Internship Week 1 Assessment

**Project Title:** Plant Disease Detection System for Sustainable Agriculture

# 1. Project Problem Statement Description

The agricultural sector forms the backbone of many economies and is deeply affected by plant diseases that can significantly reduce crop yields and quality. The early detection and accurate diagnosis of these diseases are crucial for effective management and sustainable farming. Traditional methods often rely on manual inspection, which is time-consuming, error-prone, and not scalable for large-scale agriculture.  
  
To address this challenge, our project aims to develop a Convolutional Neural Network (CNN)-based model capable of detecting and classifying plant diseases from images of crop leaves. The model will be trained on image data from crops such as apple, cherry, grape, and corn, and will be designed to differentiate between healthy and diseased leaves, while also identifying the specific disease when present.  
  
This system is a step forward in precision agriculture, enabling farmers to identify diseases at an early stage, make informed decisions, and apply targeted treatments. Ultimately, this contributes to improved crop management, reduced chemical usage, and more sustainable agricultural practices.

# 2. Project Pipeline Structure

The development of the plant disease detection system follows a structured pipeline to ensure an organized and efficient workflow. Here's an overview of the steps involved:

## A. Data Collection & Data Loading

The dataset is organized into three key folders:  
- Train: Images of various disease categories and healthy leaves used to train the CNN model.  
- Validation: Used to tune model parameters and prevent overfitting.  
- Test: Evaluates the model’s performance on unseen data.  
  
Each folder is further divided into subcategories such as:  
- Category1: Healthy  
- Category2: Diseased (e.g., Apple Scab, Grape Black Rot)

## B. Data Preparation

To ensure smooth model development on Google Colab, the dataset preparation includes the following steps:  
1. Dataset Compression: The folder structure is zipped.  
2. Drive Upload: The zip file is uploaded to Google Drive.  
3. Colab Integration: Google Drive is mounted on Google Colab.  
4. Data Extraction: The dataset is unzipped using Python code in Colab.  
5. Directory Structure Setup: The data is then ready for processing.

## C. Image Processing & Image Augmentation

Before feeding the data into the model, several preprocessing steps are applied to enhance image quality and diversity:  
- Resizing and Scaling of images to a standard format.  
- Augmentation Techniques such as:  
 - Rotation  
 - Zoom  
 - Flipping  
 - Brightness/contrast adjustment  
  
These augmentations increase the robustness of the model by simulating different real-world conditions.

## D. Model Training & Evaluation

- The training dataset is used to teach the model how to recognize patterns.  
- The validation dataset helps fine-tune model hyperparameters.  
- The test dataset is used to evaluate the final model’s accuracy and ability to generalize.  
- The model is developed using CNN layers for feature extraction and classification.