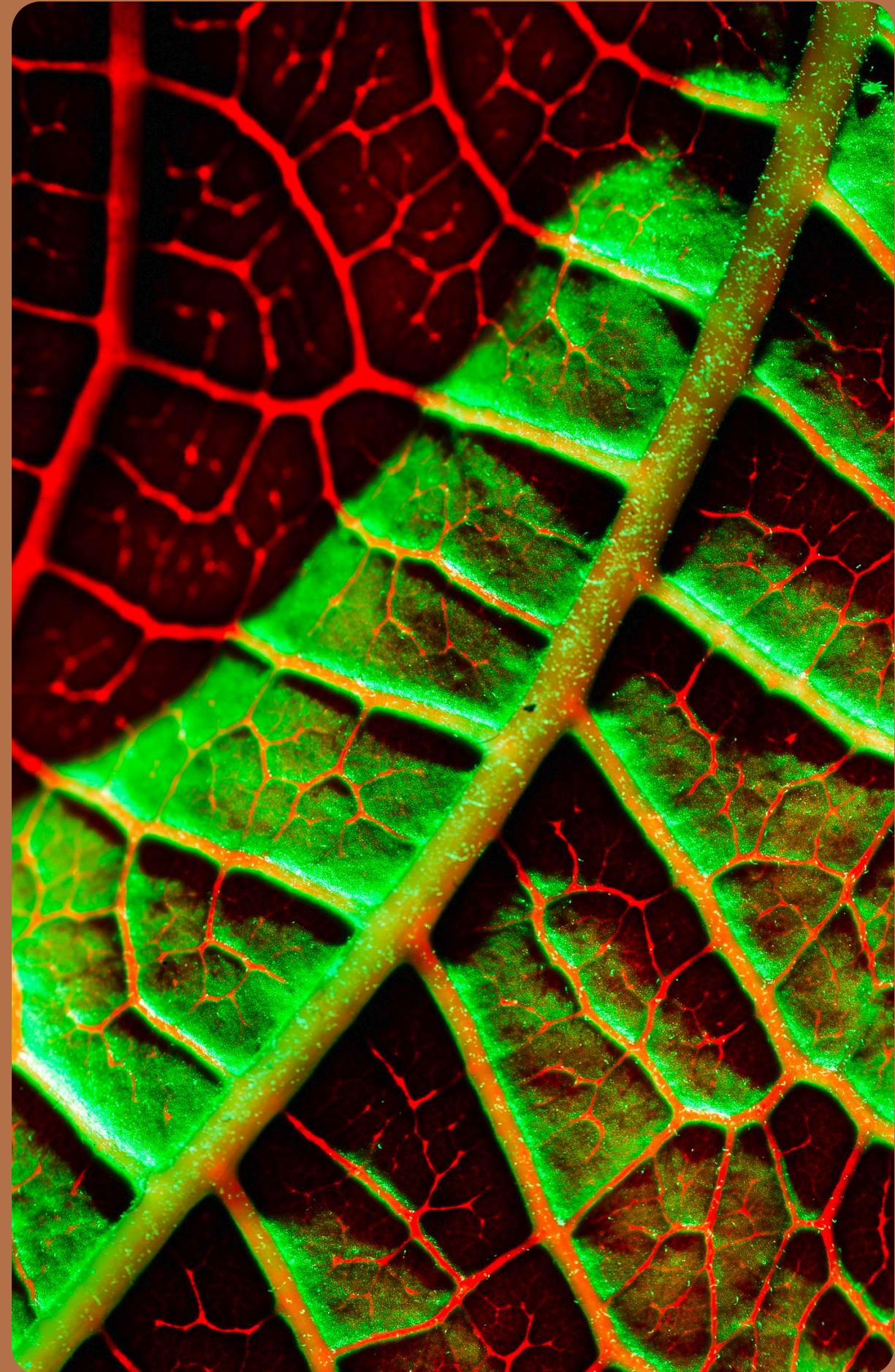
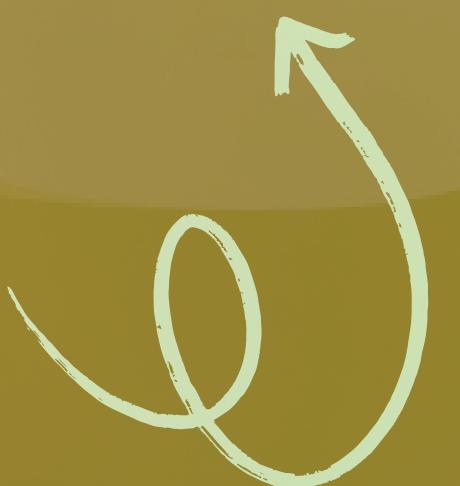


