# Crop Recommendation System Based on Soil and Weather Conditions

## Project Overview

The **Crop Recommendation System** is designed to suggest the most suitable crop for farming based on various soil and weather parameters. This system leverages machine learning models to predict which crop will yield the best results given the conditions of soil and weather, such as nutrient content and environmental factors.

## Dataset

The dataset used for this project includes the following features: (source:Kaggle)

- **N (Nitrogen)**: Nitrogen content in the soil.
- **P (Phosphorus)**: Phosphorus content in the soil.
- **K (Potassium)**: Potassium content in the soil.
- **Temperature**: The average temperature of the region.
- **Humidity**: The humidity levels in the air.
- **Rainfall**: The amount of rainfall in mm.
- **pH**: The pH value of the soil.
- **Recommended Crop**: The target label representing the most suitable crop for the given conditions.

The dataset has 2200 rows and 8 columns, which were used to train and evaluate machine learning models.

## Objective

The objective of the project is to build a model that accurately predicts the most suitable crop based on soil and weather conditions. Multiple machine learning algorithms were implemented and compared to find the best-performing model.

## Models Implemented

The following models were trained and evaluated:

1. **Logistic Regression**
2. **Decision Tree**
3. **Random Forest**
4. **Support Vector Classifier (SVC)**

### Model Comparison

After training and testing the models, the **Random Forest Classifier** yielded the highest accuracy. The model comparison based on accuracy is as follows:

- **Logistic Regression**: [0.9181818181818182]
- **Decision Tree**: [0.9863636363636363]
- **Random Forest**: [**0.9931818181818182**] (highest)
- **SVC**: [0.9681818181818181]

**Models after hyperparameter tuning**:

- **Logistic Regression**: [0.9704545454545455]
- **Decision Tree**: [0.9863636363636363]
- **Random Forest**: [**0.9931818181818182**] (highest)
- **SVC**: [0.9795454545454545]

# Best Performing Model: Random Forest Classifier

The **Random Forest Classifier** was selected as the best model due to its superior performance in terms of accuracy. It is an ensemble learning method that combines multiple decision trees to enhance prediction accuracy and reduce overfitting.

# Steps Involved

1. **Data Preprocessing**:
    - Scaled the data to ensure uniformity across features.
    - Split the dataset into training and test sets.
2. **Model Training**:
    - Trained four different machine learning models on the dataset.
    - Tuned hyperparameters to optimize model performance.
3. **Model Evaluation**:
    - Evaluated the models using metrics such as accuracy, precision, recall, and F1-score.
    - Compared the models to determine the best one for crop recommendation.
4. **Final Model**:
    - Chose the Random Forest model as it had the highest accuracy.
    - Saved the trained model for future predictions.

# Technologies Used

- **Python**
- **Pandas**: For data manipulation and preprocessing.
- **Scikit-learn**: For implementing machine learning models.
- **Matplotlib**, **Seaborn**: For data visualization.
- **Pycharm**: For development and experimentation.

# Future Enhancements

- Acquiring data for more crops and their environmental conditions.
- Integration of real-time weather and soil sensor data for dynamic crop recommendations.
- Incorporation of additional soil nutrients and climate factors to improve prediction accuracy.
- Deployment of the model as a web or mobile application for user-friendly access.