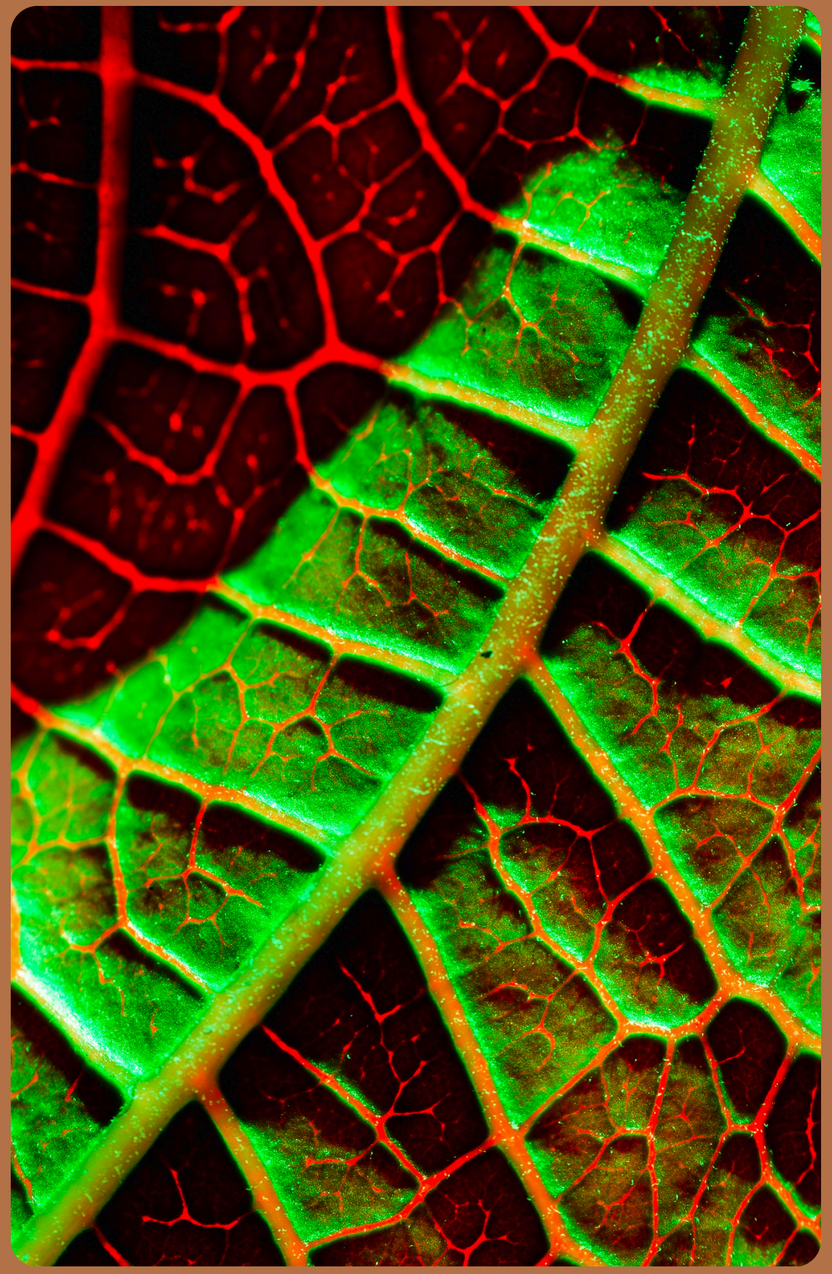
Model to Predict Coffee Disease Risk For Proactive Farm Management



Team members

JORAM MUGESA

CATHERINE KAINO

JUNE MASOLO

KHISA

Problem Background

Coffee production in Kenya faces growing risk from unpredictable diseases like Coffee Leaf Rust, which can cut smallholder yields by up to 70% and cause financial instability. Current responses are mostly reactive, relying on late symptom detection or costly, indiscriminate fungicide use, resulting in economic losses and environmental harm. There is a major need for an early-warning system that uses environmental data to enable timely, proactive disease management.

Core Question: How can we move from guesswork and reaction to science-based, proactive action?

