

## ML (Machine Learning)

- Introduction to Machine Learning
  - What is Machine Learning?
  - Types of Machine Learning: Supervised, Unsupervised, and Reinforcement Learning.
  - Common terminologies: Features, Labels, Training Data, Testing Data, Overfitting, Underfitting.
- Supervised Learning
  - Linear Regression:
    - Overview and use cases.
    - How it predicts continuous values.
  - Logistic Regression:
    - Overview and use cases.
    - Binary classification and sigmoid function.
  - Decision Trees:
    - Overview and use cases.
    - How they split data based on features.
- Unsupervised Learning
  - Clustering:
    - Overview of K-Means Clustering.
    - Applications in customer segmentation.
  - Dimensionality Reduction:
    - Introduction to Principal Component Analysis (PCA).
    - Visualizing high-dimensional data in 2D.
- Exercises
  - Supervised Learning:
    - Use a pre-trained Linear Regression model to predict housing prices using the Boston Housing Dataset.
      - Load the dataset using Scikit-Learn.
      - Split the data into training and testing sets.
      - Use the provided Linear Regression model from Scikit-Learn.
      - Evaluate performance using Mean Squared Error (MSE).
    - Use a pre-trained Logistic Regression model to classify whether an email is spam or not using the SMS Spam Dataset
      - Load the dataset and preprocess text using libraries like NLTK.
      - Fit the provided Logistic Regression model and evaluate using accuracy and confusion matrix.
  - Unsupervised Learning
    - Perform K-Means Clustering on the Iris Dataset to group flowers based on features.
      - Load the Iris dataset from Scikit-Learn.
      - Use the provided K-Means implementation.
      - Visualize clusters using scatter plots.
    - Use PCA to reduce dimensions of the MNIST Dataset and visualize it in 2D.
      - Load the MNIST dataset.
      - Apply PCA from Scikit-Learn.
      - Visualize the first two principal components.
- Projects
  - Predict Loan Approval (Supervised Learning):
    - Dataset: Bank loan data from Kaggle.
    - Goal: Predict if a loan will be approved based on customer details.
    - Steps:
      - Preprocess the dataset (handle missing values and encode categorical variables).
      - Use the pre-built Decision Tree model from Scikit-Learn.

- Evaluate metrics like accuracy and F1-score.
- Customer Segmentation (Unsupervised Learning):
  - Dataset: E-commerce transaction data.
  - Goal: Group customers into segments using clustering.
  - Steps:
    - Preprocess data (normalize numerical features).
    - Use the pre-configured K-Means clustering model.
    - Visualize clusters using PCA and interpret segments.
    -