Smart Farming: Crop Recommendation System

[Shivendra Tripathi]

November 2, 2024

Abstract

This project aims to develop a crop recommendation system using machine learning techniques to help farmers select the most suitable crop based on environmental and soil conditions. Using data on soil nutrients, temperature, humidity, pH levels, and rainfall, we applied various predictive models, including Logistic Regression, Decision Tree, and Support Vector Machine, to accurately predict crop suitability. The Decision Tree model provided the best accuracy and was chosen as the final model for deployment.

Contents

1 Introduction

Agriculture plays a crucial role in sustaining life and the economy. Smart farming techniques, which use data-driven insights, can improve crop yield and efficiency. The goal of this project is to recommend suitable crops based on soil and environmental conditions, providing farmers with valuable insights for crop selection.

This project uses a dataset that includes soil content (N, P, K), temperature, humidity, pH levels, and rainfall for different crops. Machine learning techniques are applied to build predictive models that suggest the best crop for specific climatic conditions.

2 Dataset Description

The dataset contains 2,200 samples with the following features:

- **N**: Ratio of Nitrogen in the soil
- **P**: Ratio of Phosphorous in the soil
- **K**: Ratio of Potassium in the soil
- **Temperature**: Temperature in Celsius
- **Humidity**: Relative humidity in percentage
- **pH**: pH level of the soil
- **Rainfall**: Rainfall in millimeters
- **Label**: Crop type (e.g., rice, maize, cotton)

Each crop type has 100 samples, providing a balanced dataset for analysis and model training.

3 Data Preprocessing

3.1 Data Cleaning

Data preprocessing included handling missing values and outliers.

- **Missing Values**: There were no missing values in this dataset.
- **Outliers**: No significant outliers were observed in the dataset, given the balanced nature of the data.

3.2 Data Transformation

The average values for soil content (N, P, K) and environmental factors (temperature, humidity, pH, and rainfall) were calculated to gain insights into the general requirements for each crop.

4 Exploratory Data Analysis

To understand the dataset better, various visualizations were created to observe the distribution of features and their impact on crops.

4.1 Feature Distributions

Figure 1 shows the distributions for key features:

- Nitrogen (N), Phosphorous (P), Potassium (K)
- Temperature, Humidity, pH level, Rainfall

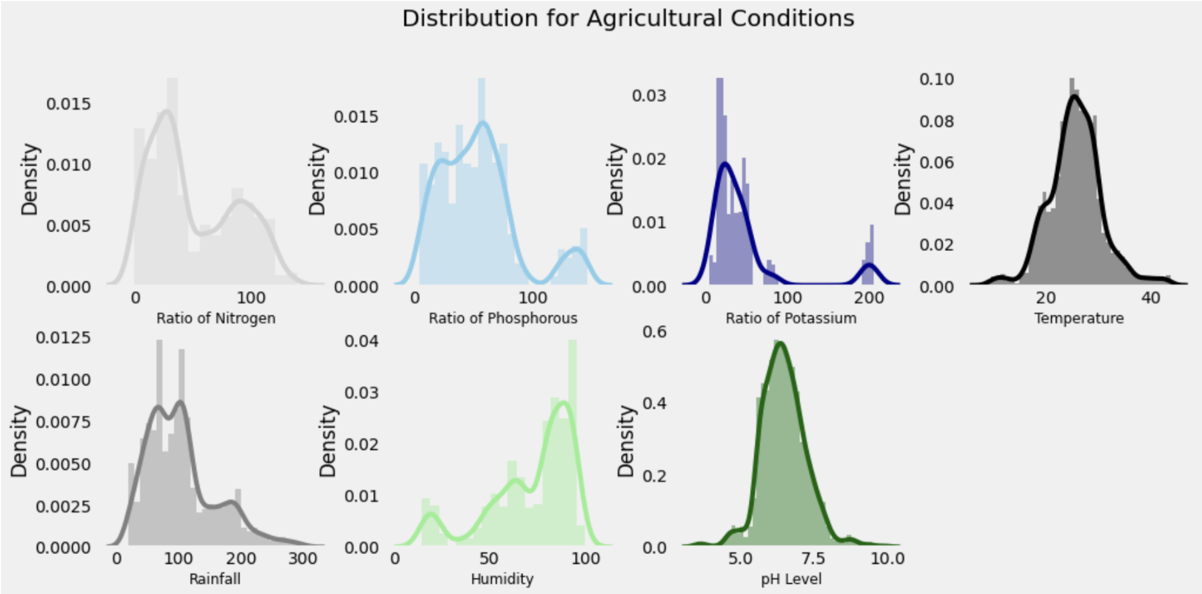

Figure 1: Distribution of Agricultural Conditions

4.2 Cluster Analysis

K-Means clustering was performed to identify groups of similar crops based on their nutrient and environmental requirements. Figure 2 shows the Elbow Method used to determine the optimal number of clusters.

