Plant Disease Detection System for Sustainable Agriculture

Problem Statement:

Develop a CNN-based model capable of detecting and classifying plant diseases from images of leaves of various crops such as apple, cherry, grape, and corn. The model should accurately identify both healthy and diseased leaves while predicting the specific type of disease. This system will aid in precision agriculture by enabling early detection and effective disease management.

Pipeline of the project:

**1. Data Collection and Data Loading**  
Collect a dataset that contains images of both healthy and diseased plant leaves (from crops like apple, grape, etc.).  
Then divide this dataset into **three parts**:

* **Training set** – used to teach the model.
* **Validation set** – used to check the model’s performance while training.
* **Test set** – used to evaluate the model after training is done.

**2. Upload Dataset to Google Drive and Mount in Google Colab**  
Since Google Colab does not store files permanently, we:

* **Compress the dataset** into a ZIP file.
* **Upload the ZIP file to Google Drive**.
* **Mount Google Drive** in Colab so we can access the dataset from there.
* **Unzip the file** in Colab to use the images for training.

**3. Image Processing and Image Augmentation**

**What it means:**

* **Image Processing**: Resize and format all images properly (like making sure they’re all the same size and color format).
* **Image Augmentation**: Apply random changes to training images (like rotation, zoom, flipping) to increase variety and prevent overfitting. It helps the model generalize better.

**4. Building the CNN Model**

* CNN (Convolutional Neural Network) is a deep learning model that is very good at analyzing images.
* Here, we create a CNN architecture with layers like convolution, pooling, flattening, and dense (fully connected) layers.
* This model will **learn patterns from the leaf images** to identify diseases.

**5. Testing and Evaluation**

**What it means:**

* Once training is done, we use the **test data** to evaluate how well the model performs on unseen images.
* We check **accuracy, precision, recall**, and maybe show a **confusion matrix** to analyze which diseases are being predicted correctly or incorrectly.