# Machine Learning for Sustainable Development Goal 12: Responsible Production and Consumption Goal 2: Zero Hunger

## 1. Introduction

**Project Objective**: To use machine learning to guide farmers to make data-driven decisions on selecting more suitable crop for better yields, aiming to support SDG 12 by optimizing land and water use and also supports SDG 2 by reducing the chances of crop failure supporting global efforts to end hunger and access to nutritious food.

**Motivation**: The agricultural sector is at the forefront of the fight against hunger and poverty. However, farmers frequently encounter obstacles in selecting the right crops due to a lack of access to scientific insights and data-driven recommendations. By utilizing machine learning, we aim to create predictive tools that can support.

## 2. Data Collection

**Data Source:** Kaggle Dataset (“Crop Prediction Dataset”)

**Dataset Description:**

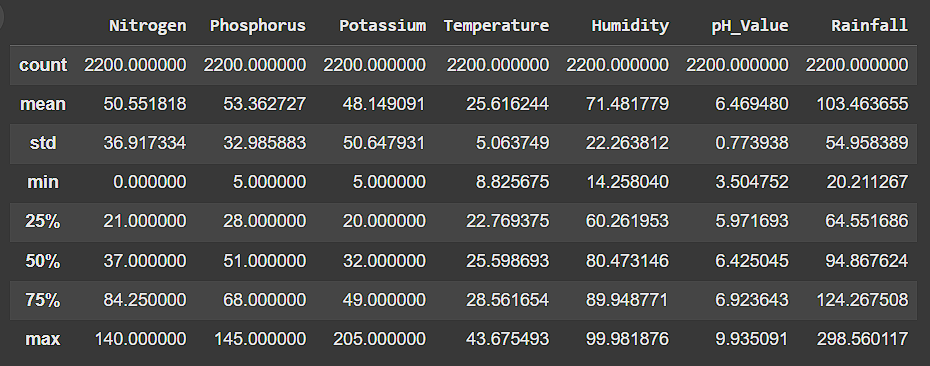
Features:

* Nitrogen - Nitrogen content in the soil
* Phosphorus - Phosphorus content in the soil
* Potassium - Potassium content in the soil
* Temperature - temperature in degrees Celsius
* Humidity - relative humidity in percentage
* pH\_Value - pH value of the soil
* Rainfall - rainfall in mm

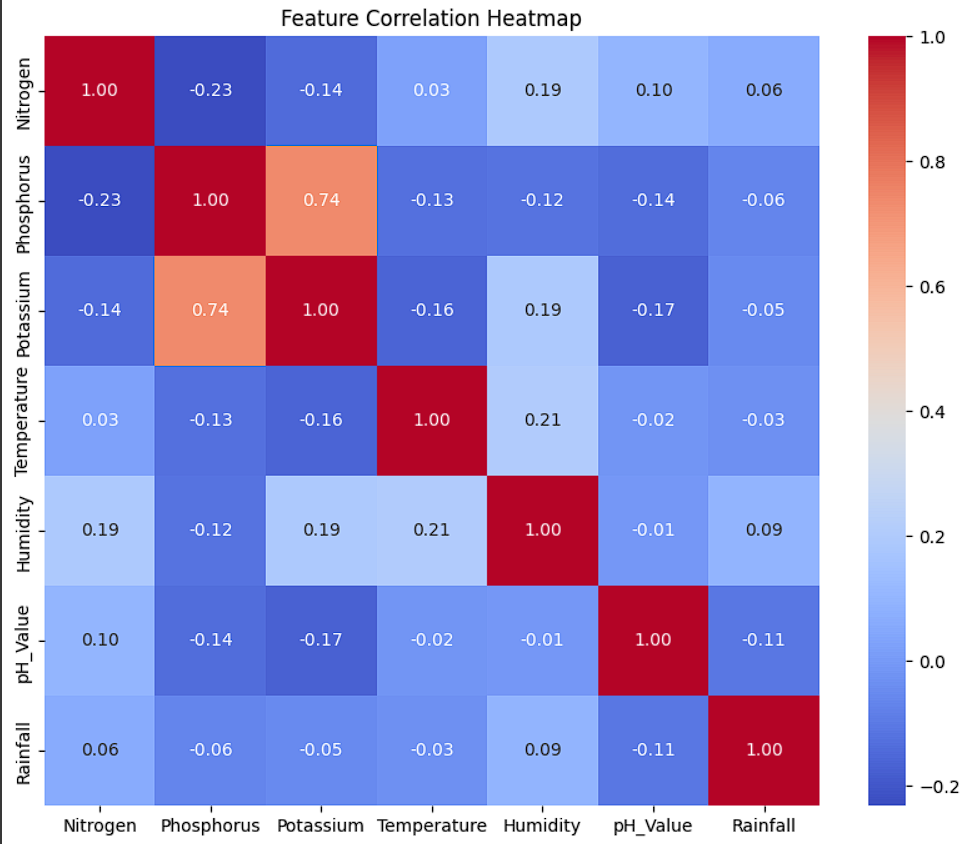
Size: 2200 rows by 8 columns  
Target Variable: Crop Name (Multi-class)