Objectives

What This Analysis Aims to Achieve

Climate Characterization

Document baseline weather patterns (2010-2020) in Kenya's coffee-growing regions

Trend Identification

Detect emerging climate patterns and shifts affecting coffee suitability

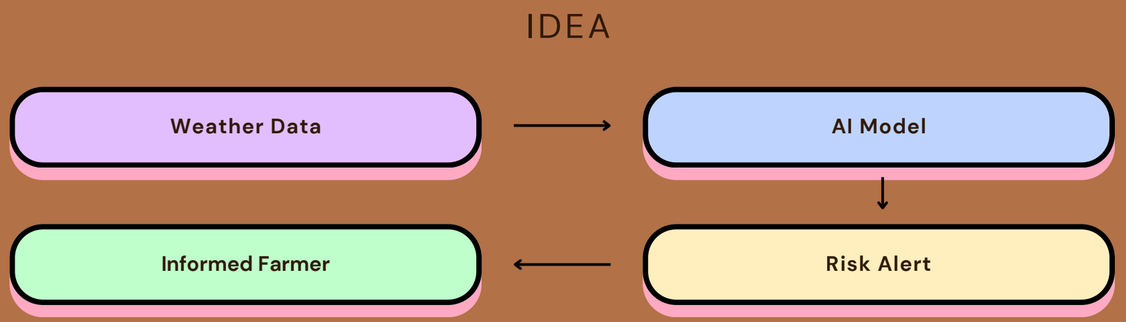
Risk Assessment

Evaluate climate-related threats to yield, quality, and farmer livelihoods

Adaptation Framework

Develop evidence-based recommendations for climate-resilient coffee production

Our Solution: A Proactive Early-Warning System



01

We built a predictive model that acts like a "weather forecast for disease."

The Big Idea

02

The model analyzes key environmental conditions and historical patterns to predict the risk level (Low, Medium, High) of a disease outbreak before it happens.

How it Works

03

It gives farmers actionable intelligence, telling them when and if intervention is truly necessary.

Key Benefit

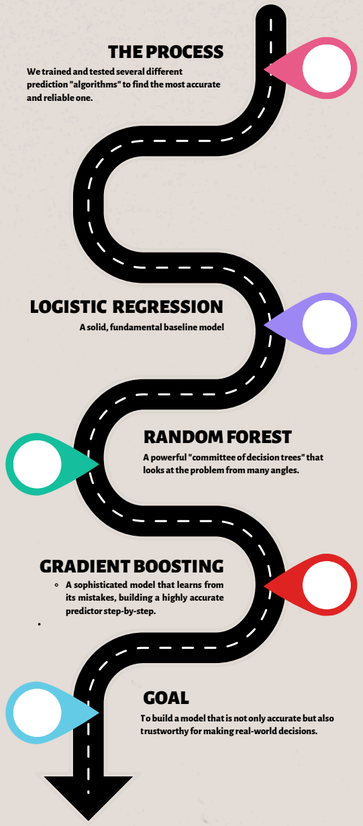
Dataset Used

Weather Data: Historical and forecast meteorological data (temperature, humidity, rainfall) from the NASA POWER API, specifically tailored for agromodeling. • Link: https://power.larc.nasa.gov/

Extracting Data from NASA POWER API (The model will use Nasa Power dataset from 01-01-2010 to 31-12-2020 focusing on Coffee plantation Kenya in Nyeri area(major coffee zone)).

Building the "Brain": Our Machine Learning Approach

A simple graphic showing three different "tools" (icons for a decision tree, a group of trees, and a boosting arrow) being tested.



