# ****Lab 10:****

**Bithon Macharia p101/1774g/21**

**Samuel Kariuki P101/1725G/21**

**David Murichu P101/1720G/21**

**Morris Kamande P101/1721G/21**

**Mbugua maxwel P101/1795G/21**

**Felix Mokaya P101/1823G/21**

**Victor Gabriel kahihu P101/1715G/21**

**Robert Mburu P101/1755G/21**

**Morompi Francis P101/1779gG21**

### ****Title****: ****Simulation and Modeling for Crop Cultivation and Weather Patterns****

### ****Abstract****

In this project, we aim to predict crop yield based on various weather and soil conditions using machine learning models. The project's goal is to help farmers and agricultural planners optimize their crop production and make informed decisions regarding agricultural planning. By utilizing advanced machine learning techniques such as Random Forest, Gradient Boosting, and XGBoost, we provide a robust prediction model that evaluates crop yields under different environmental factors. The project is implemented through a Streamlit web application, offering a user-friendly interface where users can input different weather and soil conditions and receive crop yield predictions.

### ****1. Introduction****

Agriculture plays a crucial role in ensuring food security across the world, and optimizing crop production is key to feeding a growing global population. Crop yield prediction is a vital aspect of agricultural planning, and accurate forecasting helps farmers make better decisions regarding planting, irrigation, and harvesting. Traditional methods of predicting crop yields rely heavily on manual assessments and limited data, which may not always provide accurate insights.

Recent advancements in machine learning offer an innovative solution for improving crop yield predictions by analyzing large datasets and identifying patterns in weather, soil conditions, and other environmental factors. In this project, we leverage machine learning algorithms to create a model that predicts crop yield based on various environmental parameters, enabling better decision-making in agriculture.

### ****2. Problem Statement****

Accurate crop yield prediction is essential for improving agricultural productivity, optimizing resource allocation, and ensuring food security. However, traditional prediction methods are often unreliable and do not account for the complex relationships between weather, soil conditions, and crop growth. Machine learning offers the potential to develop more accurate and dynamic models by analyzing large datasets of environmental conditions and crop yield outcomes.

This project addresses the problem of accurately predicting crop yields by leveraging machine learning algorithms and real-time data inputs, enabling farmers and agricultural planners to make more informed decisions that could lead to higher yields and more sustainable farming practices.

### ****3. Dataset Overview****

#### 3.1 Dataset Description

The dataset used for this project contains synthetic data that simulates weather and soil conditions affecting crop yield. The dataset includes the following features:

* **Temperature (°C)**
* **Rainfall (mm)**
* **Soil Quality (Rating from 1 to 10)**
* **Humidity (%)**
* **Solar Radiation (W/m²)**
* **Soil Moisture Content (%)**

The target variable is **crop\_yield**, which represents the yield (in tons per hectare) of a specific crop under the given weather and soil conditions. The dataset contains a total of 10,000 records, which were generated to reflect typical agricultural conditions.

#### 3.2 Data Preprocessing

Before training the machine learning models, several preprocessing steps were carried out:

* **Missing Data Handling**: Any missing or null values in the dataset were filled with the mean of the respective column.
* **Feature Scaling**: Some features, such as temperature and rainfall, were normalized to ensure that all features had the same scale.
* **Data Splitting**: The dataset was split into training (80%) and testing (20%) sets to evaluate the performance of the models.

### ****4. Machine Learning Models****

In this project, three different machine learning models were trained to predict crop yield based on the features in the dataset:

#### 4.1 Random Forest Regression

Random Forest is an ensemble method that uses multiple decision trees to improve prediction accuracy. Each tree in the forest is trained on a random subset of the data, and the final prediction is made by averaging the results of all trees.

Advantages:

* Handles non-linear relationships well.
* Robust to overfitting, especially when there are many trees.

#### 4.2 Gradient Boosting Regression

Gradient Boosting is an ensemble technique that builds trees sequentially, where each new tree corrects the errors made by the previous one. It combines weak learners to create a strong learner.

Advantages:

* Can handle complex data and capture non-linear relationships.
* Works well with a variety of data types.

#### 4.3 XGBoost

XGBoost (Extreme Gradient Boosting) is an optimized version of gradient boosting that incorporates additional features, such as regularization, to prevent overfitting and improve computational efficiency.

Advantages:

* High performance and speed.
* Can handle large datasets effectively.

### ****5. Model Evaluation****

The performance of each model was evaluated using the following metrics:

#### 5.1 R² Score

The R² score, also known as the coefficient of determination, measures how well the model’s predictions match the actual values. A score closer to 1 indicates better performance.

#### 5.2 Root Mean Squared Error (RMSE)

RMSE is used to measure the average magnitude of the prediction errors. Lower RMSE values indicate better model performance.

#### 5.3 Model Comparison

After training all three models, their performances were compared based on the R² score and RMSE. The model with the highest R² score and lowest RMSE was selected for further use.

### ****6. Streamlit Web Application****

To make the model accessible to users, we developed a simple Streamlit web application. The web app allows users to input various weather and soil parameters (such as temperature, rainfall, humidity, etc.) and receive real-time predictions of crop yield.

#### 6.1 Features of the Web Application:

* **User Input**: Users can input values for weather conditions (temperature, rainfall, etc.) and soil properties (soil quality, moisture).
* **Model Integration**: The best-performing model is loaded in the backend, and predictions are made in real-time.
* **Prediction Output**: The predicted crop yield is displayed to the user, providing an estimate based on the inputs.

This web app provides a user-friendly interface for agricultural planners, farmers, and researchers to obtain accurate crop yield predictions tailored to their specific conditions.

### ****7. Results and Discussion****

#### 7.1 Model Performance

The evaluation results indicated that XGBoost outperformed the other two models in terms of both R² score and RMSE. It achieved an R² score of 0.92, indicating that it was able to explain 92% of the variance in the crop yield data. The RMSE for XGBoost was the lowest among the models, indicating that its predictions were closest to the actual values.

#### 7.2 Web Application Utility

The Streamlit web application provided a simple and effective interface for interacting with the machine learning model. It successfully demonstrated the feasibility of using machine learning for real-time crop yield prediction. The application could be expanded to include more features, such as historical yield data analysis or integration with real-time weather APIs.

### ****8. Conclusion and Future Work****

This project demonstrates the power of machine learning in predicting crop yield based on various environmental factors. By using advanced algorithms such as Random Forest, Gradient Boosting, and XGBoost, we were able to build an accurate model that can be deployed in a user-friendly web application.

#### Future Work:

* **Integration with Real-Time Weather Data**: The application could be enhanced by integrating real-time weather data from APIs to provide live predictions based on current conditions.
* **Model Improvement**: Further hyperparameter tuning and additional features, such as crop type and location, could be incorporated to improve the accuracy of the model.
* **Scalability**: The project could be expanded to handle larger datasets and more complex predictions, such as multi-crop yield forecasting.