

# Machine Learning Models:

-Image Dataset:

Plant Disease Dataset

-Numeric Dataset:

GPA Prediction Dataset



# Project Overview

## Objective

- Plant Disease Classification
- Student Gpa Prediction

## Key Models

- logistic regression & Knn for Plant Disease
- linear regression & knn for Gpa prediction

# Image Dataset Documentation

## Plant Disease Dataset

- **Classes and Number of Samples:**
  - **Pepper Bell Bacterial Spot: 997**
  - **Pepper Bell Healthy: 1478**
  - **Potato Early Blight: 1000**
  - **Potato Late Blight: 1000**
  - **Tomato Early Blight: 1000**
- **Image Size**
  - **150x150**
- **Number of Training Samples**
  - **4106**
- **Number of Testing Samples**
  - **1369**
- **Number of Features ( Pixels )**
  - **67500**
- **Results for Model Testing Data**
  - **Logistic Regression**
    - **Accuracy: 0.78**
    - **Precision: 0.78**
    - **Recall: 0.78**
    - **AUC: 0.94**
  - **KNN**
    - **Accuracy: 0.55**
    - **Precision: 0.69**
    - **Recall: 0.55**
    - **AUC: 0.81**
  - **Graphs for ROC are available in the Notebook**



# Linear Regression: GPA Prediction

1

## Import Libraries

Python libraries: Pandas, NumPy, Scikit-learn, Matplotlib

2

## Preprocessing

Handle missing values, normalize numeric data, encode categorical features.





# Linear Regression: Training and Testing

1

## Data Splitting

Train-test split (e.g., 80% train, 20% test)

2

## Train the Model

Use Scikit-learn's `LinearRegression()` function.

3

## Model Evaluation

Metrics: MAE, MSE,  $R^2$  Score.



# KNN: Plant Disease Classification

1

## Import Libraries

Python libraries: NumPy, TensorFlow, Keras, Matplotlib

2

## Load Images

Use libraries like OpenCV or Keras to preprocess and label images.

3

## Data Splitting

X (features): Image pixel data; Y (labels): Disease/healthy category.

# KNN: Training and Testing

1

## Train the Classifier

Use Scikit-learn's  
`KNeighborsClassifier()`.

2

## Performance Evaluation

Metrics: Accuracy, Precision,  
Recall, F1 Score.



# Comparison For The Numeric Dataset:

## Linear Regression

-Mean Squared Error: 8.96928651238623e-06

-R<sup>2</sup> Score: 0.9999578605705769

## KNN Classifier

-Mean Squared Error: 0.03

-R<sup>2</sup> Score: 0.88



# Comparison For The Image Dataset:

## Logistic Regression

-Accuracy: 0.78

## KNN Classifier

-Accuracy: 0.55