Crop Recommendation System Using Machine Learning

Objective

To build a machine learning-based system that recommends the most suitable crop to grow based on environmental and soil conditions such as nitrogen, phosphorus, potassium, temperature, humidity, pH, and rainfall. This system is intended to assist farmers in making data-driven decisions to optimize agricultural yield.

Dataset Used

Source: Publicly available on Kaggle or similar platforms

Features (Inputs):

- N (Nitrogen content in soil)
- P (Phosphorus content in soil)
- K (Potassium content in soil)
- Temperature (in Celsius)
- Humidity (in %)
- pH (acidity of soil)
- Rainfall (in mm)

Target (Output):

- Crop label (e.g., rice, maize, lentil, banana, etc.)

Preprocessing Performed:

- Categorical label encoding of target variable (crop names to numbers)
- Feature scaling using StandardScaler
- Train-Test split

Data Preprocessing Steps

- **Handling Missing Values**
    - The dataset was checked for missing or null values.
    - In this project, the dataset did not contain missing values, so no imputation was required.
- **Label Encoding for Target Column**
    - The label column representing crop names (e.g., rice, maize, mango) was of object type.
    - These were converted to integer values using a custom dictionary-based mapping (e.g., 'rice': 1, 'maize': 2, etc.) to make the target suitable for classification models.

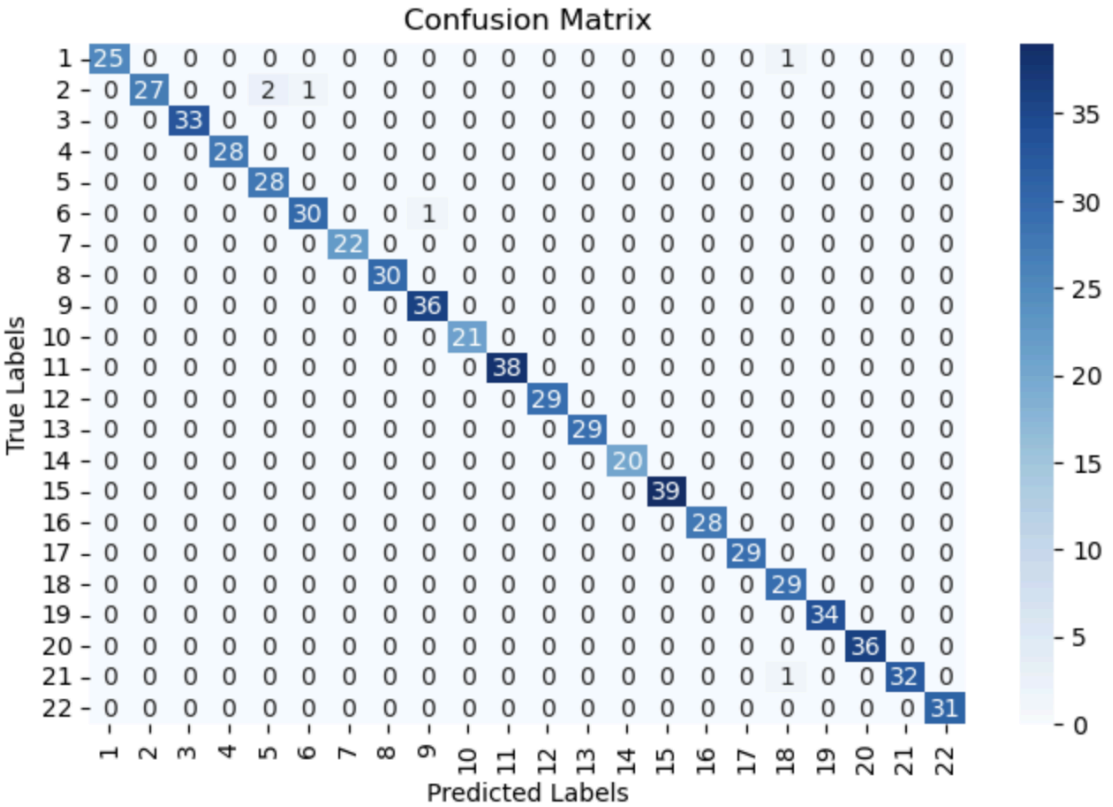
- **Feature Scaling**
    - Since the feature values (N, P, K, temperature, humidity, ph, rainfall) varied in range and units, StandardScaler was used.
    - This normalization helped improve model convergence and performance, especially for models like SVM.
- **Train-Test Split**
    - The dataset was split into training (70%) and testing (30%) sets using train_test_split.
    - This ensured proper validation of the model on unseen data, reducing the risk of overfitting.

Model

- Support Vector Machine (SVM)
- Random Forest Classifier

Model Training:

- Random Forest Classifier achieved the highest accuracy among all models.
- Hyperparameter tuning was optionally applied for better performance.
- Accuracy (Random Forest): 99%
- Very high precision and recall across all crop labels.
- Confusion matrix showed very few misclassifications, indicating a strong model.



Challenges

- Ensuring that the model did not overfit due to high accuracy
- Addressing the imbalance in class distribution (if any)
- Interpreting model results and ensuring correct encoding/decoding of crop labels

Learnings

1. Importance of preprocessing including scaling and encoding
2. Random Forest performs well on multi-class classification with tabular data
3. Confusion matrix and classification report are better indicators than accuracy alone
4. Visualization using Seaborn helped interpret the results clearly

Conclusion:

The crop recommendation model built using Random Forest Classifier is highly accurate and reliable. It can be deployed in a web-based or mobile application to assist farmers and agriculture consultants in crop planning.