# DS341 - Data Mining Project
# [10 marks (6 implementation, 4 viva)]

[CLO4---GA4---C4]

## Customer Behavior Analytics & Predictive Insights from E-Commerce Data

To provide students with practical exposure to real-world data mining by exploring a real-life e-commerce dataset. Students will apply core data mining techniques such as data preprocessing, association rule mining, classification, clustering, evaluation techniques and visual analytics to uncover and analyze customer behavior patterns and make predictive insights. This will culminate in a comprehensive presentation and formal report demonstrating the group's analytical process and findings.

### Overview

Each group will:

- Select a dataset (assigned or instructor-approved).
- Perform data cleaning, transformation, and preprocessing.
- Conduct Exploratory Data Analysis (EDA) to uncover trends.
- Apply Association Rule Mining to identify co-purchase patterns.
- Build Classification models (e.g., Decision Tree, Naive Bayes, KNN).
- Apply Clustering for customer segmentation.
- Evaluate the effectiveness of the applied data mining techniques.
- Generate business recommendations based on analytical results.
- Document their entire process in a formal report and deliver a group presentation.

### Submission Instructions

1. Create a folder named after your group, e.g., *GroupX_DataMiningProject*
2. Inside the folder, include the following files:
    a. Final Report (PDF)
    b. Complete Code File (preferably .ipynb or .py)
    c. Dataset used
3. Compress the folder into a single .zip file: *GroupX_DataMiningProject.zip*
4. Only one group member needs to upload the final zipped file on the submission portal.

### Project Report

This document must include clearly labeled sections:

1. Project Title & Group Members

2. Dataset Overview
    - Description of the dataset and source
    - Number of records and features
    - Problem statement

3. Preprocessing Details
    - Handling missing values
    - Feature encoding and transformation

4. Exploratory Data Analysis
    - Key statistics and visualizations
    - Observed trends or outliers

5. Association Rule Mining
    - Frequent itemsets and extracted rules
    - Interpretation of at least two strong rules

6. Classification Models
    - Description of algorithms used (DT, NB, KNN)
    - Evaluation and comparison of the performance of different classification models.

7. Clustering & Customer Segmentation
    - Cluster visualizations (e.g., K-means)
    - Interpretation of customer segments
    - Discussion on the quality and suitability of the obtained clusters.

8. Insights & Recommendations
    - Business value derived from patterns
    - Suggested actions

9. Challenges & Reflections
    - What was difficult? How was it solved?

10. Tools Used (libraries, packages etc)

11. Team Contribution Table (who did what)

## Dataset Requirements

If groups choose their dataset, it must meet the following:

- At least 500 rows
- Includes both categorical and numerical features
- Must be from a reliable real-world source (Kaggle, UCI, etc.)
- Toy datasets (Iris, Titanic, MNIST) are not allowed.

## Group Guidelines

- Maximum 4 members per group.
- Dataset must meet the minimum requirements.
- Task distribution and collaboration must be documented.
- Plagiarism will result in project disqualification.

***Deadline - Friday, 19th Dec 2025, 5:00pm***