## 3. Exploratory Data Analysis (EDA)

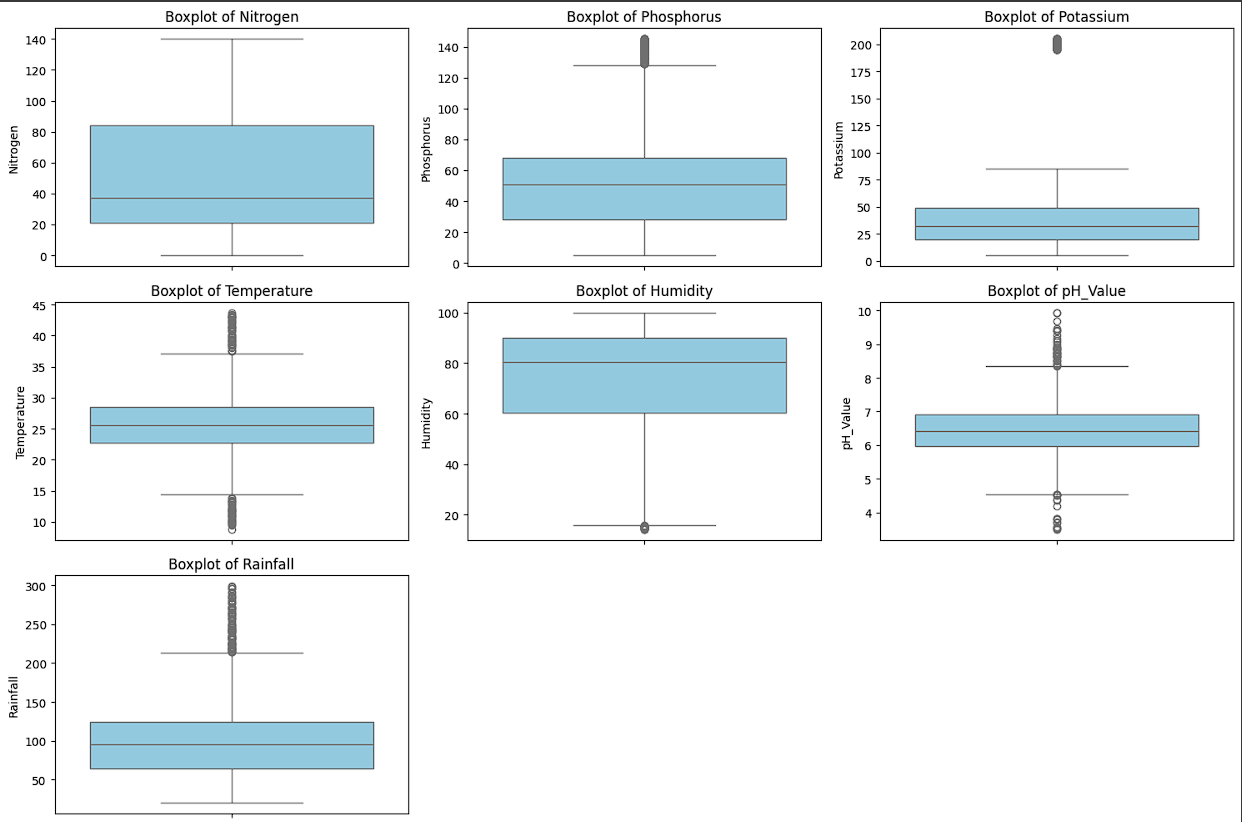
Summary Statistics: Mean, median, and distribution of each feature.



