Crop Yield Prediction using Soil and Weather Data

Mahi Chauhan — BTech (CSE: Data Science) • Mentor: Mr. Sover Singh

Web app (demo): **kishankalyaan.streamlit.app**

Introduction

A machine learning web application that predicts crop yield from soil parameters and local weather. Helps farmers plan and optimize production by providing easy-to-read predictions and insights.

Why it matters

Improves decision-making, reduces risk, and supports sustainable farming.

Core inputs

Soil nutrients, moisture, pH, temperature, rainfall, humidity.

Problem Statement

Farmers lack accessible, data-driven predictions for expected crop yield based on soil health and weather. Decisions are often reactive, causing suboptimal input use and income volatility.

- Uncertainty in yield planning
- Inefficient fertilizer and water use
- Limited access to predictive tools

Goals & Objectives

Primary Objective

Predict crop yield using soil and weather data to guide farmers.

Secondary Objectives

Provide a simple web interface, actionable insights, and model explainability.

Impact

Increase planning accuracy and promote efficient resource use.

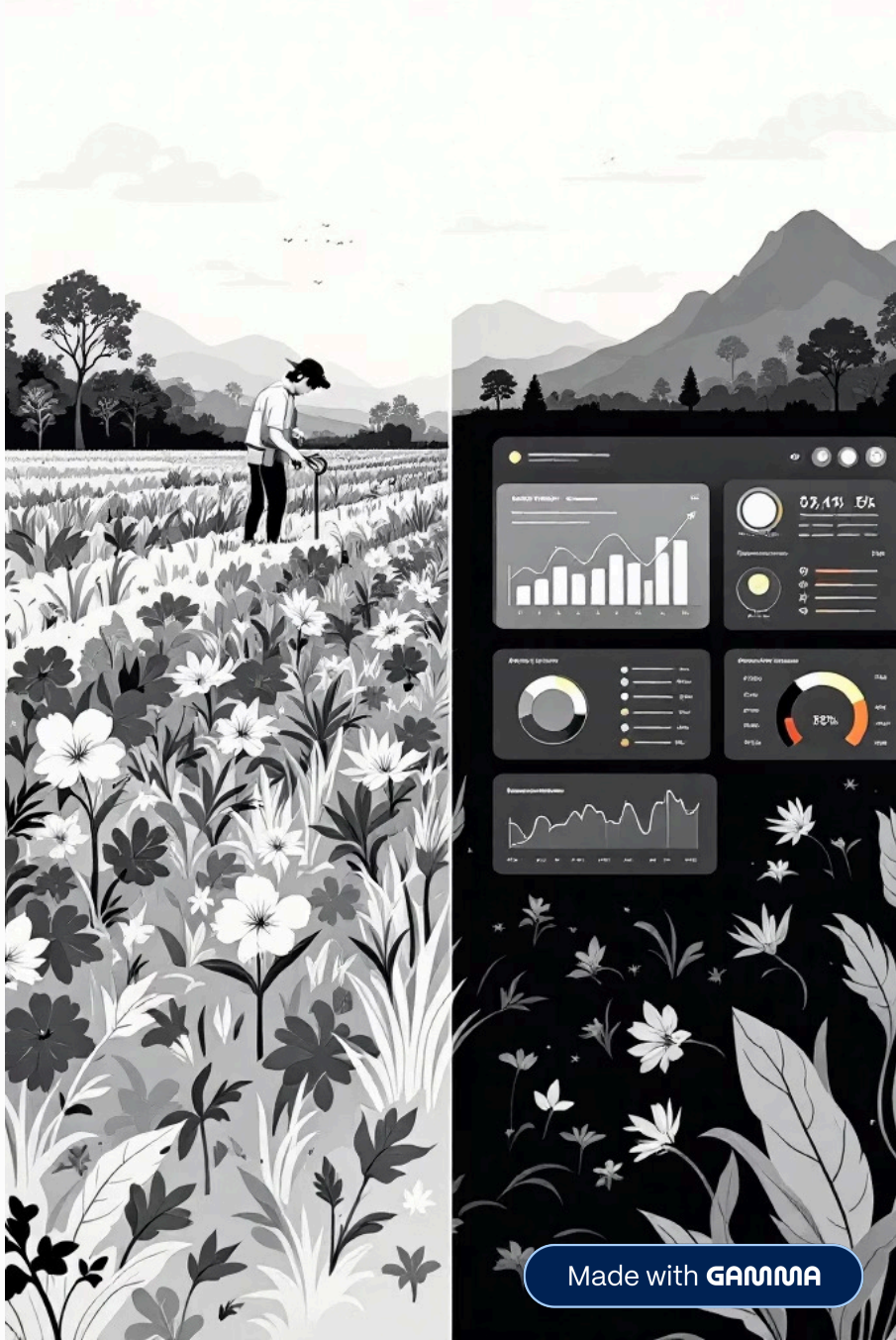
Existing vs Proposed System

Existing System

- Decisions based on experience and rough estimates
- No integrated prediction tool for smallholders
- Reactive input management

Proposed System

- Data-driven yield predictions
- Web interface for easy input and visualization
- Recommendations for better resource allocation



System Overview & Workflow

Data Collection

Soil, weather, sensors, manual inputs

Preprocessing

Clean data and create features

Model Training

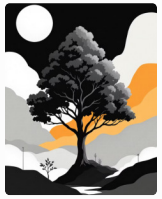
Train and validate ML model

Deployment & Output

Web app shows yield and guidance

The app ingests inputs (manual or sensor), preprocesses features, runs the trained model, and displays predicted yield with brief recommendations for inputs and risk factors.

Model, Tech Stack, Implementation & Results

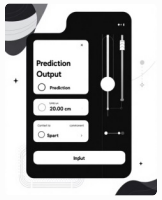


Model

Random Forest Regression — robust to noise, interpretable feature importances.

Tech Stack

Python, scikit-learn, pandas, Streamlit, NumPy, matplotlib/seaborn.



Implementation

Data cleaning → feature engineering → model training → Streamlit deployment.

Key Results

- Accurate yield estimates on test set (examples: RMSE and R^2 reported during demo)
- Model highlights top features: soil nitrogen, rainfall, temperature variance
- Live demo accessible at the provided URL

Challenges & Future Work

- Challenges: limited labeled data, variability in sensor accuracy
- Future: ensemble models, real-time sensor integration, region-specific calibration

Conclusion & References

The project delivers a practical prediction tool for farmers, combining soil and weather analytics into a simple web interface. References include academic papers on crop modeling, scikit-learn docs, and Streamlit guides.