Soil and Crop Recommendation

SAMIE AHMAD

Literature Review

The Future of Digital Agriculture: Technologies and Opportunities
Fountas, Spyros, et al.

Deep Learning in Agriculture: A Survey
Kamilaris, Andreas, and Francesc Prenafeta Boldú

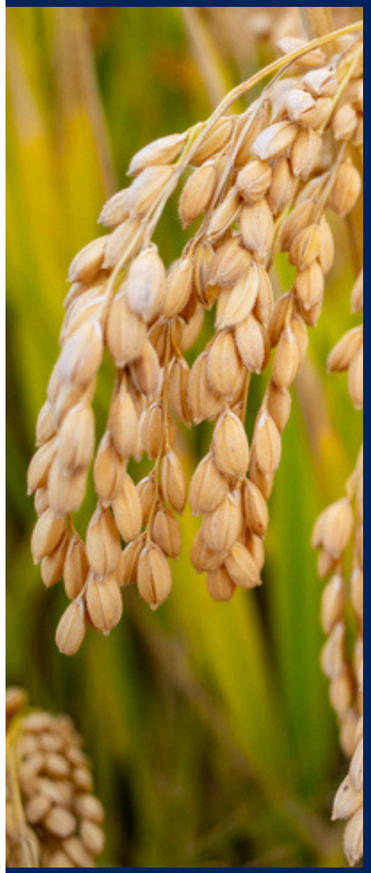
Big Data in Smart Farming – A Review
Wolfert, Sjaak, et al.

Visible, Near Infrared, Mid Infrared or Combined Diffuse Reflectance
Viscarra Rossel, Raphael, et al.

Exploring Machine Learning Models for Soil Nutrient Properties:
Folorunso, Olusegun, et al.

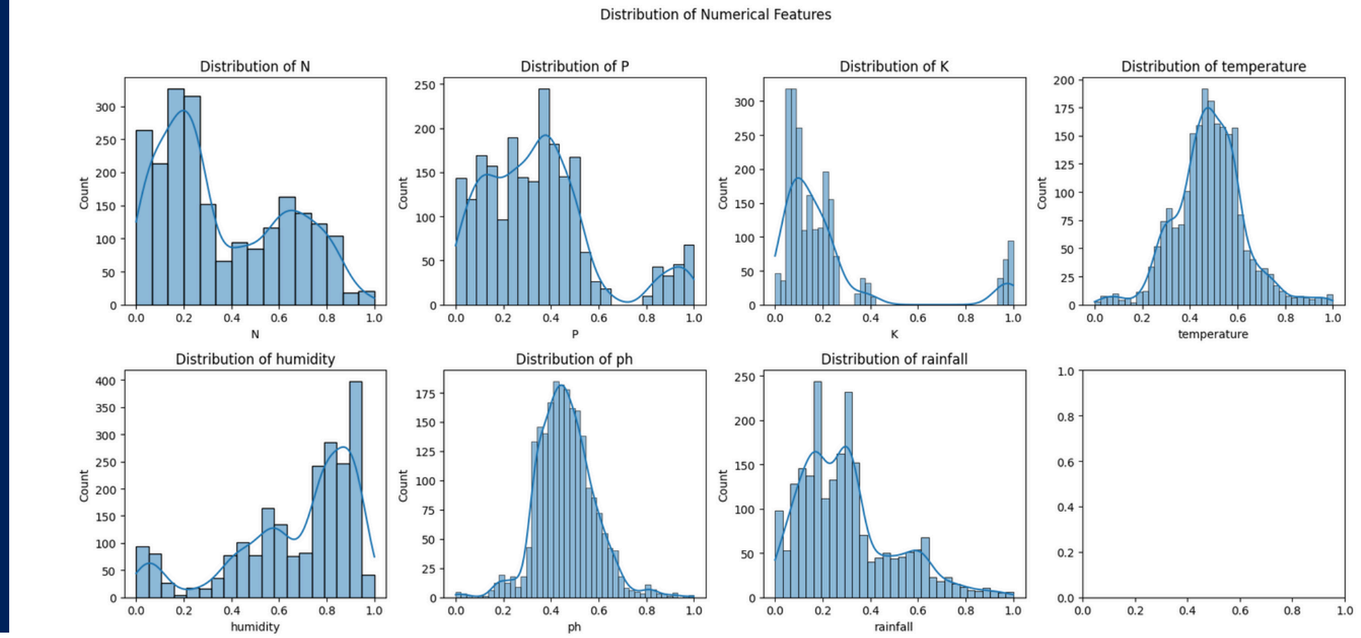
Advancements in Crop Yield Prediction Using Deep Learning Algorithms
Ajina, A., et al.

