Machine Learning to Predict Coffee Yields in Uganda

1. Background and Problem Statement

Coffee is Uganda's leading export crop, and it generates a significant amount of household income and national revenue. Despite its contribution, the sector has been struggling with yields forecasting because of variations in climatic conditions, pests, and inconsistency in data collection.
History data from the Uganda Coffee Development Authority (UCDA) and the Uganda Bureau of Statistics (UBOS) exist, but they are not projections that can help farmers and policymakers plan.

Since climate change is increasingly shaping rainfall and temperature patterns, yield prediction is now an imperative. This project tries to use machine learning techniques to predict regional coffee yields from combined datasets from agricultural, climatic, and remote sensing sources.

2. Purpose and Objectives
General Objective:

Designing a machine learning model to predict coffee yield in Uganda from multi-source datasets.

Specific Objectives:

To collect and assemble coffee datasets like production, climate, and satellite datasets.

To identify the most important features influencing coffee yield.

To test and train machine learning models for predicting yield.

To compare the performance of models from different combinations of data.

To create region-based prediction of yield to drive planning and policy making.

3. Importance of the Study

Agricultural Planning: Enables farmers and cooperatives to project yield outcomes and optimize farm management.

Policy Decision Support: Provides government authorities with timely and data-based information for export planning and resource allocation.

Climate Adaptation: Enables assessment of the impact of changing climatic trends on coffee production.

Technological Progress: Demonstrates how AI can be applied in the solution of agricultural issues in African environments.

4. Methodology

Data Collection:

The study will collect data from:

UCDA / UBOS: Coffee plot, tree count, trees in production, and household data.

FAOSTAT & USDA Reports: Annual national coffee production statistics.

Satellite & Climate Sources:

CHIRPS rainfall data

NASA POWER temperature and solar radiation data

Sentinel-2 NDVI for plant health

Data Integration:

The data will be merged by region and year to generate an integrated analytical dataset. The dataset will include structural, climatic, and satellite-based variables.

Data Analysis and Modeling

Data cleaning and feature engineering (normalization, filling missing values).

Model training using Linear Regression, Random Forest, and XGBoost.

Model evaluation using R^2, RMSE, and MAE.

Model comparison between training on individual vs. grouped datasets.

5. Expected Outcomes

Cleaned and formatted coffee yield dataset (2010–2022).

Trained and validated machine learning model for predicting yield.

Climatic and environmental factors most responsible for coffee yield.

A visualization dashboard or report of local yield predictions.

Data-driven production planning suggestions for Uganda.

6. Scope and Limitations

The study will focus on Uganda's principal coffee-growing areas (Central, Western, Eastern, Northern, and South-West).

Limitations can include incomplete historical data, varying data resolution, and challenges in matching satellite and ground data. These will be overcome by interpolation and feature aggregation techniques.

7. Expected Impact

Successful completion of this project will demonstrate how machine learning can be used to improve agricultural forecasting in Africa. The model can be scaled up to other crops or used in national agricultural early warning systems.

9. Keywords

Coffee production, machine learning, Uganda, climate data, satellite data, crop forecasting