Methodologies

5. **DEPLOYMENT**

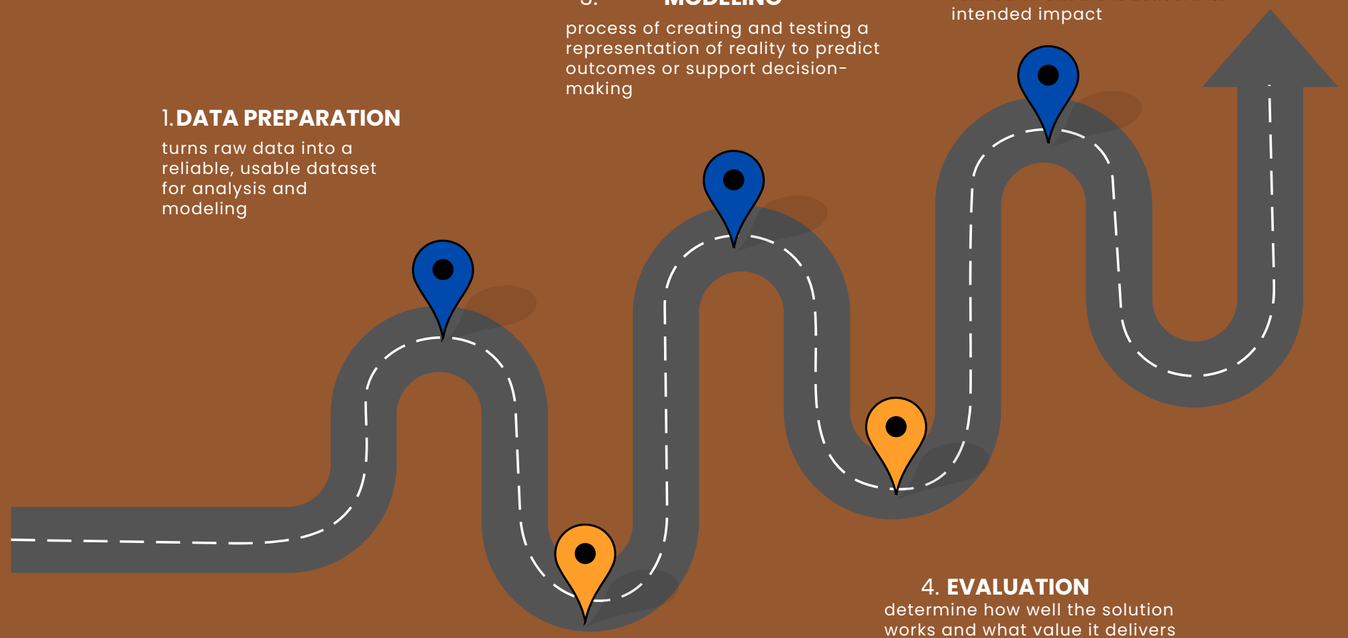
A developed solution is rolled out, used in practice, and refined to ensure it delivers its intended impact

3. **MODELING**

process of creating and testing a representation of reality to predict outcomes or support decision-making