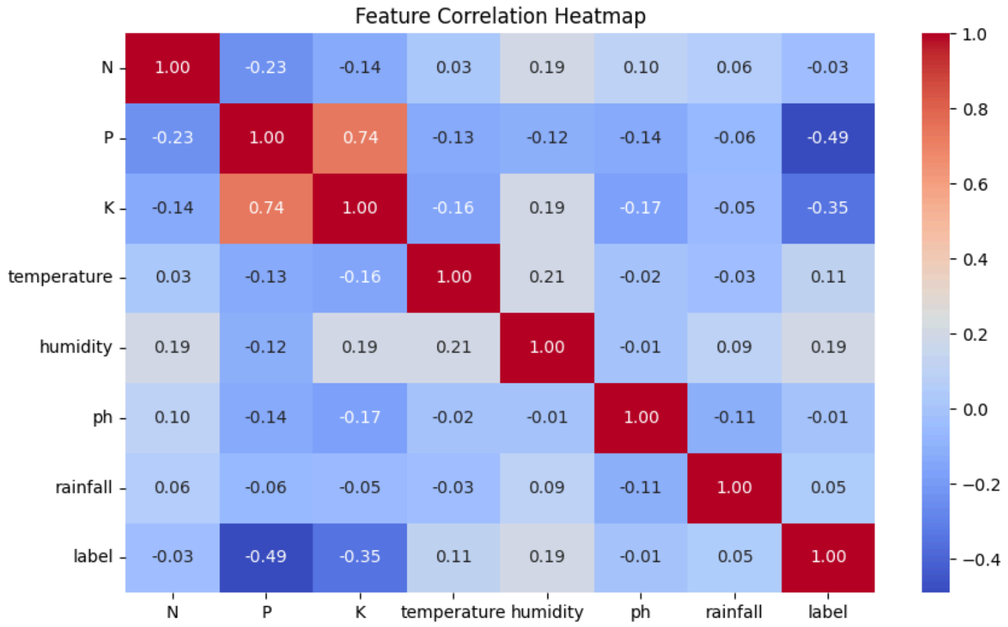
A Comprehensive Review on Automation in Agriculture Using AI
Jha, Kirtan, et al.

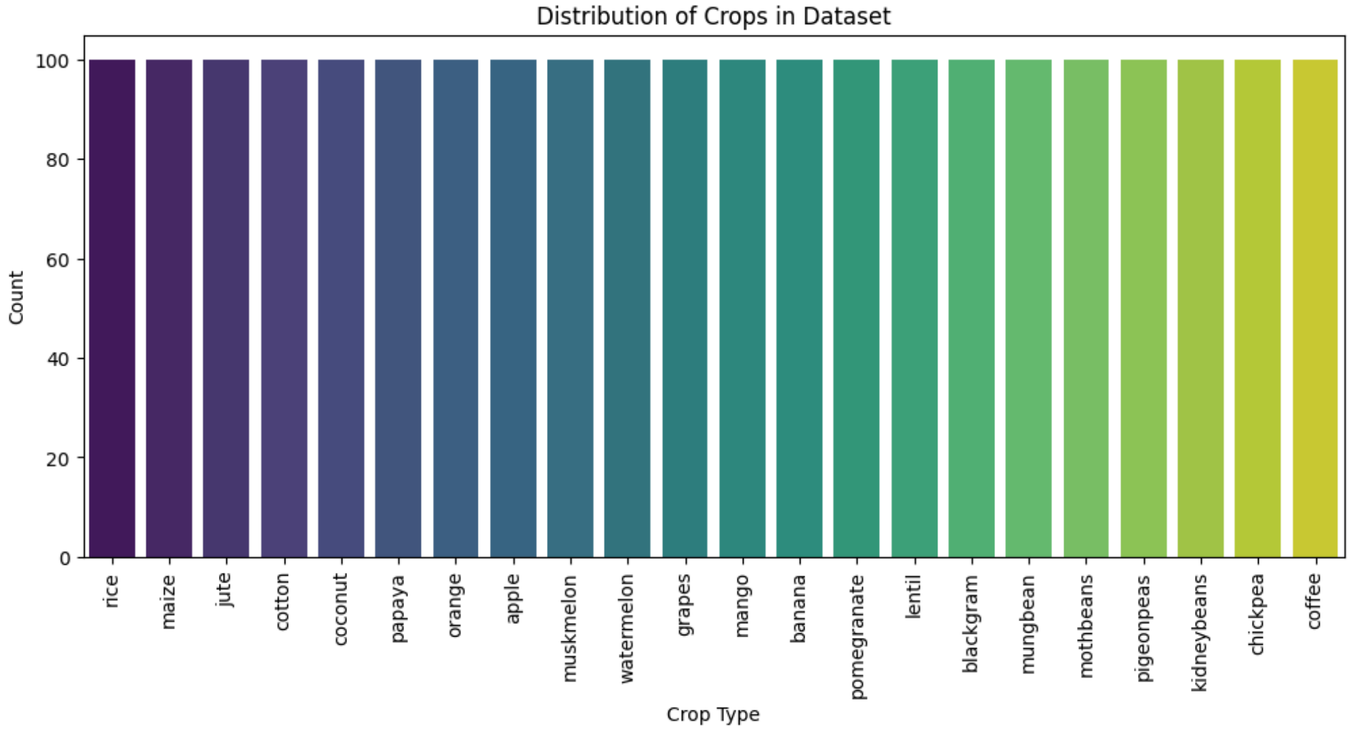


The Dataset

This dataset contains labels for the recoommended crop based on the distribution of minerals in soil, pH and humidity.







Model Implementation

Random Forest Classifier was chosen for its robustness and efficiency

Preprocessing

- Data Cleaning
- Feature Scaling
- Label Encoding
- Train-Test Split (80:20)

Optimized Hyperparameters

- n_estimators = 300
- max_depth = 25
- min_samples_split = 4
- min_samples_leaf = 2

Evaluation

High Model Accuracy

The model achieved 99.32% accuracy, indicating strong predictive performance.

Minimal Misclassification

The confusion matrix shows that most crops were classified correctly.

Feature Importance & Correlations

Soil nutrients (N, P, K), temperature, and humidity played a significant role in crop prediction.