# Model to Predict Coffee Disease Risk For Proactive Farm Management



# Team members



**JORAM MUGESA**



**CATHERINE KAINO**



**JUNE MASOLO**



**KHISA**



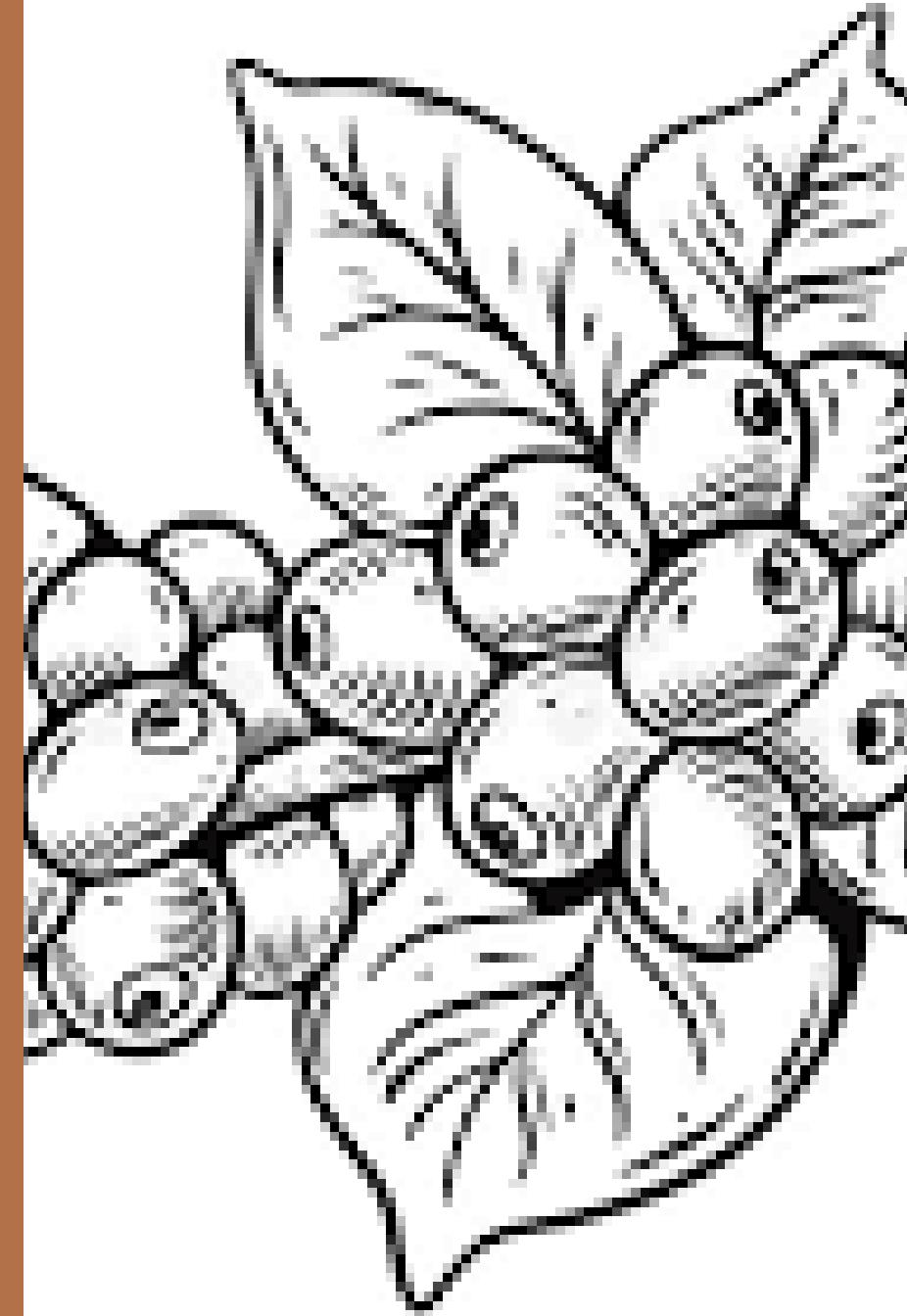