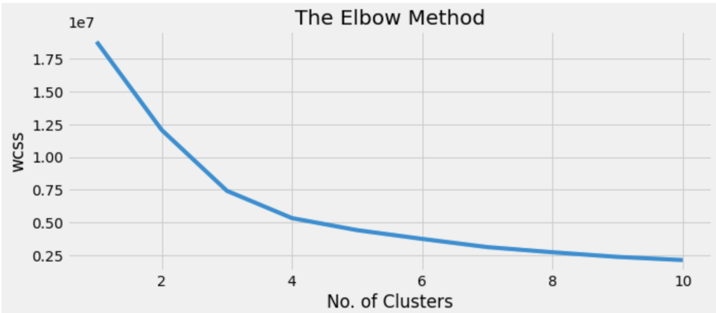

Figure 2: Elbow Method for Optimal Clusters

5 Modeling

Three machine learning models were used to predict the best crop based on the input features:

- **Logistic Regression**: Achieved an accuracy of 96.8%.
- **Decision Tree**: Achieved the highest accuracy of 98.2%.
- **Support Vector Machine (SVM)**: Achieved an accuracy of 97.7%.

The Decision Tree model was selected as the final model due to its superior accuracy.

6 Results

6.1 Model Evaluation

The performance of each model was evaluated using a Confusion Matrix and Classification Report. Figure 3 displays the Confusion Matrix for the Decision Tree model.

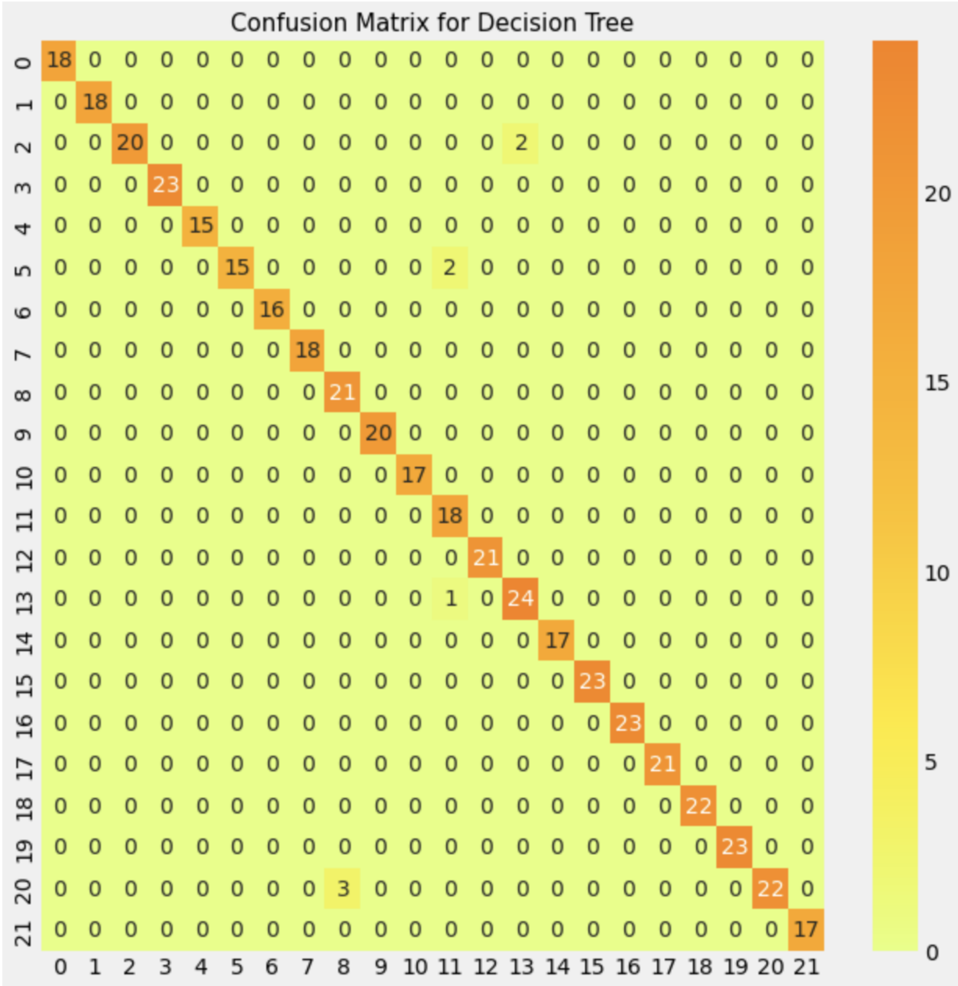

Figure 3: Confusion Matrix for Decision Tree Model

6.2 Sample Prediction

An example of real-time prediction based on user input is provided below:

```
Enter N : 100
Enter P : 50
Enter K : 40
Enter Temperature : 30
Enter Humidity : 70
Enter Ph : 6.5
Enter Rainfall : 150

Suggested Crop: Rice
```

7 Conclusions

The "Smart Farming" project successfully implemented a crop recommendation system using machine learning. The Decision Tree classifier achieved the best results, accurately predicting the best crop with an accuracy of 98.2%. This system can help farmers make informed decisions based on soil and environmental conditions.

8 Future Work

The project could be enhanced with:

- **More Crop Types**: Adding more crop data to improve the diversity of predictions.
- **Real-Time Data Integration**: Using IoT sensors to gather real-time data for more accurate recommendations.
- **Web Application Deployment**: Deploying the model as a web application for broader accessibility.

9 References

- Scikit-learn documentation: `https://scikit-learn.org/`
- Matplotlib documentation: `https://matplotlib.org/`
- Dataset Source: Provided by course materials or university resources.