99.3% ACCURACY

93.4% PRECISION

93.3% RECALL

93.3% F1 SCORE

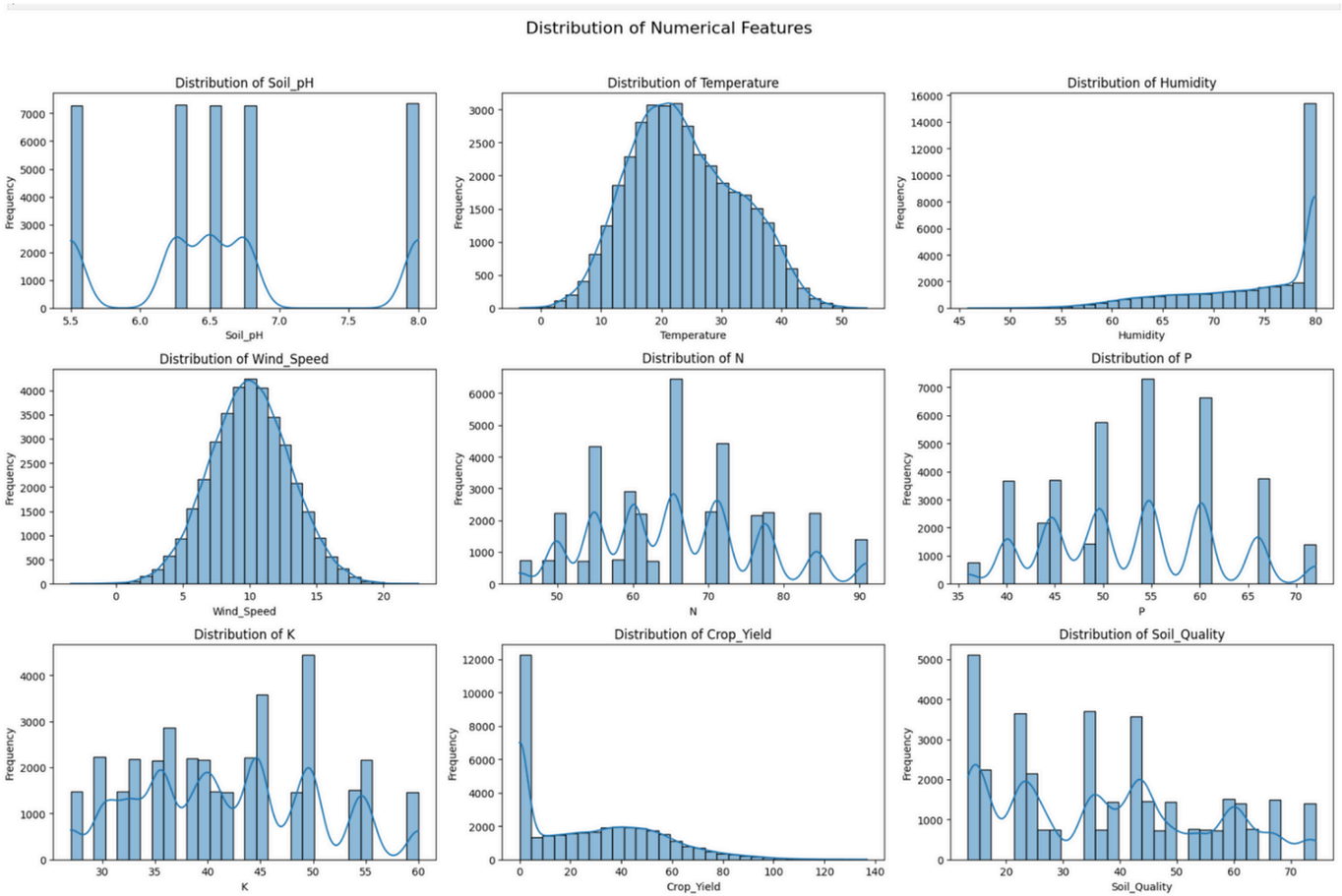
Model Phase 2

New DataSet Review

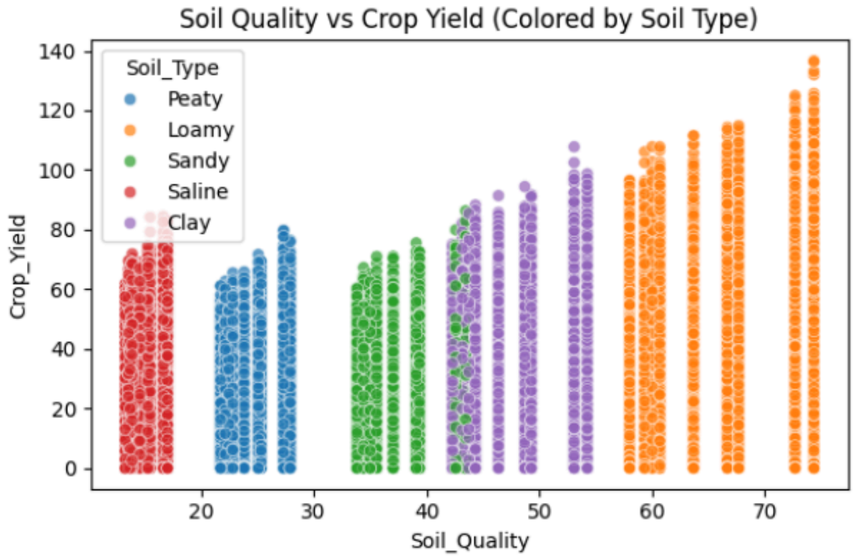
- Entries: 36,520 rows
- Features: Soil pH, Temperature, Humidity, Wind Speed, N, P, K, Soil Quality, Soil type, Crop Type, Crop Yield
- Date Range: 2014 – 2023
- Crop Types: 10
- Soil Types: 5

Exploratory Data Analysis

- **Right-skewed nutrient and yield values.**
- **Soil pH centers around neutral (6.5–7.5).**