**KENNEDY OMORO**



**ALIVIN NGENO**

# 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.**



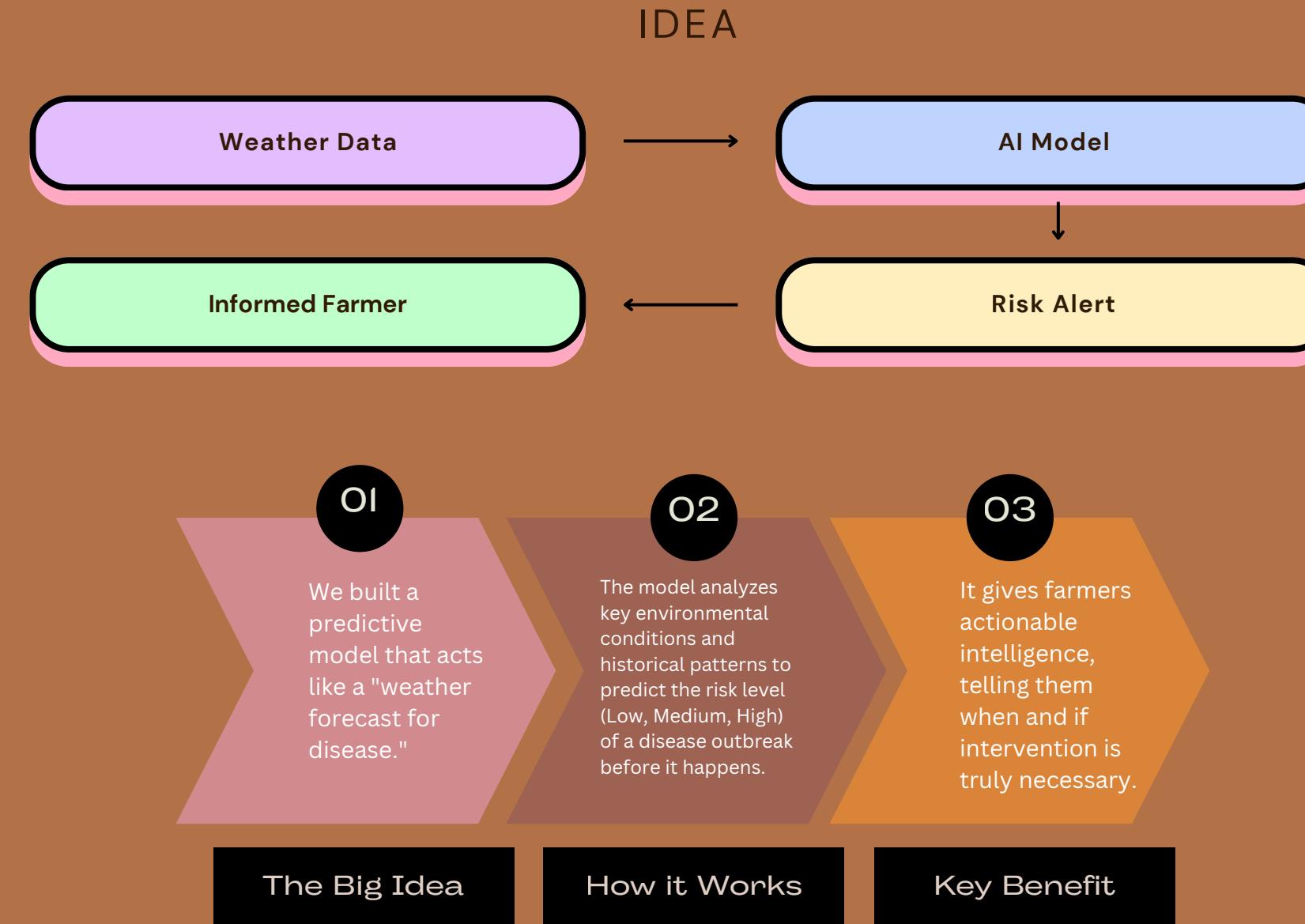
**Core Question: How can we move from guesswork and reaction to science-based, proactive action?**

