

Plot No. 2, Knowledge Park-III, Greater Noida (U.P.) –201306  
**POST GRADUATE DIPLOMA IN MANAGEMENT (2024-25)**  
**END TERM EXAMINATION (TERM -V)**

Subject Name: **Data mining for Decision making**  
Sub. Code: **PGIT51**

Time: **90 mins**  
Max Marks: **40**

**INSTRUCTIONS**

1. All questions are to be solved using Organge on individual Computers/LAPTOPS.
  - Start with New workflow : Enter Subject Name, Date, Course & Session (PGDM 2023-25), SEMESTER, Student's Name, ROLL NUMBER on top eg. # Subject Name, # Date and so on
  - Export all processes as .ows files in one folder. Folder must be named with your full name and Roll Number for example (Name\_GM-----) . submit the soft copies to the invigilator using a PD.
  - Conceptual questions must be answered by adding Text Annotation , from bottom left, within the workflow in .ows file.
2. During examination, no student is allowed to use mobile phones/Smart watch/Internet in any condition.
3. Data sheets (.csv file/Excel) will be provided as a soft copy on the Desktops/Laptops

**Attempt all Four Questions.**

**40 marks**

CO-1	Identify the business needs for knowledge discovery in order to create competitive advantages with data mining technologies appropriately in order to realize their real business value in solving business problems. (L3)
CO-2	Utilize the basic data mining concepts and their application in business context using data mining tools.(L3)
CO-3	Evaluate interesting and useful patterns from the explosive Volume of data by application of supervised and unsupervised techniques. (L3, L4, L5)
CO-4	Examine Integration of theory and application in various functional areas through interdisciplinary approach. (L4)

**Question 1**

**10 marks**

**Case scenario**

Given **Dataset New Market Basket Analysis** A supermarket wants to analyze purchase patterns to suggest complementary products. Mine the given data to

1. Find the product associations to improve recommendations using Apriori Algorithm.
2. Do a detailed analysis of the statistical values and identified patterns.

**Dataset Features:**

- TransactionID (Unique Identifier)
- Product1 (Item Name)
- Product2 (Item Name)
- Product3 (Item Name)
- Product4 (Item Name)
- Product5 (Item Name)

## Question 2

10 marks

### Case Scenario

Given Dataset **New Customer Churn** A telecom company wants to identify customers who are likely to leave their service so they can offer personalized retention plans. Mine the given data using the appropriate Classification algorithm to

1. predict if a customer will churn based on their interaction history.
2. do a detailed analysis of the statistical values and suggested retention plans.

### Dataset Features:

- CustomerID (Unique Identifier)
- Age (Numeric)
- Subscription Length (Months)
- Monthly Spending (\$)
- Customer Support Calls (Count)
- Contract Type (Monthly, Yearly)
- Churn (Yes/No) (Target Variable)

## Question 3

10 marks

### Case Scenario

Given **Dataset New House Prices**, A real estate agency wants to build a model to estimate house prices based on historical sales data. Mine the given data using the Regression/Random Forest algorithm to

3. predict the price of a house based on its features
4. do a detailed analysis of the statistical values.

### Dataset Features:

House ID (Unique Identifier)

Square Feet (Numeric)

Number of Bedrooms (Numeric)

Number of Bathrooms (Numeric)

Location (Categorical)

Age of House (Years)

Price (\$) (Target Variable)

## Question 4

10 marks

### Case Scenario

Given Dataset **New Customer Segmentation** A retail company wants to categorize customers into different groups to provide personalized marketing campaigns. Mine the given data using the K-Means Clustering, Hierarchical Clustering to

5. group customers into different segments based on their purchasing behavior.

6. do a detailed analysis of the statistical values and identified groups.

7.

### **Dataset Features:**

CustomerID (Unique Identifier)

Age (Numeric)

Annual Income (\$)

Spending Score (1-100)

Purchase Frequency (Times per Month)

### **Dataset 1**



new\_market\_basket.csv

### **Dataset 2**



new\_customer\_churn.csv

### **Dataset 3**



new\_house\_prices.csv

### **Dataset 4**



new\_customer\_segmentation.csv

Note: All Questions are mapped to all course outcomes i.e. CO1, CO2, CO3, CO4