# Crop Production Analysis and Prediction

## 1. Introduction

This project analyzes global agricultural production data using the FAOSTAT dataset. The objective is to understand crop trends, identify productivity patterns, and build predictive models to estimate crop production based on historical data.

## 2. Approach

The workflow involved data cleaning, exploratory data analysis (EDA), feature engineering, and machine learning model development. A Decision Tree Regressor was selected as the best model based on its accuracy in predicting crop production.

## 3. Data Cleaning

Column names were standardized, irrelevant or missing data was removed, and duplicates were dropped. Key columns retained include area, item, element, year, unit, and value.

## 4. Exploratory Data Analysis (EDA)

- Crop Distribution: Identified most and least cultivated crops across different regions.

- Geographical Trends: Found specific regions focusing on high-production crops.

- Yearly Trends: Observed increase in production over years for certain crops and regions.

- Yield vs Area: Detected correlation between yield and area harvested indicating productivity.

## 5. Visualizations

- Bar charts and pie charts for crop frequency by region.  
- Line plots for year-wise trends in production and yield.  
- Heatmaps to understand correlation between Area, Yield, and Production.

## 6. Predictive Modeling

Various regression models were evaluated, including Linear Regression, Random Forest, and Decision Tree. The Decision Tree Regressor achieved the best performance with approximately 94% R² score. It was used to build a Streamlit application for interactive predictions.

## 7. Actionable Insights

- Promote crops with high yield but low cultivation area to improve food security.

- Optimize resource allocation to regions with consistent yield improvements.

- Use predictive modeling to forecast future production and plan ahead.

## 8. Conclusion

This project demonstrates the power of data-driven agriculture through comprehensive analysis and predictive modeling. It helps policymakers and agricultural planners in making informed decisions based on historical crop data.