# Objectives

## What This Analysis Aims to Achieve



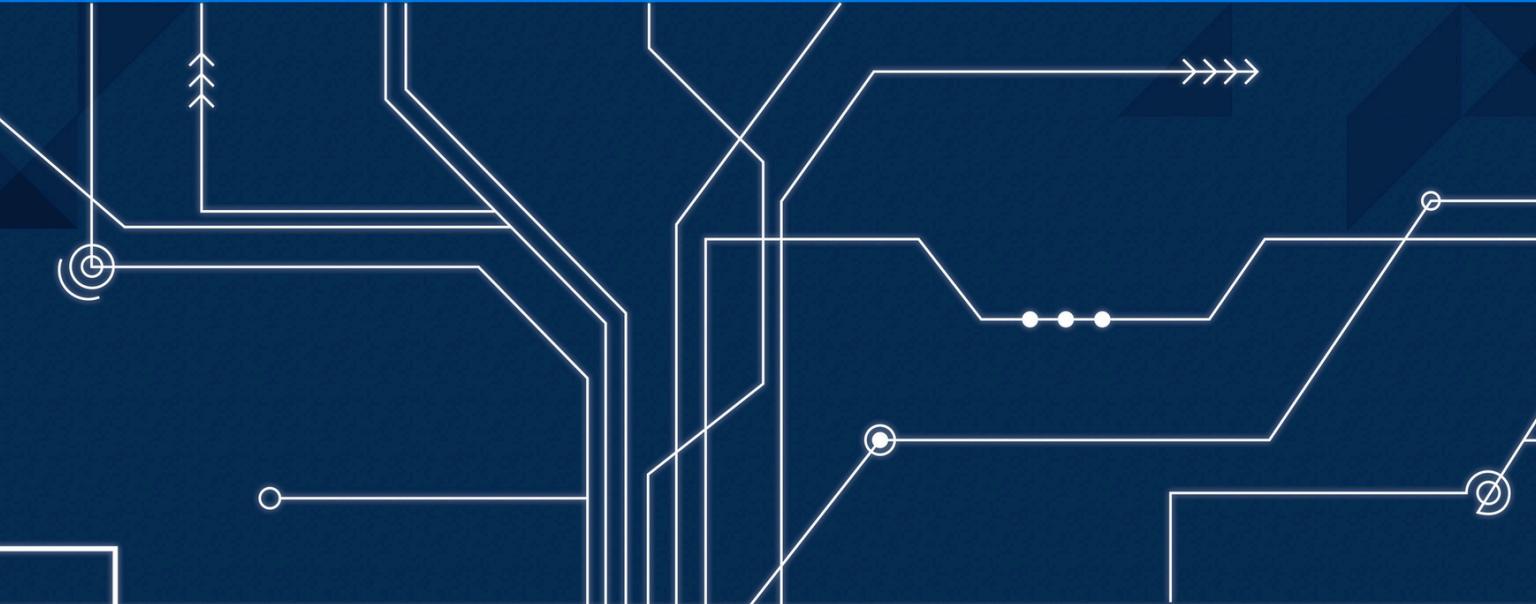
# Our Solution: A Proactive Early-Warning System



