Machine Learning Course - Project.

Part 1: Supervised Learning - Classifying Text or Images

Option A: Classify News Headlines into Categories (Text)

- **Dataset:** AG News Dataset: wangrongsheng/ag_news · Datasets at Hugging Face
- **Task:** Train a model (Logistic Regression, Decision Tree, Boosting, or KNN) to classify news headlines into categories: *World, Sports, Business, and Science/Technology*.
- **Steps:**
    1. **Preprocess text** (Tokenization, TF-IDF vectorization).
    2. **Train multiple classifiers** and compare accuracy.
    3. **Evaluate models** using precision, recall, and F1-score.

Option B: Classify Handwritten Digits (Images)

- **Dataset:** Digits Dataset (Scikit-learn) load_digits — scikit-learn 1.6.1 documentation
- **Task:** Train multiple models to recognize digits (0-9) and choose one best.
- **Steps:**
    1. **Flatten images** (Convert 8x8 pixel images to vectors).
    2. **Train multiple classifiers** and compare performance.
    3. **Evaluate models** using accuracy and confusion matrix.

Part 2: Unsupervised Learning - Clustering Text or Images

Option A: Clustering Movie Scripts into Genres (Text)

- **Dataset:** IMDB Movie Reviews IMDB Dataset of 50K Movie Reviews
- **Task:** group similar movie scripts into clusters (genres).
- **Steps:**
    1. **Convert text into numerical features** using TF-IDF.
    2. **Apply clustering algorithms** and find optimal hyperparameters.
    3. **Visualize clusters** using PCA or t-SNE.

Option B: Clustering Fashion Items (Images)

- **Dataset:** Fashion MNIST: Fashion MNIST
- **Task:** group fashion images into categories (shoes, shirts, bags, etc.).

- **Steps:**
    1. **Preprocess images** (convert grayscale to vectors).
    2. **Use PCA for dimensionality reduction** before clustering.
    3. **Interpret clusters** and check how well they match known categories.

Part 3: Association Rule Learning - Discovering Patterns

Market Basket Analysis on Grocery Data

- **Dataset:** Groceries Dataset: Groceries dataset
- **Task:** discover frequent item combinations in supermarket purchases.
- **Steps:**
    1. **Preprocess data** (convert transaction lists into a matrix format).
    2. **Apply Apriori algorithm** to find strong association rules (e.g., "People who buy milk often buy bread").
    3. **Visualize results** using network graphs.

Deliverables

1. **GitHub link for “Each” Part**
2. **Short Report** (Markdown format inside the notebook):
    - **Insights from EDA**.
    - **Comparison of supervised models**.
    - **Findings from clustering & association rule learning**.