Crop Recommendation using Machine Learning

Objective

The primary objective of this project is to develop a Machine Learning model that can accurately **Recommend the most suitable crop** for a particular location or region based on its soil and climatic conditions. The model uses features like nitrogen (N), phosphorous (P), potassium (K), temperature, humidity, pH level, and rainfall to make predictions.

The project also includes a user-friendly web interface developed using Streamlit, where users can input values and get real-time crop suggestions. This system is beneficial for farmers, agricultural researchers, and organizations promoting smart farming techniques.

Dataset Used

The dataset for this project was taken from **Kaggle's Crop Recommendation Dataset**. It contains over **2200 records** with the following features:

- **N**: Nitrogen content in the soil
- **P**: Phosphorous content in the soil
- **K**: Potassium content in the soil
- **Temperature**: Average temperature in °C
- **Humidity**: Relative humidity in percentage
- **pH**: pH value of the soil
- **Rainfall**: Average rainfall in mm
- **Label**: The recommended crop (target variable)

The dataset was clean and well-balanced, with **no missing values**.

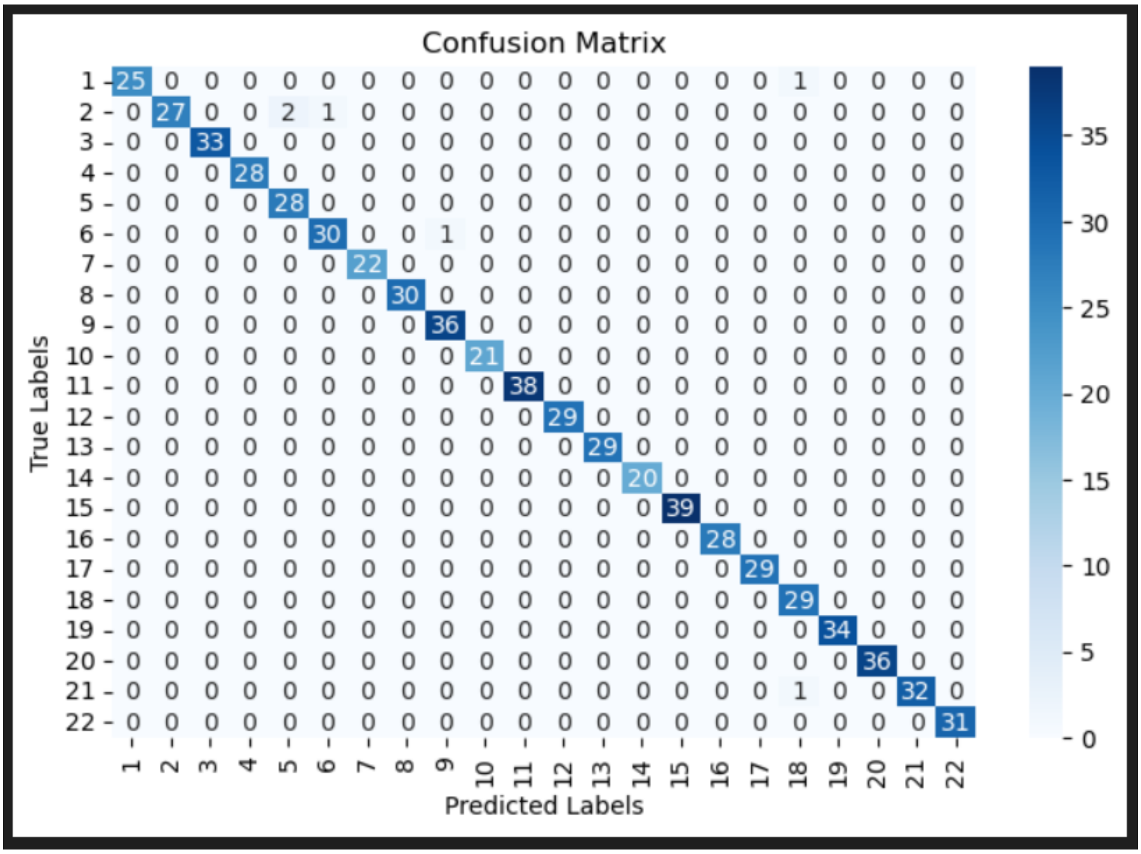
Model

- The **Random Forest Classifier** algorithm was chosen for this classification task.
- This model is highly effective for multi-class classification problems and provides robust performance on structured data.
- It works by building multiple decision trees and combining their results to get better accuracy and reduce overfitting.

Performance Metrics

The model was trained using 80% of the data and tested on the remaining 20%. The following metrics were used to evaluate performance:

- **Accuracy**: 99%
- **F1-Score**: 0.99
- **Confusion Matrix**: Minimal misclassifications, showing excellent generalization



The model performs exceptionally well across all classes, it is due to the well-distributed dataset and the power of ensemble learning.

Challenges

While working on this project, several challenges were encountered:

1. **Multi-class Classification**: Since the target variable consists of more than 20 crop categories, it was essential to ensure the model did not favor a few classes over others.
2. **Model Selection**: Initial experiments with Logistic Regression and SVM resulted in lower accuracy. Random Forest provided better results with less preprocessing.
3. **Feature Importance**: Understanding which features most influenced the prediction was key to trusting the model's output.

Learnings

Through this project, we gained a practical understanding of applying machine learning to real-world agricultural problems. Some key takeaways include:

- Improved skills in **data preprocessing**, **exploratory data analysis**, and **feature selection**.
- Learned how to build and deploy models using **Scikit-learn** and serve them using **Streamlit**.
- Realized the potential of AI and ML in assisting farmers to make **data-driven decisions**.
- Understood how **Random Forest** works internally and its advantages over other models in classification problems.