Exploratory Data Analysis

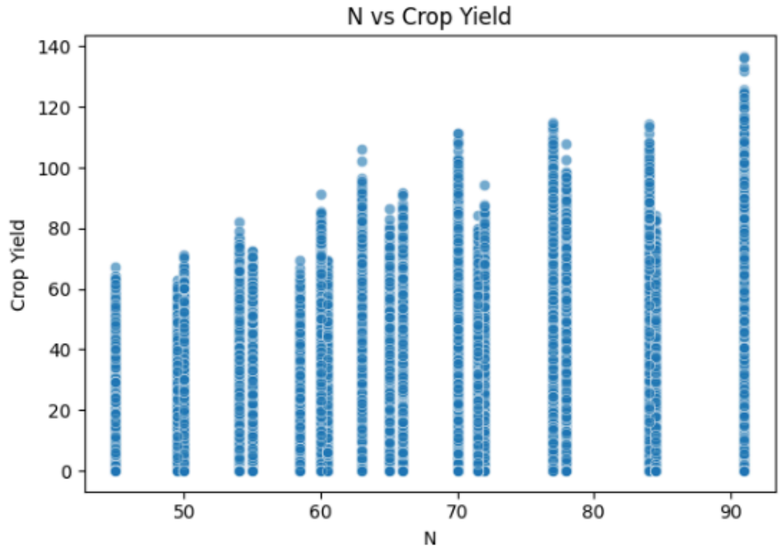


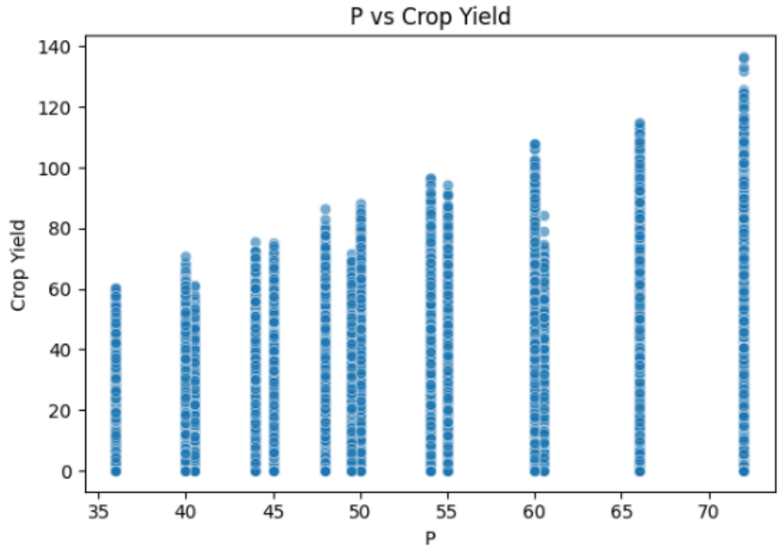
Soil Quality vs. Yield

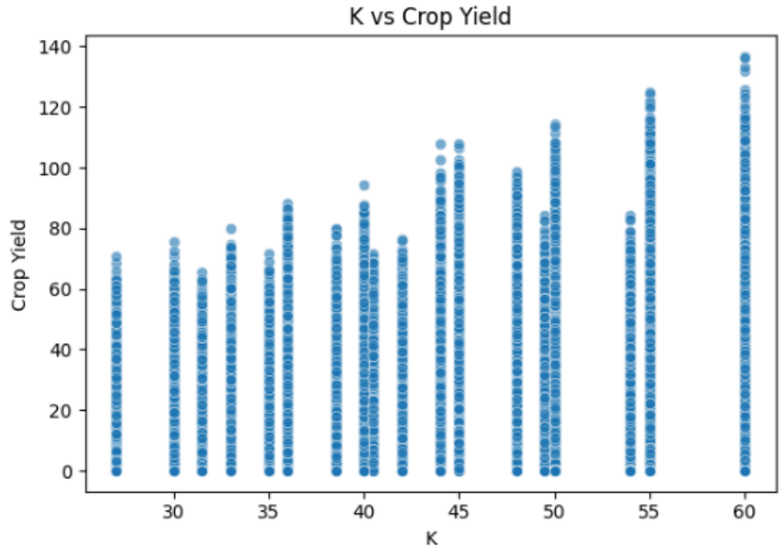
- **Positive trend between soil quality and crop yield.**
- **Soil Type influences this relationship.**

NPK vs Crop Yield

- **Yield increases with nutrients up to a threshold.**
- **After a certain level, yield stabilizes or drops → non-linear effect.**







