Crop Recommendation Project

This project uses machine learning to recommend optimal crops. It considers soil and environmental features. The goal is to maximize agricultural yield.

by Maxwell Njoroge

Problem Statement

Crop Selection Challenges

Farmers struggle to choose the best crop. Local soil and environmental conditions vary widely.

Machine Learning Solution

We developed a model to predict suitable crops. This guides farmers' planting decisions effectively.

Dataset Overview

Features	Nitrogen, Phosphorus, Potassium, Temperature, Humidity, pH, Rainfall
Target	Recommended crop label (22 classes)
Size	2200 samples
Data Quality	No missing values or duplicates

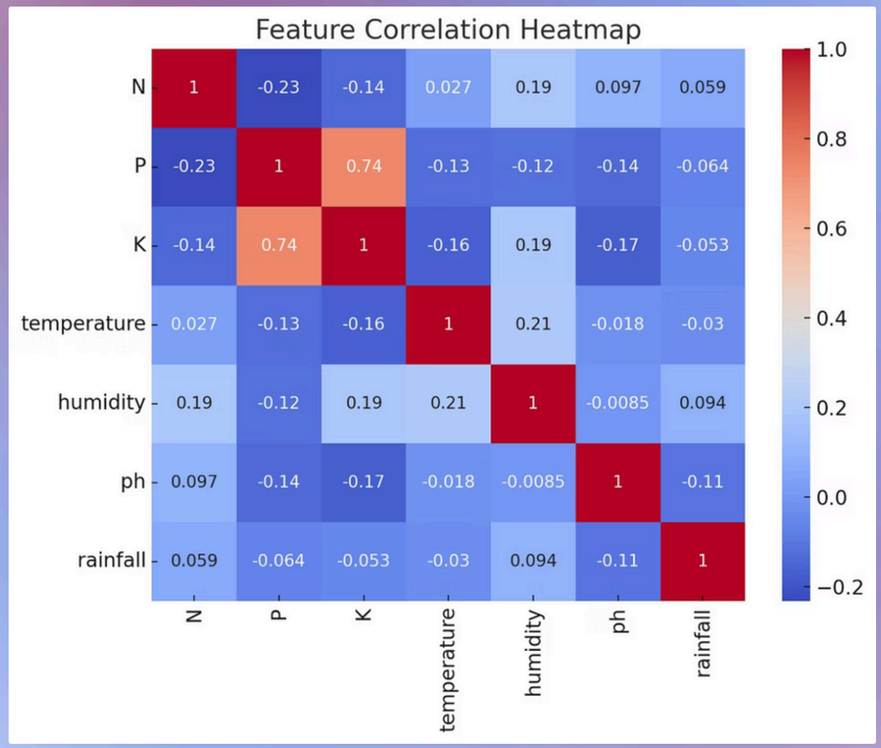
The dataset provides comprehensive information. It covers critical soil and environmental factors. This allows for robust crop recommendations.

Exploratory Data Analysis: Correlation Heatmap

The correlation heatmap visualizes relationships between features. It reveals key dependencies and independent variables. This guides feature selection for modeling.

- **Most features are weakly correlated.** This is a good sign for machine learning because it means each feature brings unique information.
- **Temperature and Humidity** show some mild correlation (they often rise together), which makes sense environmentally.
- **Soil nutrients** (N, P, K) are relatively independent from environmental factors like Rainfall or pH.

Implication: There's minimal redundancy between features, so all variables likely contribute uniquely to the crop classification.



Modeling Approach

Classifiers Utilized

- Logistic Regression (baseline)
- Decision Tree
- Random Forest (tuned)

