**Maharshi Karve Stree Shikshan Samstha's**

**Smt. Hiraben Nanavati Institute of Management & Research for Women**

**INTERNAL EXAMINATION – Sem IV – ( Jan- July 2026)**

**SUBJECT – DATA MINING** **SUBJECT CODE – 101**

**Date : 08-05-2026** **Pattern : 2024**

**Duration : 150 min** **Max Marks : 50**

**Instructions for students :**

- *Marks are indicated for each question.*
- *Handwriting should be eligible for evaluation.*
- *Marks will be given for quality, not quantity.*

---

**Q.1]** Solve any five . ( 10 Marks)

a) Define Big Data and mention any two needs of Big Data Analytics.
b) What is Data Mining? State any two real-world applications of Data Mining.
c) Differentiate between structured and unstructured data.
d) Define classification and clustering with one example each.
e) What is data normalization? Name any two normalization techniques.
f) Define support and confidence in association rule mining.
g) What is anomaly detection? Mention one practical application.
h) Define precision and recall in classification models.

**Q.2]** Attempt any Two – 5 Marks Each

A] A retail company wants to analyze customer purchase behavior using Business Intelligence and Data Mining techniques.
Explain:

- Scope of Business Intelligence (BI)
- Relationship between BI and Data Mining
- Steps involved in the Data Mining process

Illustrate your answer with a suitable example.

B] An online education platform stores the following student information:

- Student Name
- Course Enrolled
- Course Rating
- Number of Videos Watched

- Completion Status

Identify the most suitable attribute/data type for each field and justify your answer based on the nature of data.

C] A telecom company wants to detect fraudulent SIM card usage using Data Mining techniques.

Explain:

- Data preprocessing
- Pattern discovery
- Evaluation phase
- Role of anomaly detection in fraud prevention

**Q.3]** Attempt any One – 10 Marks

A] Describe various classification techniques used in Data Mining. Explain any two of the following in detail with suitable examples:

- Decision Tree Classification
- Rule-Based Classification
- k-Nearest Neighbor (k-NN)
- Support Vector Machine (SVM)

Also explain the importance of:

- Accuracy
- Precision
- Recall
- Confusion Matrix

B] A digital marketing company wants to analyze customer purchase records for predicting buying behavior.

The dataset contains:

- Customer_ID
- Age
- Monthly_Spending
- Purchase_Frequency
- Membership_Type
- Purchase_Status

Using suitable classification techniques:

i) Explain the preprocessing steps required before building the model.

ii) Describe how training and testing datasets are created.

iii) Explain how Random Forest and Boosting techniques can improve prediction performance.

iv) Explain cross-validation and its importance.

v) Suggest the most appropriate evaluation metrics for this problem.

**Q.4]** Attempt any One – 10 Marks

A] Describe the various clustering techniques used in Data Mining.
Explain the following methods with suitable examples:

- k-Means Clustering
- Hierarchical Clustering
- Density-Based Clustering (DBSCAN)
- Graph-Based Clustering

Also explain:

- Internal cluster validation
- External cluster validation
- Applications in customer profiling and market segmentation

B] A supermarket chain wants to segment customers based on annual spending and purchase frequency.

The management wants to:

- Identify premium customers
- Detect irregular buying patterns
- Design targeted marketing campaigns

Answer the following:

i) Explain how clustering can help in customer segmentation.
ii) Differentiate between agglomerative and divisive hierarchical clustering.
iii) Explain the working of k-means clustering with suitable steps.
iv) Discuss density-based clustering methods and their advantages.
v) Explain the practical applications of clustering in business analytics.

**Q.5]** Attempt any One – 10 Marks *(10 Marks)*

A] An e-commerce company wants to analyze customer buying patterns using association analysis.
The following transaction data is available:

| Transaction ID | Items Purchased |
|---|---|
| T1 | Bread, Milk, Butter |
| T2 | Bread, Diapers, Milk |
| T3 | Milk, Diapers, Butter |
| T4 | Bread, Milk, Diapers, Butter |
| T5 | Bread, Butter |

Answer the following:

i) Explain the Apriori Algorithm and its working.
ii) Calculate:

- Support
- Confidence
- Lift

for the association rule:
**Bread → Milk**
iii) Explain sequential pattern mining.
iv) Discuss applications of association analysis in:

- Medical Informatics
- Telecommunication Alarm Diagnosis
- Customer Buying Path Analysis

v) Explain ethical concerns related to Data Mining and Big Data Analytics.

B] As a Data Analyst in a multinational organization, explain how Big Data Analytics and Data Mining techniques can be applied in the following domains:

- Banking and Fraud Detection
- Healthcare Analytics
- E-commerce Recommendation Systems
- Mobile Environment Analytics
- Customer Relationship Management (CRM)

Also explain:

- Privacy concerns in Data Mining
- Data ownership and responsible usage
- Corporate responsibilities in handling Big Data systems.

***