**4. Neural Network:**

- **Q.** Draw the topology of a multilayer, feed-forward Neural Network / Explain ANN. [Diagram required – refer to the original question paper] **(WINTER 2023, 4 Marks; WINTER 2025, 7 Marks OR; SUMMER 2026, 4 Marks)**

**5. Accuracy Evaluation:**

- **Q.** Explain various methods of evaluating accuracy of classifier / Discuss cross-validation method. **(SUMMER 2023, 3 Marks OR; SUMMER 2024, 4 Marks OR; WINTER 2025, 3 Marks; WINTER 2025, 4 Marks OR)**

### Other Important Questions:

**1. Classification Theory:**

- **Q.** What is classification? Explain classification as a two step process with diagram. [Diagram required – refer to the original question paper] **(SUMMER 2026, 4 Marks OR)**
- **Q.** Explain the issues regarding the classification and prediction. **(SUMMER 2022, 7 Marks OR)**
- **Q.** Explain proximity measures used in distance based classifiers. **(WINTER 2025, 4 Marks OR)**
- **Q.** Write down short note on backpropagation. **(SUMMER 2024, 4 Marks)**
- **Q.** What are the hierarchical methods used in classification? **(WINTER 2022, 4 Marks)**
- **Q.** Describe in detail about Rule based Classification. **(WINTER 2022, 7 Marks)**

**2. Decision Tree Concepts:**

- **Q.** What is attribute selection measure? **(WINTER 2022, 3 Marks OR)**
- **Q.** What do you mean by Tree Pruning? Explain with example. **(WINTER 2023, 3 Marks)**
- **Q.** Explain the following as attribute selection measure: (i) Information Gain (ii) Gain Ratio **(WINTER 2023, 4 Marks)**

**3. Bayesian Classification:**

- **Q.** What are Bayesian Classifiers? **(WINTER 2022, 3 Marks)**
- **Q.** What is Bayesian theorem? **(SUMMER 2022, 4 Marks)**
- **Q.** Demonstrate how Bayesian classification helps in predicting class membership probabilities. **(SUMMER 2022, 7 Marks)**
- **Q.** Explain Naïve Bayes Classifier with example. **(WINTER 2023, 7 Marks OR)**

**4. Classification Problems (Numerical):**

- **Q.** The following table shows the midterm and final exam grades obtained by students in a database course. Use the method of least squares to find an equation for the prediction of a student's final exam grade based on the student's midterm grade in the course. Predict the final exam grade of a student who received 86 grade in the midterm exam. [Diagram required – refer to the original question paper] **(SUMMER 2023, 4 Marks)**
- **Q.** Apply ID3 classification algorithm on the following data and construct a decision tree. Show all the stepwise calculations clearly. [Diagram required – refer to the original question paper] **(SUMMER 2023, 7 Marks)**
- **Q.** Predict class label of the tuple X = (age = youth, income = medium, student = yes, credit_rating = fair) with the help of Naïve Bayesian classification method and the following data. [Diagram required – refer to the original question paper] **(SUMMER 2023, 7 Marks OR)**

## Chapter 5: Cluster Analysis

### Repeated Questions:

**1. K-Means and K-Medoids:**

- **Q.** Explain K-means algorithm / Discuss the K-means clustering algorithm using examples. **(SUMMER 2022, 7 Marks; WINTER 2023, 7 Marks; WINTER 2025, 7 Marks)**
- **Q.** How K-Mean clustering method differs from K-Medoid clustering method? **(WINTER 2022, 4 Marks OR; SUMMER 2023, 4 Marks OR; WINTER 2023, 3 Marks; SUMMER 2024, 4 Marks OR; SUMMER 2026, 4 Marks)**

**2. Clustering Basics:**

- **Q.** Differentiate classification and clustering. **(SUMMER 2023, 3 Marks; WINTER 2025, 4 Marks)**
- **Q.** Explain the requirements of Clustering in Data Mining. **(SUMMER 2022, 7 Marks OR; WINTER 2022, 3 Marks; SUMMER 2023, 4 Marks OR; WINTER 2023, 7 Marks OR; SUMMER 2024, 7 Marks OR)**

**3. Hierarchical Clustering:**

- **Q.** Differentiate between partitioning and hierarchical methods for clustering. **(WINTER 2023, 4 Marks OR; SUMMER 2024, 4 Marks)**
- **Q.** Explain hierarchical and agglomerative method of clustering. **(SUMMER 2023, 3 Marks OR; WINTER 2025, 7 Marks)**