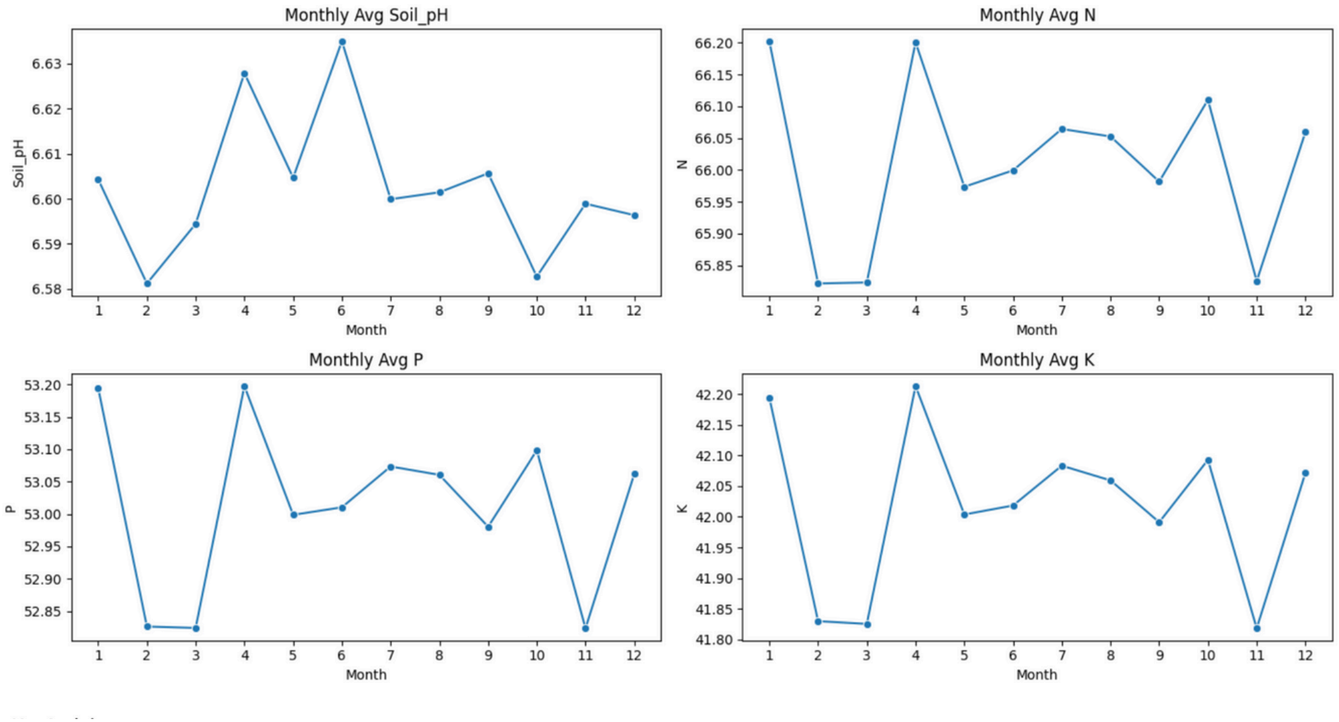
Temporal Analysis

• NoP and K levels be sight seasonait 6.58-6.64 - a neutral range ideal for most crops.

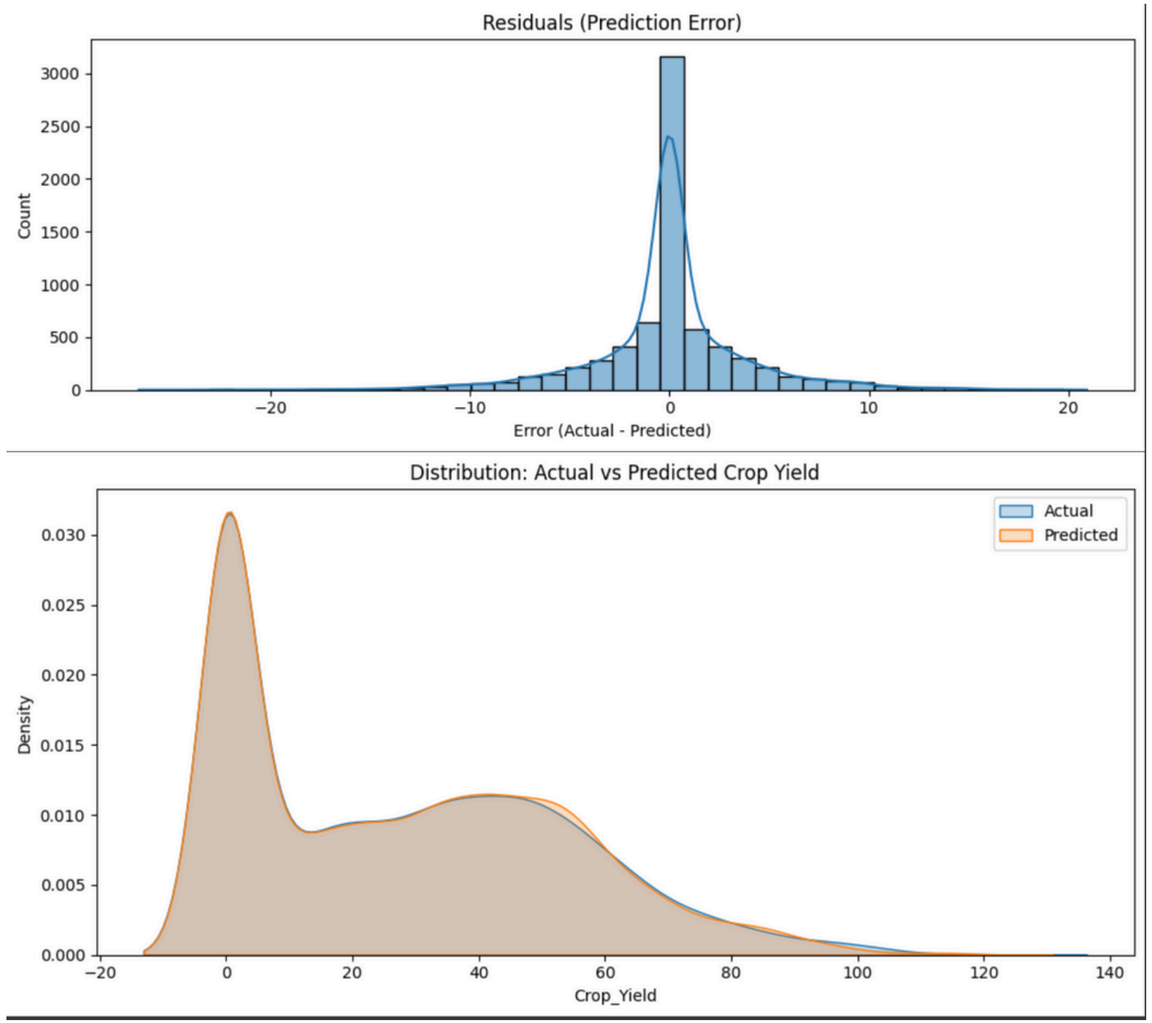
- Nutrient spikes around April-June.
- Dips in February and November, suggesting these months are less fertile.
- These patterns likely reflect fertilizer application cycles or natural soil respiration.