Visualizations:  
- Correlation heatmap to understand relationships between variables.



- Boxplots for outlier detection.

  
- Histograms to assess the distribution of each variable.

## 4. Data Preprocessing

Encoding Categorical Variables: Encodin32g the output variable using ‘Label Encoder’.  
Feature Scaling: Standardized features using `StandardScaler` for better performance in machine learning models.

## 5. Machine Learning Model Selection

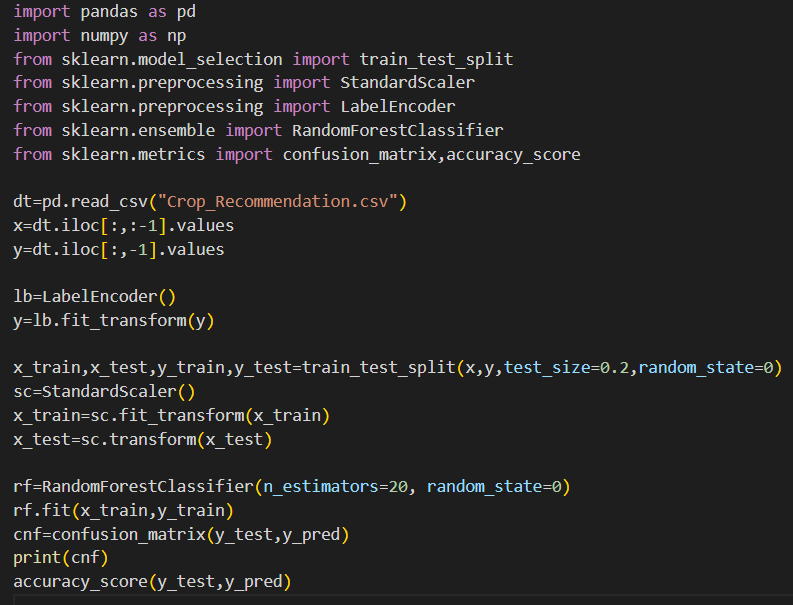
Model Choices:  
- Selected Random Forest Classifier which is best for multi class classifier which uses multiple

## 6. Model Implementation

Data Splitting: Split dataset into 80% training and 20% testing sets using `train\_test\_split` from Scikit-Learn.

**n\_estimators = 20**: This parameter specifies the number of decision trees in the forest.

***Code:***



## 7. Results and Evaluation

Model Performance:  
- Random Forest achieved an accuracy of 98%, F1-score of 99%   
Confusion Matrix: Visualized true vs. predicted values to identify common misclassifications.

## 8. Conclusion and Future Work

Key Takeaways: Machine learning models effectively predict type of crop based on Environment and Soil conditions.   
Future Improvements:  
- Real-time Prediction – developing web based or mobile application.  
- Expanding to a broader dataset covering various conditions.  
- Incorporate Expert Knowledge.

## 9. References

- Kaggle Dataset  
- Scikit-Learn Documentation