PROJECT ABSTRACT

Crop Recommendation System using Machine Learning

Agriculture employs half of India's workforce, yet which crop to sow is decided by habit or last year's prices.

01 Title

Crop Recommendation System using Machine Learning

02 Introduction / Background

Agriculture employs half of India's workforce, yet which crop to sow is decided by habit or last year's prices. A mismatched crop in unsuitable soil or rainfall can destroy a season's income. Precision-agriculture tools remain the preserve of agribusinesses, leaving smallholders without guidance. AgriGuide meets that need: a machine-learning recommender mapping seven conditions to suitable crops.

03 Problem Statement

Farmers lack a way to choose crops for soil and weather from data. Advisories are generic, extension services are stretched thin, and commercial tools are priced beyond smallholders' reach. Wrong choices cut yields and waste fertilizer, causing distress. This project replaces guesswork with explainable recommendations in a web app.

04 Objectives

- Build a classifier mapping seven soil/climate inputs to crops at ~99% accuracy.
- Benchmark Random Forest, XGBoost and Logistic Regression rigorously.
- Deliver top-three recommendations with suitability scores via the AgriGuide app.
- Attach fertilizer and soil-health tips to every recommended crop.
- Ship a retrainable pipeline fully defensible in a viva.

05 Proposed System / Working

AgriGuide trains on the crop-recommendation dataset: roughly 2,200 samples across 22 crops (rice, wheat, maize, cotton, sugarcane, pulses). Seven features are used: nitrogen, phosphorus, potassium, temperature, humidity, soil pH and rainfall. The pipeline covers cleaning, stratified 80/20 splitting and five-fold cross-validation. Random Forest, XGBoost and Logistic Regression are benchmarked with confusion matrices and F1-scores. Tree ensembles suit small tabular data: they beat neural networks, train in seconds on a laptop, and give interpretable feature importance.

Serialized models are served through the Flask app.

The models reach ~99% accuracy on the test set with matching F1-scores. In the demo, seven inputs — nitrogen-rich soil with high rainfall — return rice, jute and sugarcane with suitability percentages and sowing guidance. Feature-importance plots show rainfall, potassium and nitrogen dominating, matching agronomic intuition — viva material.

06 Technologies / Components Used

Python 3 · **pandas** · **NumPy** · **scikit-learn** · **XGBoost** · **Seaborn** · **Flask** · **HTML/CSS frontend** · **Jupyter notebooks** · **pickle/joblib**

07 Key Features

- Crop prediction from seven inputs: N, P, K, temperature, humidity, pH, rainfall.
- Top-three crops ranked by suitability score, not one answer.
- Model comparison with accuracy, F1-score and confusion matrices.
- Feature importance showing which inputs drive decisions.
- Fertilizer and soil-health tips for every recommended crop.
- Seasonal filtering of recommendations (Kharif / Rabi).
- Retrainable notebook pipeline for regional datasets.

08 Applications / Use Cases

- Direct value for smallholder farmers choosing crops, extension services needing scalable advisories, agri-startups building decision-support tools, and students studying soil–crop relationships.
- The pipeline also templates other tabular recommenders, such as fertilizer-dose advice.

10 Conclusion

Machine learning delivers social value on ordinary hardware with a modest dataset. Benchmarking plus a usable farmer-facing app turns soil data into deployed decision support — a viva-ready story from data to product.