### Other Important Questions:

**1. Clustering Concepts:**

- **Q.** Define “clustering”? Mention any two applications of clustering. **(SUMMER 2026, 3 Marks)**
- **Q.** How to evaluate clustering algorithm? **(WINTER 2025, 3 Marks OR)**
- **Q.** Discuss data matrix and dissimilarity matrix with respect to clustering. **(SUMMER 2023, 4 Marks)**
- **Q.** What is meant by hierarchical clustering? **(WINTER 2022, 3 Marks OR)**
- **Q.** Differentiate Agglomerative and Divisive Hierarchical Clustering? **(WINTER 2022, 4 Marks)**

**2. Clustering Algorithms:**

- **Q.** Explain PAM algorithm. **(WINTER 2025, 7 Marks OR)**
- **Q.** Explain following clustering algorithm in details: 1) CLARA 2) BIRCH **(SUMMER 2024, 7 Marks)**

**3. Outlier Analysis:**

- **Q.** What is an outlier? **(SUMMER 2022, 3 Marks)**
- **Q.** Briefly explain outlier analysis. **(WINTER 2025, 3 Marks)**
- **Q.** Discuss outlier analysis as a data mining functionality with the help of an example. **(SUMMER 2023, 3 Marks OR)**

**4. K-Means (Numerical):**

- **Q.** Suppose that the data mining task is to cluster the following eight points (with (x, y) representing location) into three clusters: A1(2, 10), A2(2, 5), A3(8, 4), B1(5, 8), B2(7, 5), B3(6, 4), C1(1, 2), C2(4, 9). The distance function is Euclidean distance. Suppose initially we assign A1, B1, and C1 as the center of each cluster, respectively. With the help of k-means algorithm calculate, (i) The three cluster centers after the first round execution (ii) The final three clusters **(SUMMER 2023, 7 Marks)**

## Chapter 6: Web mining and other data mining

### Repeated Questions:

**1. Web Mining:**

- **Q.** What is web log? Explain web structure mining and web usage mining in detail. **(WINTER 2023, 7 Marks OR; SUMMER 2026, 7 Marks OR)**
- **Q.** Explain web usage mining. **(SUMMER 2022, 4 Marks; WINTER 2022, 7 Marks OR; SUMMER 2023, 7 Marks OR)**
- **Q.** Discuss web structure mining. **(SUMMER 2023, 3 Marks; SUMMER 2026, 7 Marks OR)**
- **Q.** Write a short note: Web content mining. **(SUMMER 2022, 4 Marks OR; WINTER 2022, 7 Marks; SUMMER 2024, 7 Marks OR)**

**2. Temporal and Multimedia Mining:**

- **Q.** Explain temporal mining. **(SUMMER 2022, 3 Marks; SUMMER 2024, 3 Marks)**
- **Q.** Discuss multimedia mining. **(SUMMER 2022, 3 Marks OR; SUMMER 2023, 4 Marks)**

**3. Distributed Data Mining and Hadoop:**

- **Q.** Discuss the Applications of Distributed and parallel Data Mining. **(SUMMER 2024, 3 Marks OR; SUMMER 2026, 4 Marks OR)**
- **Q.** Discuss the main features of Hadoop Distributed File System / Why Hadoop is important? **(WINTER 2023, 4 Marks OR; SUMMER 2026, 7 Marks OR)**

### Other Important Questions:

**1. Spatial and Web Mining:**

- **Q.** What is spatial data mining? Explain with its application. **(WINTER 2025, 3 Marks)**
- **Q.** Discuss spatial Data Mining, Temporal Mining, and Multimedia Mining with their application. **(WINTER 2025, 7 Marks OR)**
- **Q.** Discuss application of web mining. **(WINTER 2025, 4 Marks OR)**
- **Q.** Explain web mining using example. **(SUMMER 2026, 3 Marks OR)**
- **Q.** Give difference between text mining and web mining. **(WINTER 2023, 3 Marks OR)**

**2. Hadoop and MapReduce:**

- **Q.** Explain Hadoop storage - HDFS. **(SUMMER 2026, 3 Marks)**
- **Q.** Explain MapReduce. Explain any example using MapReduce. **(SUMMER 2026, 7 Marks)**

**********