Weather-Based Crop Recommendation System

Mohith Banka

September 18, 2025

1 Project Overview

This project develops a machine learning-powered system to recommend optimal crops for farmers based on environmental and soil conditions. By analyzing key factors like soil nutrients (Nitrogen, Phosphorus, Potassium), temperature, humidity, pH, and rainfall, the model predicts the most suitable crop from 22 common varieties (e.g., rice, maize, apple, banana). This promotes sustainable agriculture by maximizing yield and minimizing resource waste. The system leverages Random Forest and LSTM neural networks, with GPU acceleration and a Flask-based RESTful API for real-time predictions.

2 Dataset

- Source: Augmented from ∼2,200 to 10,000 records using Gaussian noise (2% variance) to simulate real-world variations, maintaining balance across 22 crop classes.
- Features (7 numerical): Nitrogen (N), Phosphorus (P), Potassium (K), temperature (°C), humidity (%), pH, rainfall (mm).
- Target: Categorical label (22 crop classes, e.g., 'rice', 'maize', 'coffee').
- Balance: ∼455 samples per class post-augmentation.

3 Methodology and Preprocessing

1. Loading Data: Read augmented CSV using Pandas.
2. Feature-Target Split: Separated features (X) and labels (y).
3. Encoding: Applied LabelEncoder to convert crop labels to integers (0–21).
4. Scaling: Used StandardScaler for feature normalization (zero mean, unit variance).
5. Train-Test Split: 80/20 split (4,000 train, 1,000 test) with stratification.
6. GPU Setup: Utilized NVIDIA GeForce GTX 1650 via PyTorch for efficient computation.

4 Model Training

- Random Forest (RF): Ensemble classifier for interpretable predictions.

- LSTM: Sequential neural network for capturing patterns in environmental data.
- Visualization: Seaborn heatmaps and Matplotlib histograms for feature correlations and crop distributions.

5 Results

The models were evaluated on the test set (1,000 samples). Results are summarized below:

Table 1: Model Performance Metrics

Model	Training Accuracy (%)	Test Accuracy (%)
Random Forest	95.0	92.0
LSTM	93.0	90.0

- Confusion Matrix: Visualized via Seaborn, showing high precision for most classes, with minor confusion between crops with similar environmental needs (e.g., rice and jute).
- Performance Notes: Random Forest outperformed LSTM slightly due to its robustness with tabular data, while LSTM captured sequential patterns effectively.

6 Deployment and Inference

Deployed as a Flask API on localhost:5000:

- Endpoint: POST /predict
- Input: JSON with features (N, P, K, temperature, humidity, pH, rainfall, model type).
- Output: Predicted crop and confidence score.
- Example: Input {N: 90, P: 40, K: 40, temperature: 24.5, humidity: 80, pH: 6.5, rainfall: 200, model: "lstm"} yields {prediction: "rice", confidence: 0.92}.

7 Key Achievements and Impact

- Scalability: Handles 10,000+ records with 2–5x GPU speedup.
- Accuracy: High precision across diverse climates.
- Real-World Use: Integrates with weather APIs for live data; accessible via mobile/web apps.
- Future Work: Add SHAP for explainability, multi-crop ranking, climate change simulations.

8 Source Code

Available at: https://github.com/mohithbanka/Weather-Based-Crop-Recommendation-System