AI Final Project

CLARA: A Lightweight Contrastive Embedding Approach for Retail E-Commerce Clustering

Group - 1

Instructor: Prof. Qazi

Student: You-Rui Feng **Student ID:** 1137141

GitHub: https://github.com/Yourui-droid/AI-final-project

Project Summary

CLARA improves clustering by combining baseline K-Means with an LDA projection learned from K-Means pseudo-labels. Workflow: preprocess & standardize features → K-Means → LDA embedding → K-Means in LDA space → evaluate (Elbow, Silhouette) and visualize (PCA 2D, LDA 2D).

Key Results (Silhouette Score)

Dataset	K-Means	CLARA
Mall Customers	0.289	0.469
Amazon	0.212	0.411
Flipkart	0.204	0.531
Walmart	0.039	0.209

Execution Instructions (Reproducibility)

1) Clone

git clone https://github.com/Yourui-droid/AI-final-project.git

cd AI-final-project

2) Install

python -m venv .venv

pip install -r requirements.txt

3) Dataset Use the existing CSVs under dataset/ (or follow dataset/README.md to download/place files).

4) Run Run notebooks in notebooks/ in numeric order (01 → 04 …).

5) Outputs Figures: results/figures/ | Metrics: printed and/or saved (if implemented).

Notes

- If Jupyter is missing: pip install jupyter, then run jupyter notebook.
- If datasets are not committed, download links and placement rules are provided in dataset/README.md.