Technical Documentation: Crop Recommendation System

Major Project Report

Student Name: PRINCE GURUNG
Roll No: 2317057
Date: 31 January 2026
GitHub Repository: https://github.com/princegurung421-dev/Crop-Recommendation

1 Project Overview

Agriculture is one of the most important sectors for food security and economic growth. However, selecting the correct crop for cultivation is a complex decision that depends on both soil nutrients and environmental conditions. Many farmers rely on traditional knowledge, which can be difficult when climate patterns change or when the soil quality differs across regions.

The **Crop Recommendation System** is a Machine Learning based application that recommends the most suitable crop for cultivation based on:

- **Soil Nutrients:** Nitrogen (N), Phosphorus (P), Potassium (K)
- **Soil Chemistry:** pH level
- **Environmental Factors:** Temperature, Humidity, Rainfall

The main objective is to provide a fast and accurate crop suggestion that can help increase productivity and reduce the risk of crop failure.

2 Dataset Analysis

The dataset used for this project contains **2200 records** and includes **7 input features** with **1 target label**. The target label represents the crop type, and the dataset contains **22 unique crops**.

2.1 Features and Target

Input Features:

- Nitrogen (N)
- Phosphorus (P)
- Potassium (K)
- Temperature

- Humidity
- pH
- Rainfall

Target Output:

- Crop label (22 crop categories)

2.2 Data Distribution

The dataset is well balanced, with each crop having an equal number of samples. This is an advantage because it reduces the risk of model bias towards a particular crop class.

2.3 Correlation and Nutrient Patterns

To understand how features relate to each other, correlation analysis is performed. For example, certain crops may show strong relationships between nutrient requirements and rainfall. Similarly, pH levels strongly influence the type of crops that can grow in specific soil types.

In addition, crops have different nutrient profiles:

- Some crops require high Nitrogen and moderate rainfall.
- Some crops thrive in low Nitrogen but high humidity.
- pH can significantly affect crop suitability.

These variations make Machine Learning a strong solution for crop selection.

3 Machine Learning Model

This system uses an ensemble-based supervised learning model. Based on model performance and suitability, a **Random Forest Classifier** is used.

3.1 Why Random Forest?

Random Forest is selected because it provides strong performance for classification problems involving non-linear patterns.

- **High Accuracy:** Works well with complex relationships between soil and environment.
- **Reduced Overfitting:** Uses multiple decision trees and averages results.
- **Handles Mixed Feature Effects:** Performs well even when features interact.
- **Feature Importance:** Helps identify the most influential factors for crop selection.

3.2 Mathematical Concepts

3.2.1 Feature Scaling

Feature scaling ensures that all features contribute fairly during model training. Two common methods are used:

Min-Max Scaling:

$$X_{scaled} = \frac{X - X_{min}}{X_{max} - X_{min}}$$

Standardization:

$$X_{standardized} = \frac{X - \mu}{\sigma}$$

Where μ is the mean and σ is the standard deviation.

3.2.2 Entropy and Information Gain

Random Forest is built from Decision Trees. Decision Trees split data using entropy and information gain.

Entropy:

$$H(S) = -\sum p_i \log_2(p_i)$$

Information Gain:

$$IG(S, A) = H(S) - \sum \frac{|S_v|}{|S|} H(S_v)$$

These concepts help the model choose the best splits for classification.

4 System Architecture

The system is implemented using a simple web-based architecture:

1. **Frontend:** HTML5, CSS3 (Glassmorphism UI), Bootstrap 5
2. **Backend:** Flask (Python)
3. **ML Model:** Scikit-learn trained model stored as `model.pkl`

4.1 Prediction Workflow

The prediction pipeline follows these steps:

1. User enters values for N, P, K, temperature, humidity, pH, and rainfall.
2. Frontend sends the form data to Flask backend.
3. Backend preprocesses values using saved scalers:
   - `minmaxscaler.pkl`
   - `standscaler.pkl`
4. Model generates the crop recommendation.
5. Backend returns the predicted crop name to the UI.

5 Model Evaluation

The model is evaluated using classification metrics and visual tools.

5.1 Confusion Matrix

A confusion matrix provides a clear overview of how many samples were correctly and incorrectly classified. It helps identify which crops may be confused with others due to similar soil/environment patterns.

