Fertilizer Recommendation System Based on Soil & Crop Type using Random Forest

1. Introduction

In modern agriculture, choosing the right fertilizer is crucial for improving crop yield, maintaining soil health, and ensuring sustainable farming. However, farmers often lack the scientific tools needed to make these decisions. This project aims to build a machine learning model that recommends the most suitable fertilizer based on soil type, crop type, and various environmental and nutrient parameters using a Random Forest Classifier.

2. Problem Definition

This is a classification problem where the goal is to predict the best fertilizer category for a given set of features. The output variable (target) is categorical (e.g., Urea, DAP, 14-35-14), making it suitable for RandomForestClassifier.

3. Data Collection

The dataset was created synthetically to reflect realistic values of environmental parameters like temperature, humidity, moisture, soil type, crop type, and soil nutrient levels (Nitrogen, Phosphorus, Potassium). It consists of 100 rows and 9 columns.

4. Exploratory Data Analysis (EDA)

EDA was conducted to understand the distribution and relationship of features. Count plots were used to examine fertilizer, soil, and crop type distributions. Boxplots and heatmaps revealed how various nutrients and environmental factors varied across crops and fertilizers.

- Nitrogen and Soil Type are dominant features in fertilizer decision-making.

- Crop Type contributes significantly to the prediction.

- Features like Humidity and Temperature had lesser influence.

5. Data Preprocessing

Missing values were handled using mean imputation. Categorical features (Soil_Type, Crop_Type, Fertilizer) were encoded using LabelEncoder to convert them into numeric form. No irrelevant features were present, so all features were retained.

6. Feature and Target Definition

Features: Temperature, Humidity, Moisture, Soil_Type, Crop_Type, Nitrogen, Phosphorus, Potassium

Target: Fertilizer

7. Train-Test Split

The dataset was split into training and testing sets using an 80-20 ratio with stratified sampling to ensure even class distribution.

8. Model Building

A Random Forest Classifier was trained on the processed data. The model leverages ensemble learning by constructing multiple decision trees and averaging their results to improve accuracy and prevent overfitting.

9. Predictions and Evaluation

The model predictions were evaluated using accuracy, classification report, and confusion matrix. The model performed well, with high accuracy and precision.

10. Feature Importance

Feature importance was extracted from the Random Forest model to understand which features influenced the decision most. The top contributing features were Nitrogen, Soil_Type, and Crop_Type.

11. Hyperparameter Tuning

GridSearchCV was used to optimize the model by searching over different hyperparameter combinations such as n_estimators, max_depth, and max_features. The best parameters were selected based on cross-validation performance.

12. Model Saving

The trained model was saved using joblib, making it ready for future use or deployment in real-world applications.

13. Conclusion

This project successfully demonstrates the use of machine learning for fertilizer recommendation. By leveraging environmental and crop-specific features, the model can guide farmers toward optimal fertilizer usage, promoting efficiency and sustainability in agriculture.