Title of the Project: Crop Recommendation System using Machine Learning

Team Member: Manthan Bhegade

Name of Guide: Dr. Sachin Jadhav

Introduction:

Agriculture is one of the essential sectors that influence food security and the economy. Farmers usually do not know how to choose a suitable crop for their fields with the changing weather conditions, the health of their soil, and the uncertainty of weather patterns. Crop selection with traditional approaches depends on experience and intuition, but this might not always be accurate.

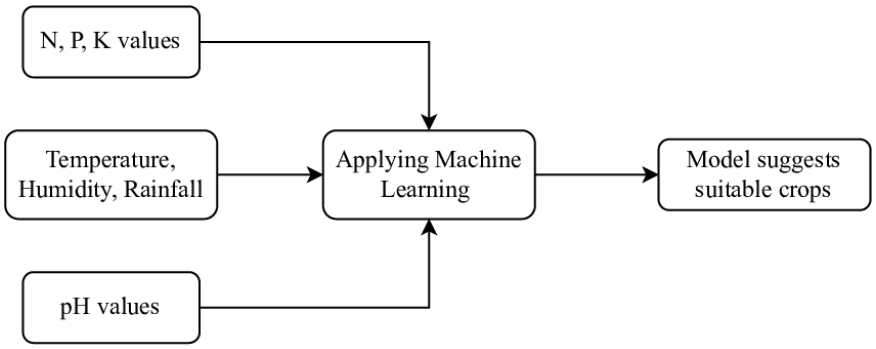
The envisioned Crop Recommendation System utilizes machine learning to examine influential environmental parameters such as soil nutrients (N, P, K), temperature, humidity, pH, and precipitation to make evidence-based crop suggestions. The system helps farmers make better decisions, maximize productivity, and boost profitability.

Proposed System Architecture:

The system is organized into five major phases:

1. **Data Collection & Preprocessing:**
    - Gather soil and climate data, such as Nitrogen (N), Phosphorus (P), Potassium (K), temperature, humidity, pH, and rainfall.
    - Clean and normalize the data to improve model accuracy.
2. **Feature Selection & Engineering:**
    - Choose relevant features that have a strong effect on crop growth.
    - Use data transformation methods to enhance model learning.
3. **Machine Learning Model Training:**
    - Train various machine learning models like Random Forest, Decision Tree, SVM, XGBoost, Logistic Regression and KNN.
    - Compare the performance of the models to choose the best one for recommendations.
4. **Crop Prediction & Recommendation:**
    - Based on input parameters, the system predicts the most appropriate crop.
    - Displays recommendations through a web-based interface using Flask.

5. **User Interface & Deployment:**
    - The system offers an easy-to-use web application where users can enter soil and climate parameters.
    - Outputs suggested crops with information about soil health and environmental conditions.



Methodology:

1. **Dataset Collection:**
    - The features of the dataset used include N, P, K, temperature, humidity, pH, and rainfall, with crop type as the target variable.
    - The data is derived from publicly accessible agricultural datasets and government research reports.
2. **Why This Dataset?**
    - The reason for choosing this dataset is that it offers a wide variety of soil and climate conditions influencing crop yield.
    - The dataset consists of historical crop data across various regions to ensure improved model generalization.
    - Using this dataset enables the system to provide accurate recommendations through learning from actual agricultural data.
3. **Data Preprocessing & Feature Engineering:**
    - Missing value handling, scaling, and transformation.
4. **Model Selection & Training:**
    - Testing Random Forest, Decision Tree, SVM, XGBoost, Logistic Regression and KNN to identify the top-performing model.
    - Model comparison based on accuracy, precision, recall, and F1-score.
5. **Web Application Development:**
    - Using Flask for backend processing.
    - Creating an easy-to-use interface using HTML, CSS, and JavaScript.
6. **Performance Evaluation & Optimization:**

- Fine-tuning model parameters for better accuracy.
- Testing with real-world agricultural data.

Machine Learning Algorithm Used in this Project:

1. **Random Forest (Best Model):**
   - **Accuracy:** 99.32%
   - **Uses:** Best for complex datasets, provides stable and reliable predictions.
2. **Decision Tree:**
   - **Accuracy:** 98.64%
   - **Uses:** Simple to understand, good for quick decision-making.
3. **Support Vector Machine (SVM):**
   - **Accuracy:** 96.82%
   - **Uses:** Works well for small datasets with clear patterns.
4. **K-Nearest Neighbors (KNN):**
   - **Accuracy:** 95.68%
   - **Uses:** Best for small datasets but slower with large data.
5. **XGBoost:**
   - **Accuracy:** 98.64%
   - **Uses:** High performance, suitable for large datasets**.**
6. **Logistic Regression:**
   - **Accuracy:** 96.36%
   - **Uses:** Good for binary classification, easy to interpret.

Random Forest is selected as the final model since it gives the highest accuracy and supports various types of data.

Dataset Used in this Project:

- **Attributes:**
  - **N (Nitrogen):** Necessary for plant growth and leaf development.
  - **P (Phosphorus):** Necessary for root growth and seed formation.
  - **K (Potassium):** Helps in plant metabolism and immunity against diseases.
  - **Temperature (°C):** Determines seed germination and crop yield.
  - **Humidity (%):** Regulates the rate of plant transpiration.
  - **pH Level:** Decides the soil acidity or alkalinity and impacts the availability of nutrients.
  - **Rainfall (mm):** For irrigation and hydrating plants.

    - **Crop Type (Label – Target Variable):** Represents the recommended crop.
- **Source:**
    - Government research center public agricultural datasets.
    - Public domain agricultural datasets on platforms such as Kaggle.
    - From GitHub - https://github.com/611noorsaeed/Crop-Recommendation-System-Using-Machine-Learning/blob/main/Crop_recommendation.csv.

Limitations:

- Accuracy relies on the dataset used to train and the variety and quality of that dataset.
- Crop suggestions can differ depending on actual climate variations in real-time.
- Internet connectivity is necessary for web-based access.

Conclusion:

The Crop Recommendation System offers an AI-driven solution to help farmers choose the most appropriate crop according to soil and climatic conditions. With machine learning, the system improves decision-making, optimizes productivity, and ensures sustainable agriculture. The solution can be instrumental in minimizing crop failure, optimizing resource use, and enhancing agricultural profitability.