Implications:

- Best months for planting: April to July, when nutrient levels peak.
- Avoid planting in February or November unless fertilizer intervention is planned.
- These trends could be used to simulate fertilization calendars or time-aware crop recommendation systems.



Regression Using XGBoost and Random Forest pipelines



Random Forest -

RMSE: 16.12,

MAE: 2.34,

R^2: 0.9758

XGBoost -

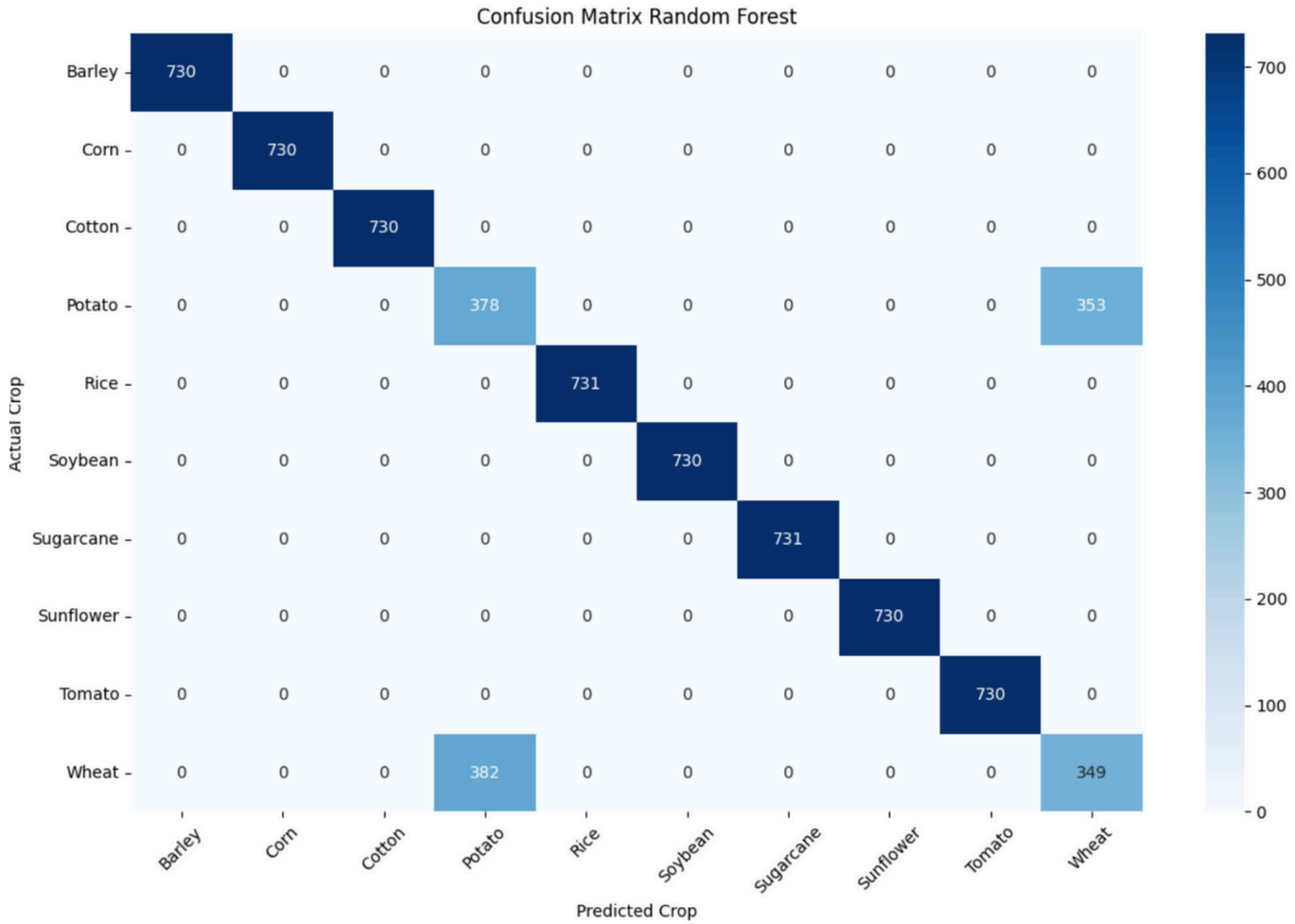
RMSE: 17.21,

MAE: 2.56,

R^2: 0.9742

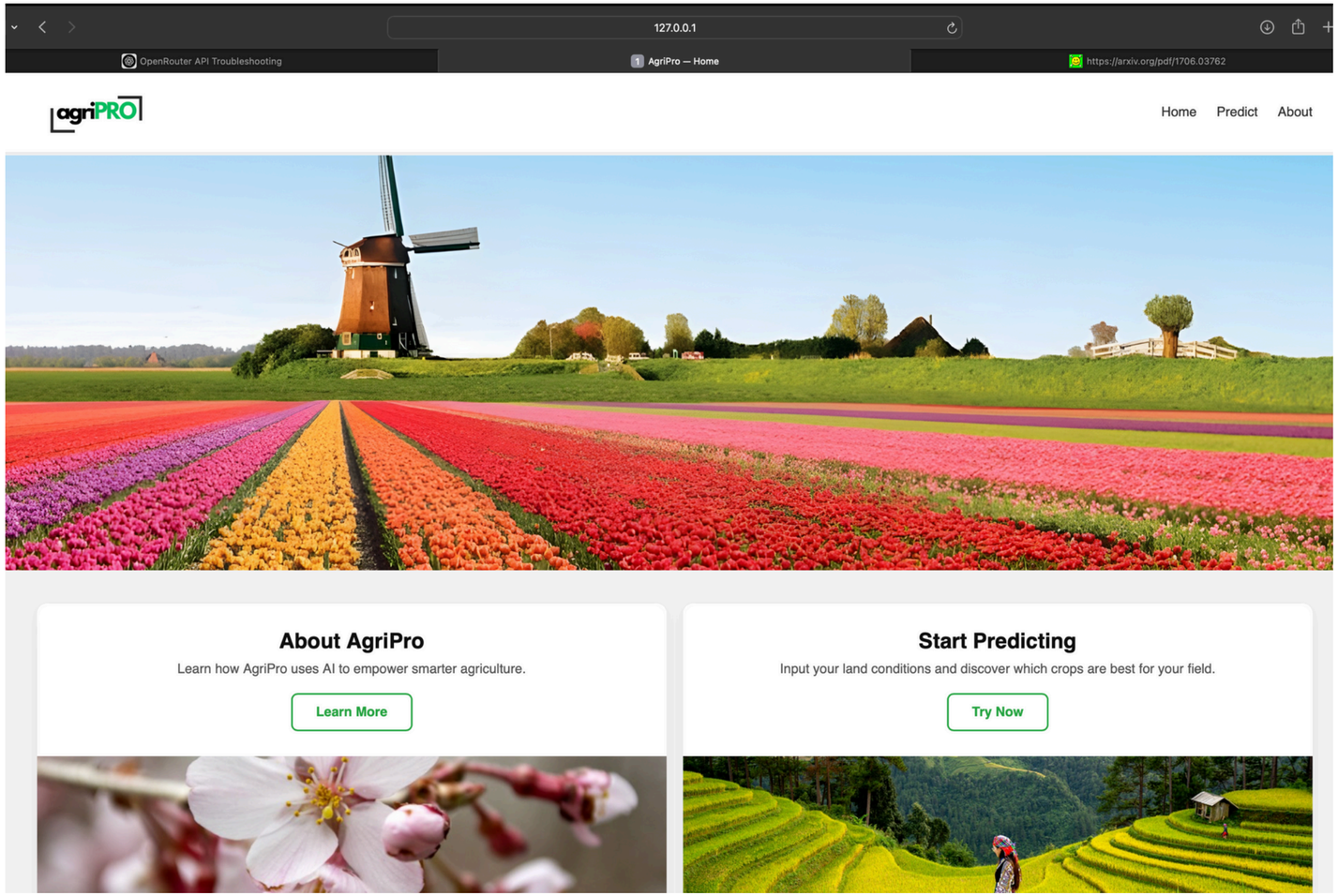
Prediction Model Implementation

- **Model: Random Forest**
- **Accuracy: 89.9%**
- **Top-2/Top-3 Accuracy: 100%**
- **Confusion Matrix: Good prediction for most crops**
- **Struggles with: Potato, Wheat**



