ARI 510 HW2 – Unsupervised Learning and Feature Selection

Project Description

This project uses the Human Activity Recognition Using Smartphones dataset to explore:

- Standardization of features
- PCA for dimensionality reduction
- Clustering with K-Means and DBSCAN
- Feature selection (PCA + Random Forest importances)
- Supervised classification with Logistic Regression

Requirements

- Python 3.10+
- Libraries:
    - numpy
    - pandas
    - scikit-learn
    - matplotlib
    - seaborn

Install with:

- Create virtual environment
    - python -m venv venv
- Activate it
    - # Windows:
    - venv\Scripts\activate
    - # Mac/Linux:
    - source venv/bin/activate
- # You should see (venv) at the start of your terminal line
- # Install packages
    - pip install numpy pandas matplotlib seaborn scikit-learn jupyter

- Files
    - Har_analysis.ipynb – main notebook with all steps (EDA, PCA, clustering, feature selection, classification).
    - AlexG HW2 Report.pdf – written report with figures and analysis.

How to Run

- Create and activate a virtual environment .
- Install dependencies.
- Open har_analysis.ipynb in VS Code (or Jupyter).
- Run all cells from top to bottom.