1. **DATA PREPARATION**

turns raw data into a reliable, usable dataset for analysis and modeling



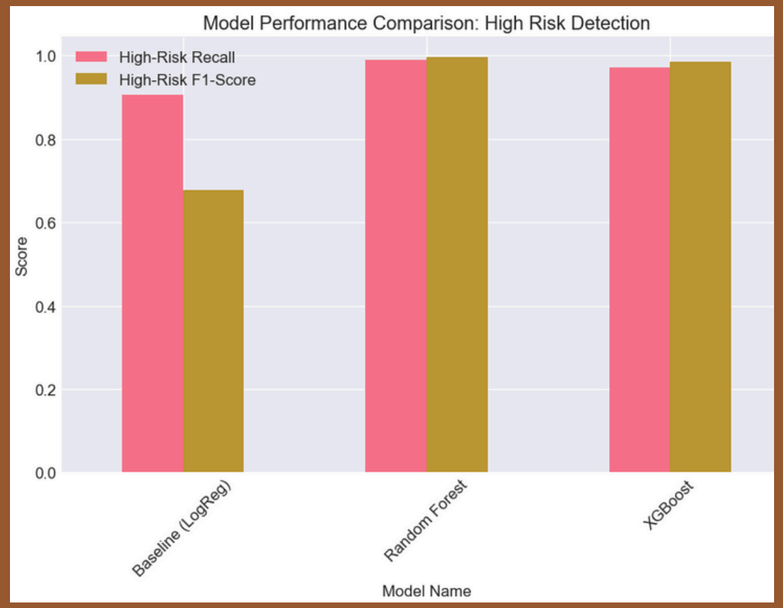
4. **EVALUATION**

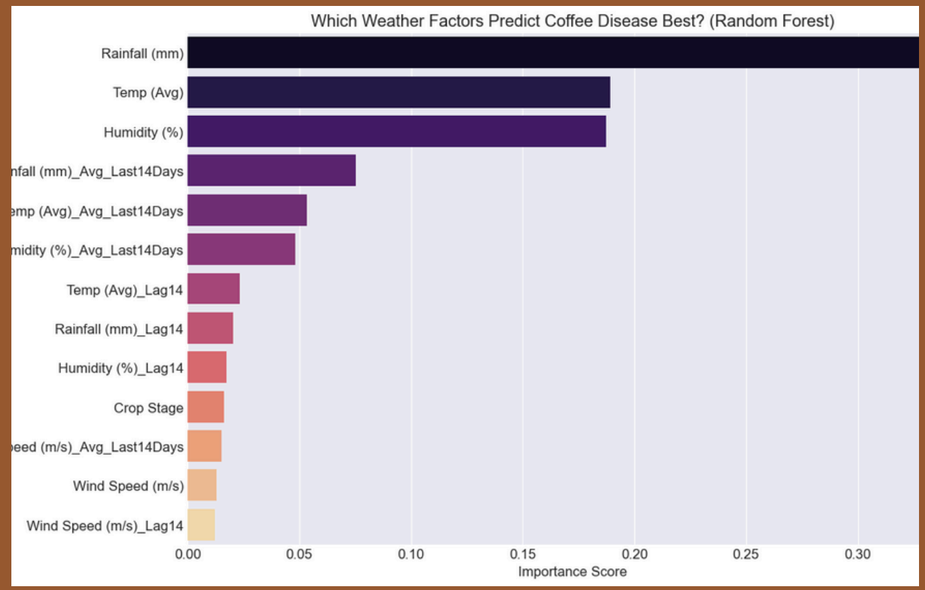
determine how well the solution works and what value it delivers