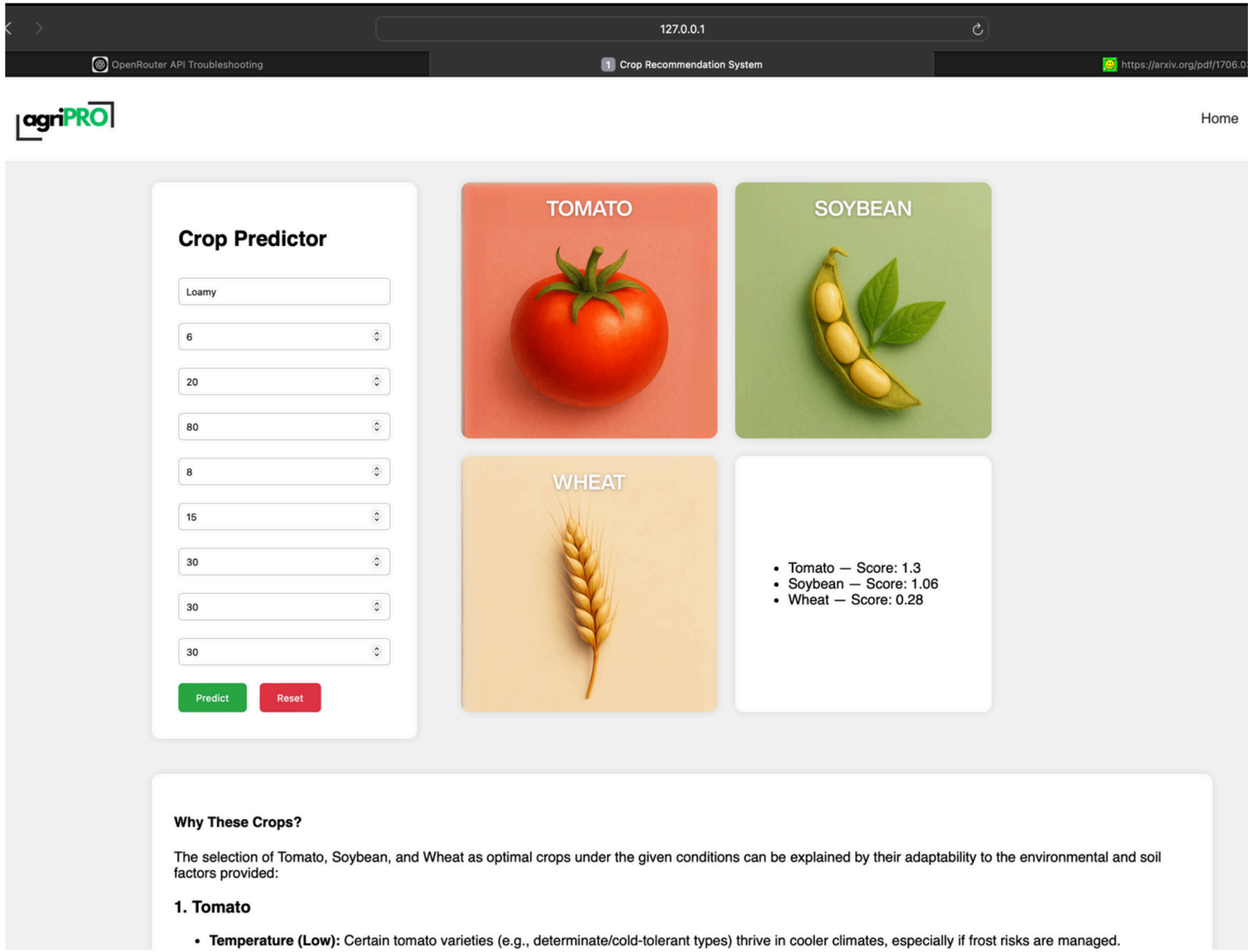
Implementation of Portal - AgriPRO

- **Built Front End using HTML**
- **Backend using Flask**
- **Model tweeked and deployed**
- **Named the portal AgriPRO**



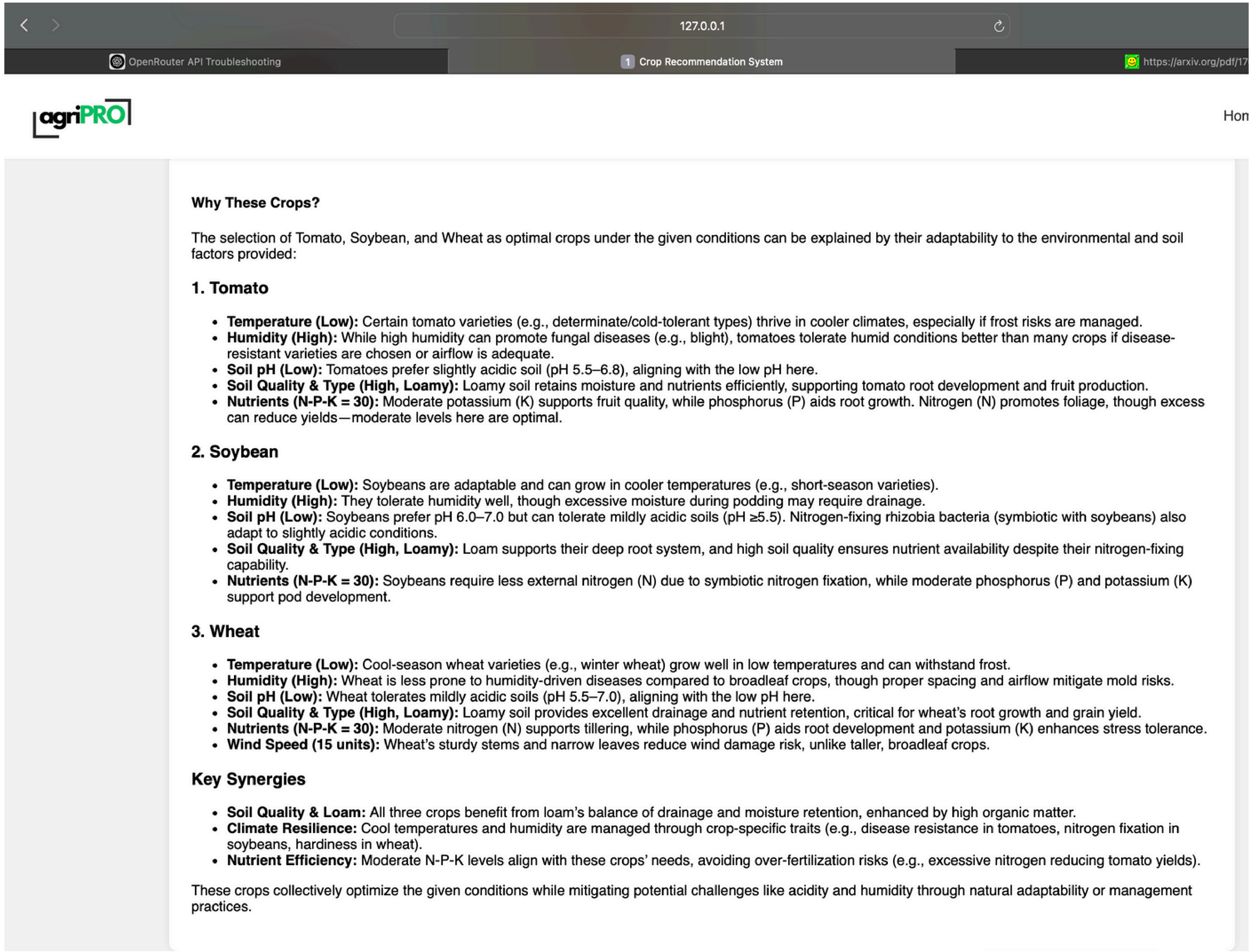
Implementation of Portal - AgriPRO

- User fills a form entering details for prediction
- The Top 3 crops based on the data are suggested
- Made bespoke icons for each crop in database, for a better vusual experince
- Scores for each crop also returned
- Explaination Module



DeepSeek Explainability Module

- **Model: R1**
- **Used to explain how the crops that were selcted are indeed the best option**
- **Works hand in hand with prediction model**
- **Integrated via System Prompt and API**



Going forward

- More feature engineering
- Sorting out the issues with Tier 1 and Tier 2 crops
- Integrating more AI based insights on the portal