Performance Metrics

- Accuracy score
- Classification report
- Confusion matrix

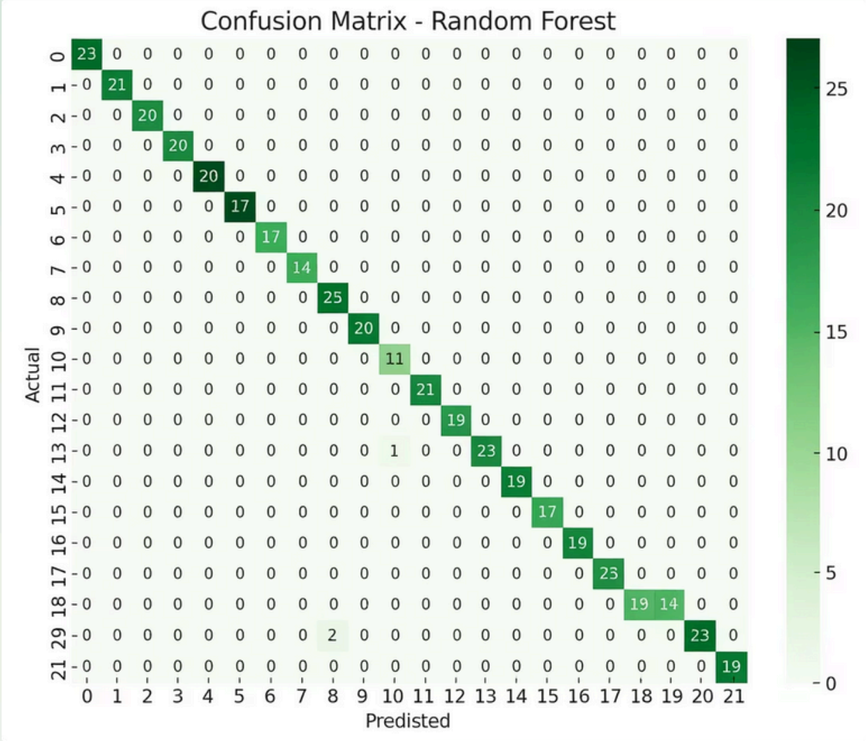
We selected a variety of models. This ensures a comprehensive evaluation. Hyperparameter tuning optimized Random Forest performance significantly.

Model Evaluation: Confusion Matrix

The confusion matrix quantifies model performance. It details correct and incorrect predictions per crop. This helps identify misclassifications.

- Most predictions fall along the diagonal, meaning **high accuracy**.
- Some confusion exists between similar crops, especially those that grow under overlapping conditions (e.g., **mothbeans vs. pigeonpeas**).
- A few classes had minor misclassifications, likely due to environmental overlap or limited sample data.

Implication: The model performs well overall but may benefit from fine-tuning or feature engineering to distinguish borderline cases better.



Top Feature Importances

Humidity Impact

Humidity ranks as a highly important feature. It significantly influences crop suitability.

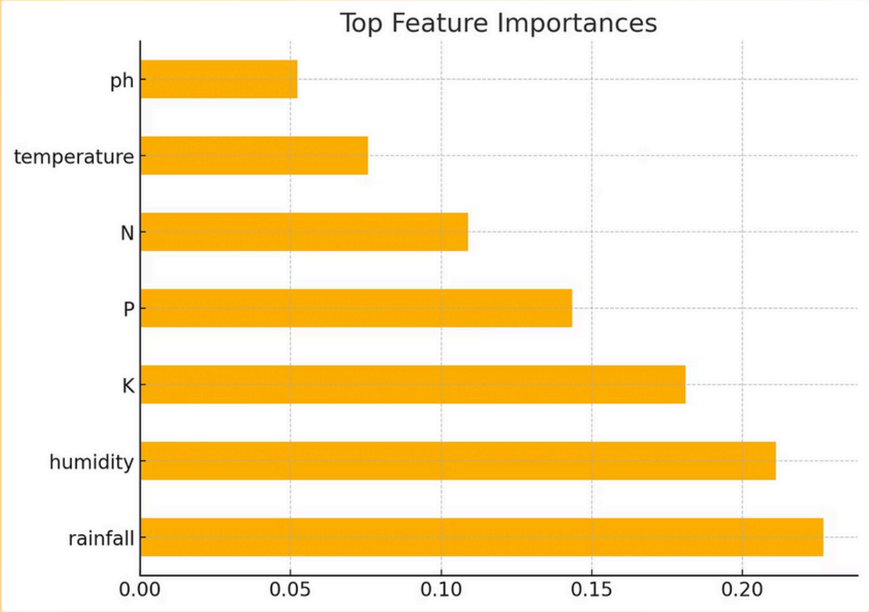
Temperature Influence

Temperature is another critical factor. It directly affects crop growth and viability.

Nutrient Significance

Nitrogen, Potassium, and Phosphorus are vital nutrients. They strongly correlate with optimal crop yield.

Feature importance highlights key environmental factors. Understanding these helps in targeted agricultural practices. This also informs future data collection efforts.



Conclusion & Recommendation

- **Recommendation:** Utilize the Random Forest model for robust crop recommendations.
- **Impact:** This tool optimizes agricultural planning, boosting farm productivity.
- **Deployment:** Develop a user-friendly mobile or web application for farmers.

The Random Forest model excels at crop prediction. It provides actionable insights for farmers. This technology can revolutionize farming practices, making them more data-driven and efficient.