5.2 Feature Importance

Feature importance helps interpret the model decision-making. For example:

- Rainfall and humidity may strongly affect tropical crop selection.
- Nitrogen and potassium often influence cereal crops.
- pH can be critical for crop survival.

6 Deployment

The application is deployed as a Flask web application. It exposes a REST endpoint:

POST `/predict`

The endpoint accepts form input, preprocesses it, and returns the predicted crop.

This makes the system usable as a web-based decision support tool for agriculture.

7 Conclusion

The Crop Recommendation System demonstrates how Machine Learning can be applied to agriculture for smarter crop planning. By analyzing soil nutrients and environmental factors, the system provides a reliable crop suggestion. The Random Forest model offers both strong performance and interpretability through feature importance.

In the future, the project can be extended using real-time weather APIs and larger datasets for better regional adaptation.

A Appendix

A.1 Evaluation and Visual Analysis

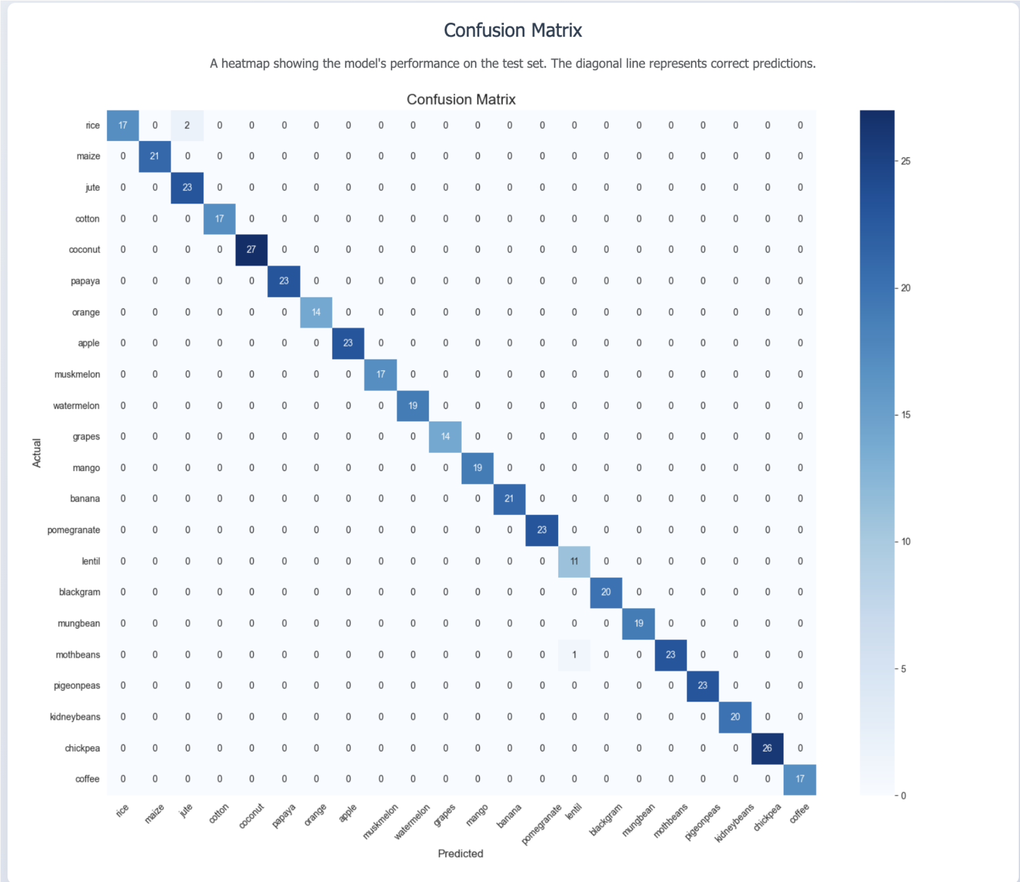

Figure 1: Confusion Matrix of Crop Recommendation Model

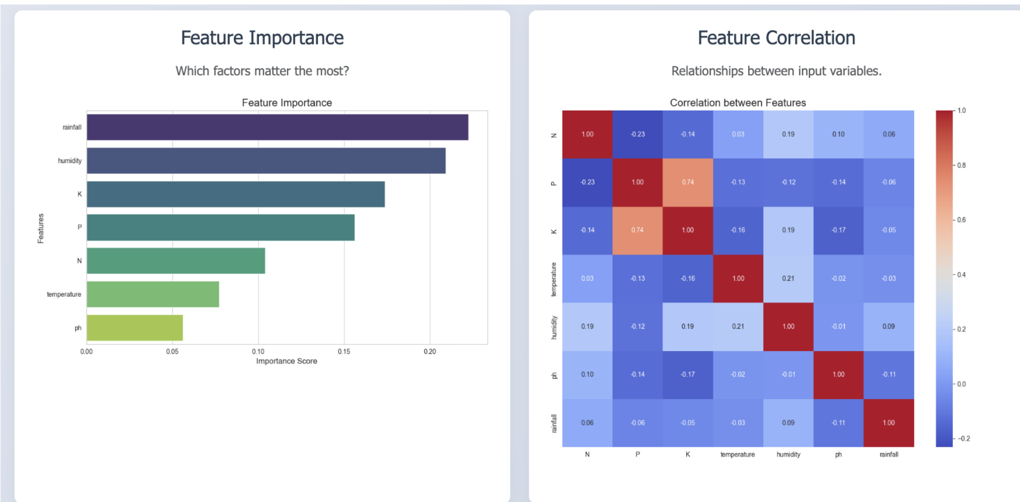

Figure 2: Feature Importance Plot from Random Forest Classifier

A.2 Crop Distributions Across Different Parameters

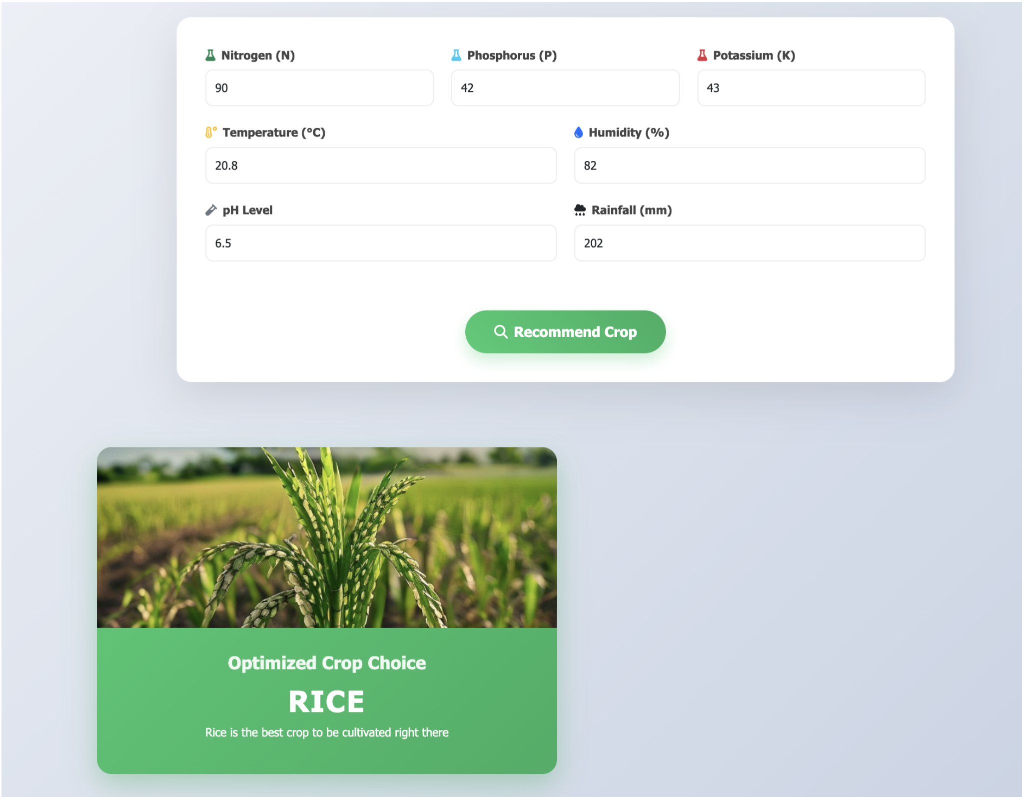

Figure 3: Crop Variation Across Different Soil and Weather Parameters (Plot 1)

