Implementation of First Module

Weekly Date

7/10/24 to 12/10/24

Dataset Collection & Preprocessing:

Objective:

To collect agricultural data (soil, crop, weather, etc.) and prepare it for machine learning model training and prediction.

Step 1: Dataset Collection

- **Data Sources:**
  - Historical agricultural datasets from Kaggle/Government open data platforms.
  - Custom datasets collected through surveys or IoT sensors.

- **Key Features Collected:**
  - Soil Nutrients: Nitrogen (N), Phosphorus (P), Potassium (K)
  - Soil pH and Moisture Level
  - Crop Type (categorical)
  - Weather Data: Rainfall, Temperature, Humidity
  - Farm Size (in acres or hectares)
  - Target Variable: Recommended Fertilizer Type

◆ **Step 2: Data Cleaning**

- **Handling Missing Values:**
  - Use mean/median for numerical fields (e.g., pH, moisture).
  - Use mode or drop records with missing categorical fields.

- **Removing Duplicates:**
    - Ensure no repeated entries distort model accuracy.

- **Fixing Inconsistencies:**
    - Standardize units (e.g., temperature in °C, rainfall in mm).
    - Convert string values to lowercase or encoded labels.

🔹 Step 3: Data Preprocessing

- **Encoding Categorical Variables:**
    - Apply Label Encoding for crop types and fertilizer types.

- **Feature Scaling / Normalization:**
    - Normalize numerical features like NPK, pH, moisture using MinMaxScaler or StandardScaler from scikit-learn.

Output:

A clean, normalized dataset ready for training the XGBoost or Random Forest model.