CROP RECOMMENDATION SYSTEM

A Machine Learning Project Report

Submitted By:
TANISH CHAUDHARI (1AUA23BCS186)
YUG LIMBACHIYA (1AUA23BCS211)

Department of Computer Science and Engineering
Adani University

Course: Machine Learning Essentials

Faculty Supervisor: Madhvi Ramrakhiyani, Neha Jagwani

Date of Submission: November 12, 2025

Abstract

Agriculture is one of the most essential sectors for food production and economic stability, yet farmers often face challenges in selecting the right crop due to varying soil nutrients, climatic conditions, and rainfall patterns. This project presents a Machine Learning-based Crop Recommendation System that predicts the most suitable crop for cultivation based on soil nutrient values (N, P, K), temperature, humidity, rainfall, and pH. The dataset used consists of real agricultural records and contains essential soil and climate features linked to different crop types.

Three machine learning models — Logistic Regression, Random Forest, and XGBoost — were trained, tested, and compared. XGBoost achieved the highest accuracy of 98.40%, slightly outperforming Random Forest with 98.36%, while Logistic Regression achieved 95.60%. Feature importance and SHAP (Explainable AI) were used to understand how each feature influences the final prediction. Additionally, a fertilizer suggestion module was developed to recommend adjustments in nitrogen, phosphorus, and potassium levels for soil improvement.

The system can be deployed using Flask or Streamlit as a web interface and can be further integrated with cloud platforms like AWS or GCP. This solution helps farmers make data-driven decisions, reduce crop failure risk, and promote sustainable farming practices.

Introduction

Agriculture supports the livelihood of millions of people, especially in developing countries. However, farmers often rely on traditional knowledge instead of scientific methods to select crops, leading to inconsistent yields and economic losses. Due to changing weather conditions and soil fertility levels, recommending the right crop has become a major challenge.

Incorrect crop selection results in poor productivity, excessive fertilizer use, soil damage, and financial loss. An intelligent system that analyzes soil nutrients and environmental conditions can help farmers choose the best crop scientifically rather than depending on guesswork.

- To build a machine learning model that accurately recommends suitable crops.
- To analyze soil and climate data using preprocessing and visualization techniques.
- To compare multiple models (Logistic Regression, Random Forest, XGBoost).
- To provide fertilizer suggestions based on nutrient deficiencies.
- To design a simple user interface for prediction.

Literature Review / Related Work

1. **Crop prediction using Random Forest and Naïve Bayes** achieved up to 95% accuracy but lacked interpretability.
2. **Deep learning models like ANN and LSTM** were used for yield prediction but required large datasets and computing power.
3. **IoT-based systems** monitor soil moisture and weather but do not provide crop recommendations.
4. **Clustering-based soil classification models** grouped soil types but did not predict crops.

Research Gaps

- Limited feature sets (mostly only N, P, K, temperature, rainfall).
- Lack of interpretability using SHAP or feature importance.
- No fertilizer recommendation modules.

Contribution of This Project

- Uses multiple models and compares results.
- Uses feature importance and SHAP for explainability.
- Provides fertilizer recommendations.
- Deployable using Flask/Streamlit.

Methodology

5.1 Dataset Description

- Total records: ~2200 rows
- Features: Nitrogen, Phosphorus, Potassium, Temperature, Humidity, pH, Rainfall
- Target: Crop (22+ types like rice, wheat, maize, etc.)

5.2 Data Cleaning

- No missing values
- Checked for duplicates and outliers
- Encoded target labels
- Applied scaling for Logistic Regression

5.3 Exploratory Data Analysis (EDA)

- Correlation heatmap

- Distribution plots for N, P, K
- Crop frequency bar chart
- Scatter plots of feature relationships

5.4 Feature Engineering

- Created nutrient imbalance score
- StandardScaler for scaled models
- Feature importance from Random Forest and XGBoost

5.5 Model Building

Model	Reason
Logistic Regression	Baseline model
Random Forest	Handles non-linearity
XGBoost	High accuracy, boosting technique

Data split: 80% train / 20% test

5.6 Hyperparameter Tuning

- Random Forest: n_estimators, max_depth
- XGBoost: learning_rate, subsample, max_depth

5.7 Model Evaluation

Model	Train Accuracy	Test Accuracy	F1-Score (Weighted)
Logistic Regression	0.9527	0.9363	0.9364
Random Forest	0.9980	0.9931	0.9931
XGBoost	0.9905	0.9840	0.9840

5.8 Deployment

- Implemented using **Streamlit GUI**
- Input: N, P, K, Temperature, Humidity, pH, Rainfall
- Output: Recommended Crop

5.9 Cloud Integration

- Streamlit

5.10 GUI Interface

- Input fields for soil data
- Predict button
- Output displayed in UI with crop name.

Results and Discussion

Random Forest performed best with accuracy above 98%.

- Confusion Matrix showed minimal misclassifications.
- Accuracy bar graph showed Random Forest >XGBoost > Logistic Regression.
- Random Forest is best for accuracy
- Random Forest is best for interpretability
- System is practical and scalable

Nitrogen and Phosphorus had the highest influence on predictions, followed by rainfall and temperature.

- High nitrogen pushed predictions towards rice and maize.
- Low rainfall favored pulses like lentil and mothbeans.
- pH was important for coconut and pomegranate.

Conclusion and Future Work

This system predicts suitable crops with high accuracy using soil and climatic data. It supports farmers in decision-making and promotes sustainable agriculture.

Limitations

- Dataset is not real-time
- No integration with IoT sensors
- Market price not considered

Scope for improvement

- Real-time weather API and IoT sensors
- Mobile app with voice assistant in local languages
- Fertilizer quantity calculation

- Profit-based crop prediction using mandi prices
- Deployment using AWS/Google Cloud

References

Smart Farming Data 2024 (SF24)

Available: https://www.kaggle.com/datasets/datasetengineer/smart-farming-data-2024-sf24

Tools and Lused:
1) Python
2) Numpy
3) Pandas
4) Seaborn
5) Matplotlib
6) Sklearn (scikit learn)
7) XGboost
8) Random
9) Streamlit
10) Warnings
11) RE
12) Joblib
13) Requests