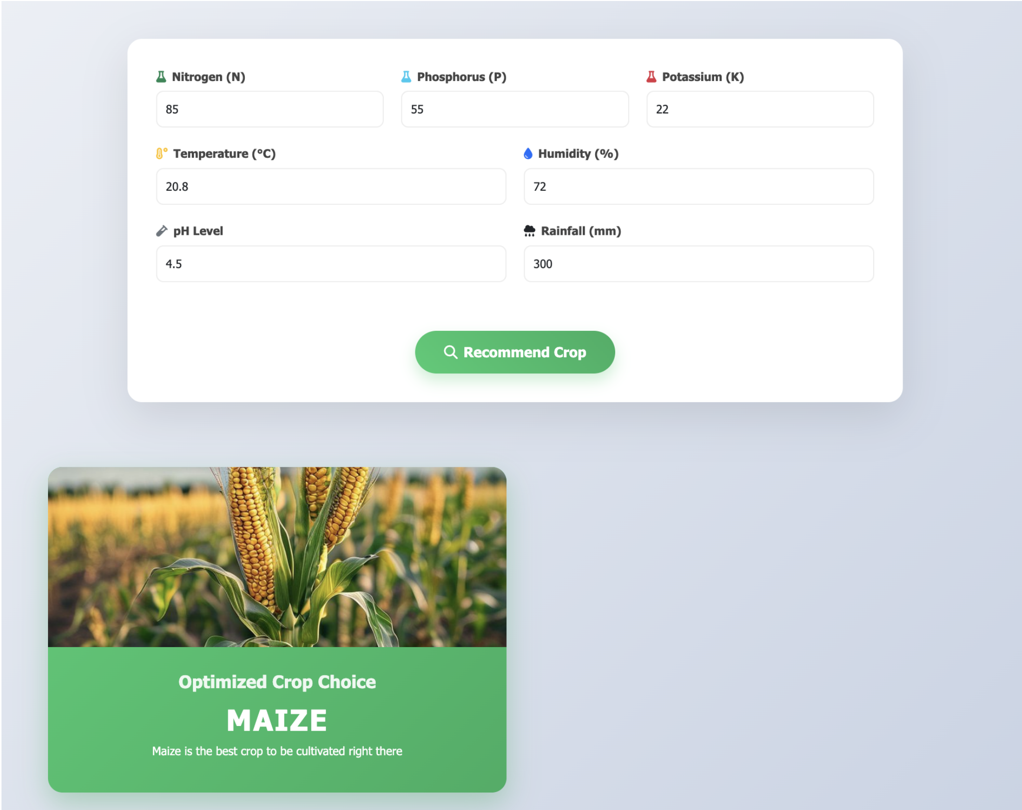

Figure 4: Crop Variation Across Different Soil and Weather Parameters (Plot 2)

