Darshan University

Data Mining – 2301CS503 | Project

Project Overview

Objective: To help students apply data mining techniques (Association Rule Mining, Classification, Clustering) on real datasets using only core libraries (NumPy, Pandas, and visualization).

Tools Allowed: NumPy, Pandas, Matplotlib / Seaborn only. **No** high-level ML libraries (e.g., scikit-learn, TensorFlow).

Learning Outcome: Understand the full lifecycle of data mining — from preprocessing and exploration to model building and evaluation.

Datasets Description

Sr.	Dataset	Link	Algorithm
1	Online Retail	https://www.kaggle.com/datasets/vijayuv/onlineretail	Apriori
2	Heart Disease	https://www.kaggle.com/datasets/johnsmith88/heart-disease-dataset	ID3
3	Credit Card Customers	https://www.kaggle.com/datasets/sakshigoyal7/credit-card-customers	K-Means

Timeline & Weekly Plan

Task	Date Range	Task	Description
1,2,3	*16 Jun – 3 Aug*	*Data Preprocessing of all three Dataset*	*Clean & explore retail data. Handle missing values, outliers, and perform data transformations.*
4	04 Aug – 09 Aug	Apply Apriori on Online Retail Dataset	Implement Apriori algorithm. Generate frequent itemsets and association rules.
5	11 Aug – 16 Aug	Evaluate Apriori Results	Use support, confidence, lift for evaluation.
6	18 Aug – 23 Aug	Apply ID3 on Heart Disease Dataset	Implement ID3 decision tree. Train on preprocessed data.

7	25 Aug – 30 Aug	Evaluate Classification Results	Evaluate using accuracy, precision, recall; create decision boundaries and visualize.
8	01 Sep – 06 Sep	Apply K-Means on Credit Card Dataset	Implement K-Means. Use preprocessed features to cluster customer types
9	08 Sep – 13 Sep	Evaluate Clustering Results	visualize clusters

Preprocessing Tasks (Weeks 1–6):
25 mini tasks or questions per week (e.g., handling missing values, outlier detection, scaling, encoding).
Algorithm Implementation (Weeks 7–12):
No scikit-learn or built-in models allowed.
Only use NumPy, Pandas, and Matplotlib/Seaborn.
Clear modular implementation and visualizations are expected.

Progress of student is evaluated on every week.