2. **EXPLORATORY DATA ANALYSIS (EDA)**

make sense of the data and uncover insights before modeling

Evaluation Results

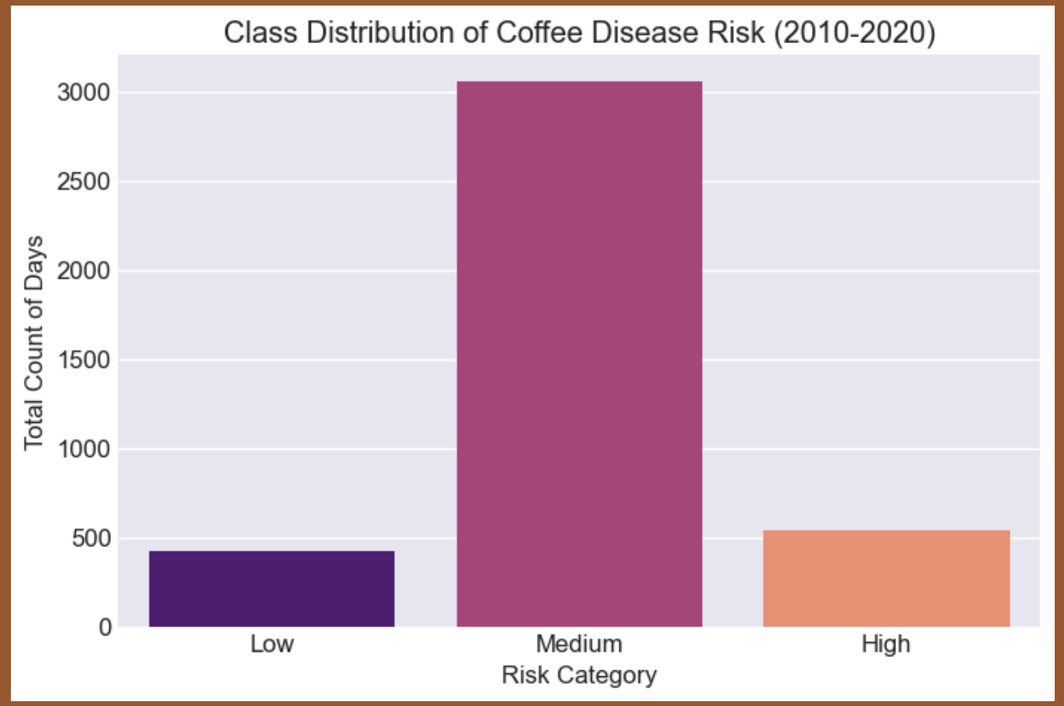




The Key Finding:
The baseline logistic regression model provides moderate accuracy but lacks reliability due to low precision, leading to many false outbreak warnings despite catching most real cases. In contrast, Random Forest and XGBoost perform exceptionally well, achieving near-perfect accuracy and precision. Random Forest is best at detecting nearly all high-risk outbreaks, while XGBoost delivers the highest overall accuracy across all classes.

The Key Finding: Central Tendency and Consensus

Key EDA Finding

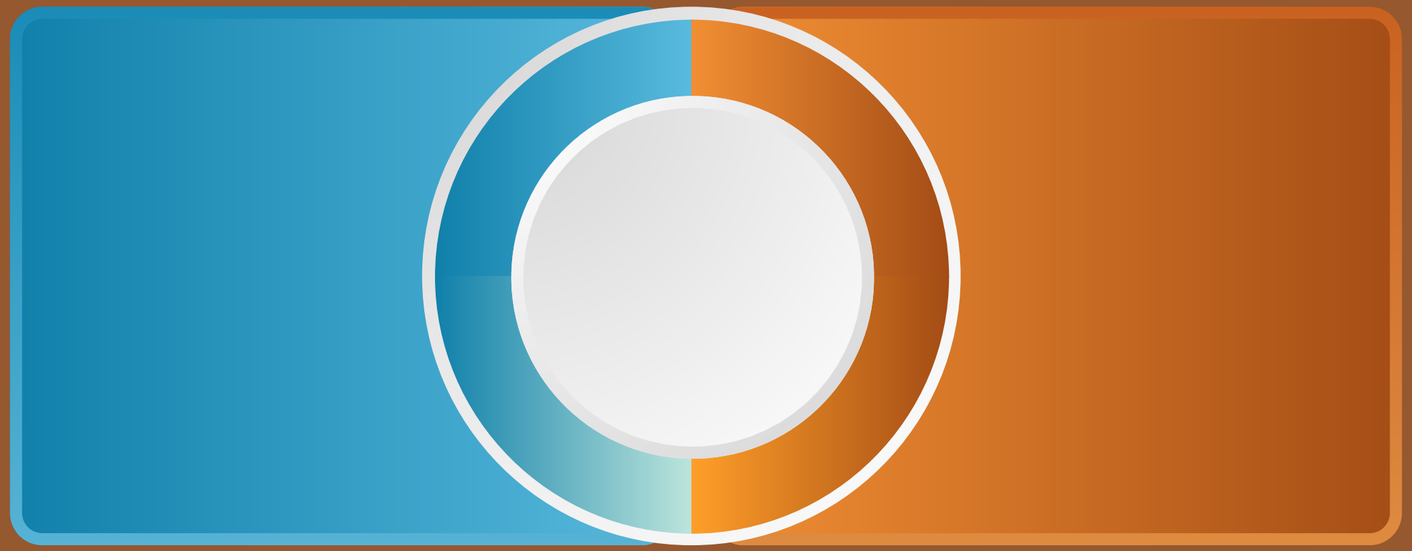


The data shows a moderate class imbalance, with far more “Medium” risk days than “High” risk days. To address the business impact of missing high-risk events, the model will use balanced class weights, ensuring that failing to detect a high-risk day is treated as significantly more costly than misclassifying a medium-risk day.