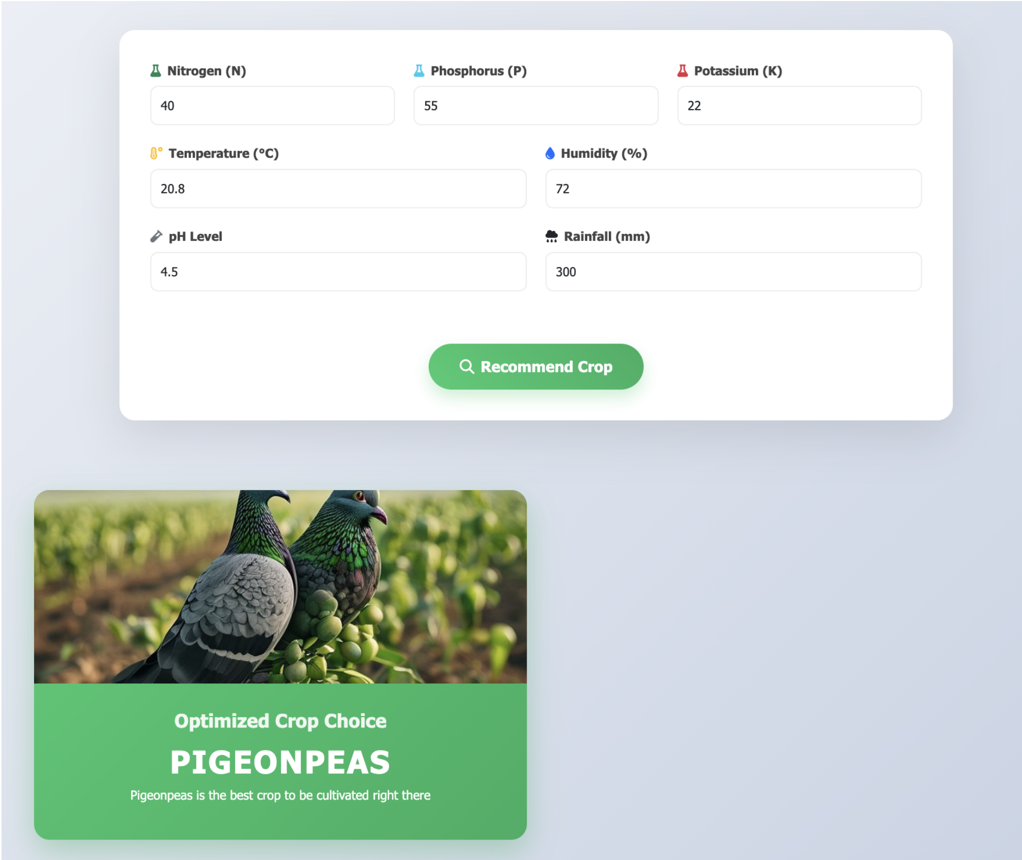

Figure 5: Crop Variation Across Different Soil and Weather Parameters (Plot 3)

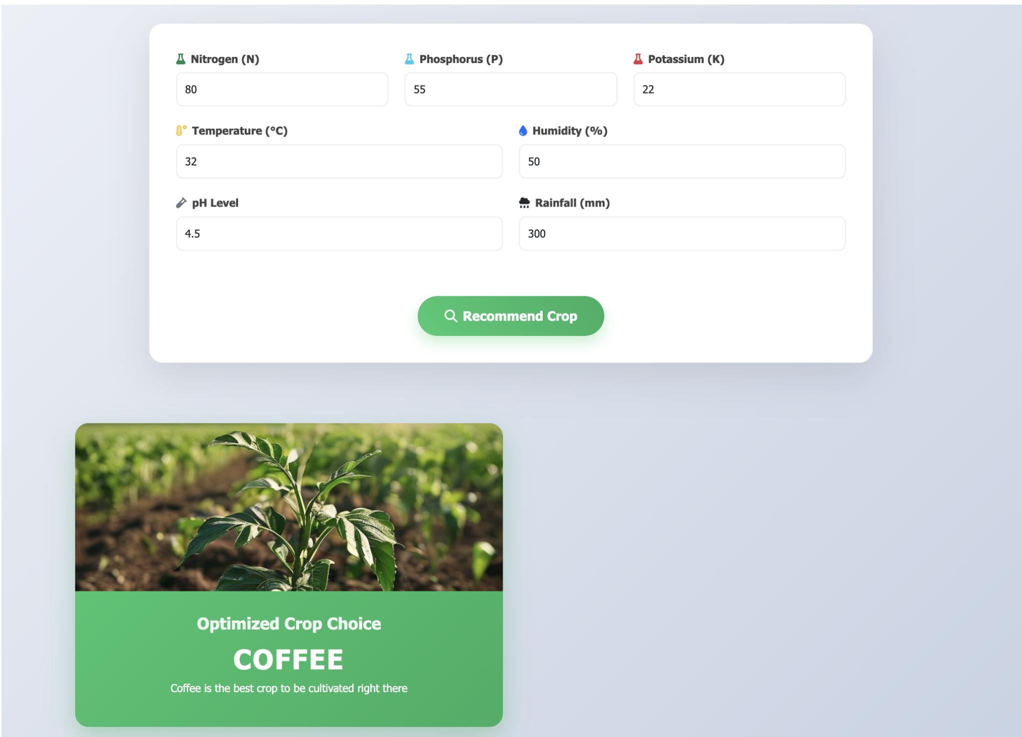

Figure 6: Crop Variation Across Different Soil and Weather Parameters (Plot 4)