



# Department of Computer Science

Krantiguru Shyamji Krishna Verma Kachchh University

University Campus, Mundra Road, Bhuj, Gujarat - 370 001

<http://cs.kskvku.ac.in>

02832 235 036

[cs@kskvku.ac.in](mailto:cs@kskvku.ac.in)

Course: M.Sc. CA&IT – 6 (NEP)

Subject: Data Mining

Assignment Date: 19-02-2026

Assignment ID: DMASEMMSC62

Submission Date: 26-02-2026

## Instructions:

- Execute the following Programs in your **Jupyter Notebook** and write **all the code and outputs** in your **Practical Notebook**.
- Upload your **final Jupyter Notebook** on **GitHub repository**. Stick the **QR Code** of that repository on your blank page (left side) of your practical notebook.
- **Stick the outputs of graph or dataset** on your **blank page (left side)** of your **practical notebook**.
- The journal should include the student's **Name** on the cover page.
- A Journal should have a **properly formatted index** with all **program names, Date of experiment, Date of Submission**.
- Display the output on the **blank page opposite (in front of)** the page where the question and program code are written.

## Programs (Supermart\_Grocery\_Sales):

1. Load the Dataset of **Supermart\_Grocery\_Sales.csv** on Jupyter Notebook and print the first 10 rows.
2. Print that data whose profit is higher than 150 ( $\geq 150$ ) and display Customer, Category, Sales, Discount columns only.
3. Write a code to perform Data cleaning by:
  - a. Fill the null values of **Sales** column using **Average (Mean)** method.
  - b. Fill the null values of **Discount** column using **Backward Fill** method.
  - c. Fill the null values of **Profit** column using **Forward Fill** method.



4. Make a new column named **Sales\_Norm** and apply **Min-Max Normalization** on Sales column.
5. Convert the given dataset into transaction format by grouping products according to each Order ID. After preparing the transactional data, apply the Apriori algorithm using a suitable library to find frequent item combinations with a minimum support of 0.02. Then generate association rules from these frequent itemsets and display only those rules whose confidence is greater than or equal to 0.5. (Hint: You can use *mlxtend* library for this question)

At last, draw a scatter plot to show the relationship between Support and Confidence of the generated rules.

6. Create a new target variable named **Profit\_Category** by classifying Profit into "High" and "Low" based on the median value. Select suitable numerical features such as Sales, Discount, and Quantity. Split the dataset into training and testing sets, implement the **Naive Bayes classifier**, and display the Accuracy score and Confusion Matrix.

  
19/12/20