Your paragraph text

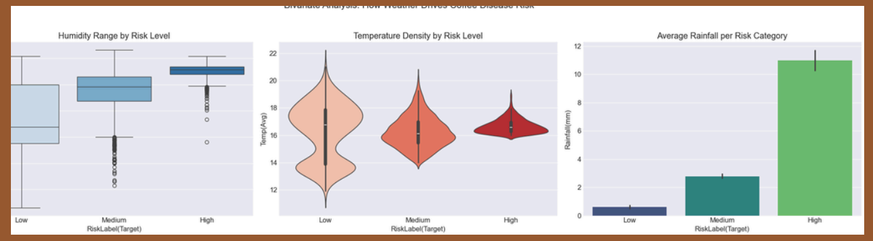
Modeling

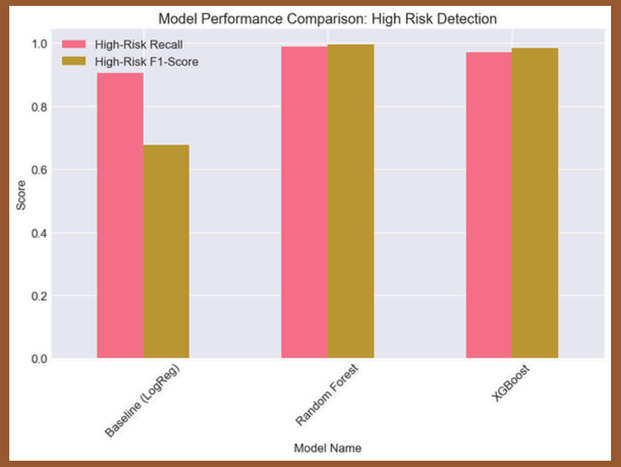
entails building three models(Logistic regression, Random Forest and XGBoost Classifier) using ML pipelines.

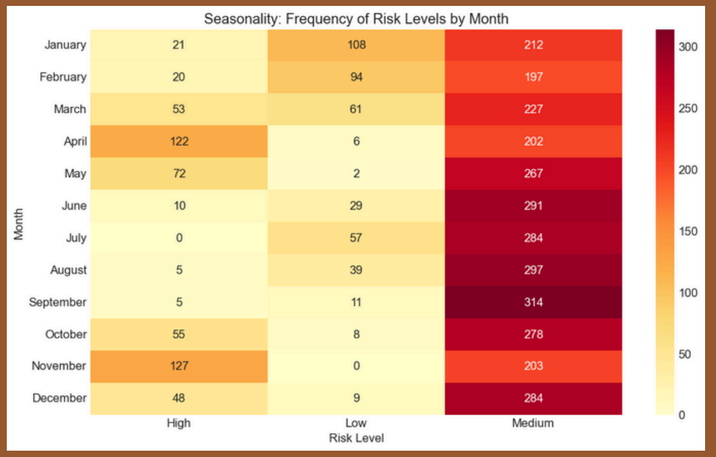


create a loop to run all three models. This allows us to see the Baseline (Logistic Regression) vs. the Ensemble models (Random Forest and XGBoost) side-by-side.

Does It Work? Validating Our Model







- **How We Measure Success**: We evaluated the model using key business-focused metrics:
  - **Precision**: When we say "High Risk," how often are we right? (Minimizes false alarms).
  - Recall: Can we catch most of the actual High-Risk periods? (Minimizes missed outbreaks).
  - **F1-Score**: The balance between Precision and Recall – our overall measure of reliability.
- **Takeaway:** The model achieved strong, balanced scores, confirming it is ready to provide dependable guidance.

From Prediction to Action: The Farmer's Dashboard

  - **"ALERT: High Disease Risk predicted for your zone in next 5 days. Consider fungicide application."**
  - **Dashboard View:** A map of the farm with a "Low Risk" green zone, or a simple calendar with red/yellow/green days.
- **The Final Output:** The model doesn't just give a number. It translates into a clear, actionable alert for the farmer.
- **Impact:** This direct communication bridges the gap between complex data science and practical farm management.

Tangible Benefits: Why This Matters

Economic Win for Farmers

Reduces unnecessary chemical costs by 30-50%, applying fungicides only when needed. Protects their primary source of income.

Economic Win for Farmers

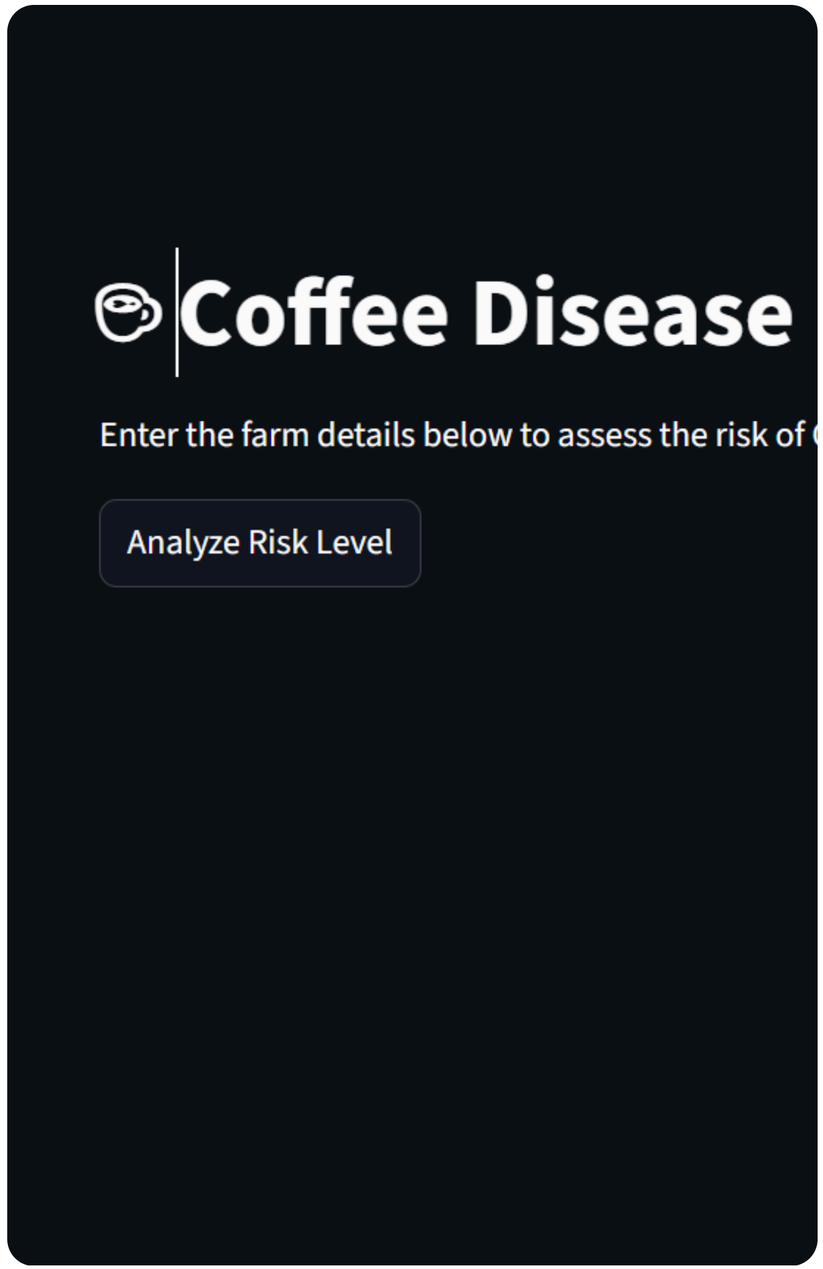
Minimizes chemical runoff, protecting soil health and local ecosystems. Promotes sustainable farming.

Economic Win for Farmers

Protects crop yields by enabling early, targeted intervention, securing Kenya's coffee production.

Conclusion & Vision

- **Headline: Transforming Coffee Farming with Data-Driven Decisions.**
- **Our Vision: To make proactive, science-based farm management accessible to every smallholder coffee farmer in Kenya.**
- **Final Message: This project demonstrates that modern data science can be harnessed to solve critical, real-world agricultural challenges, creating a more sustainable and prosperous future for a vital industry.**



Thank You!

We appreciate your questions and engagement