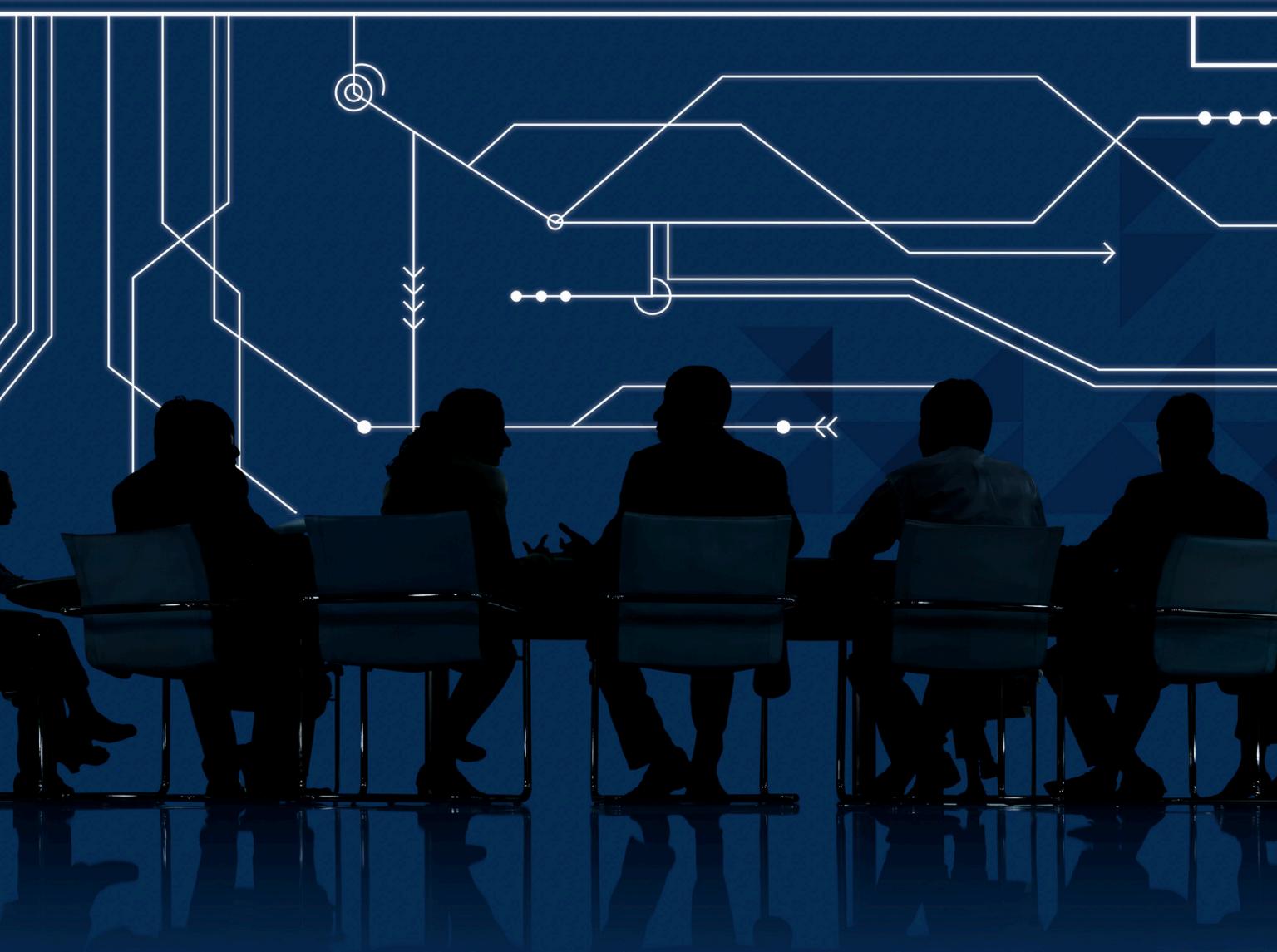
# Dataset Used

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)).

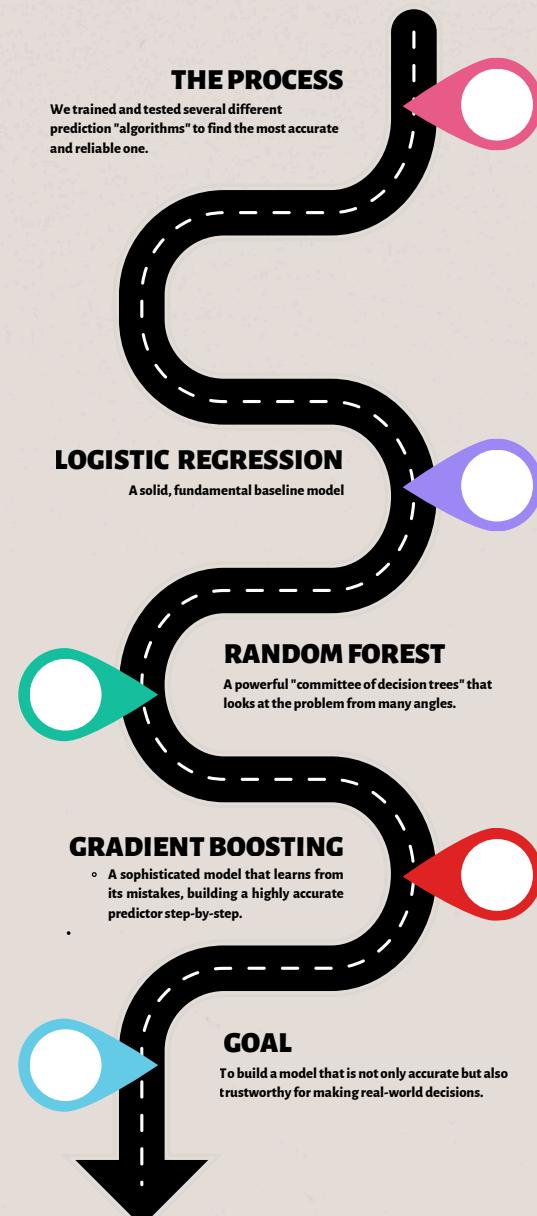


# Database



# Building the "Brain": Our Machine Learning Approach

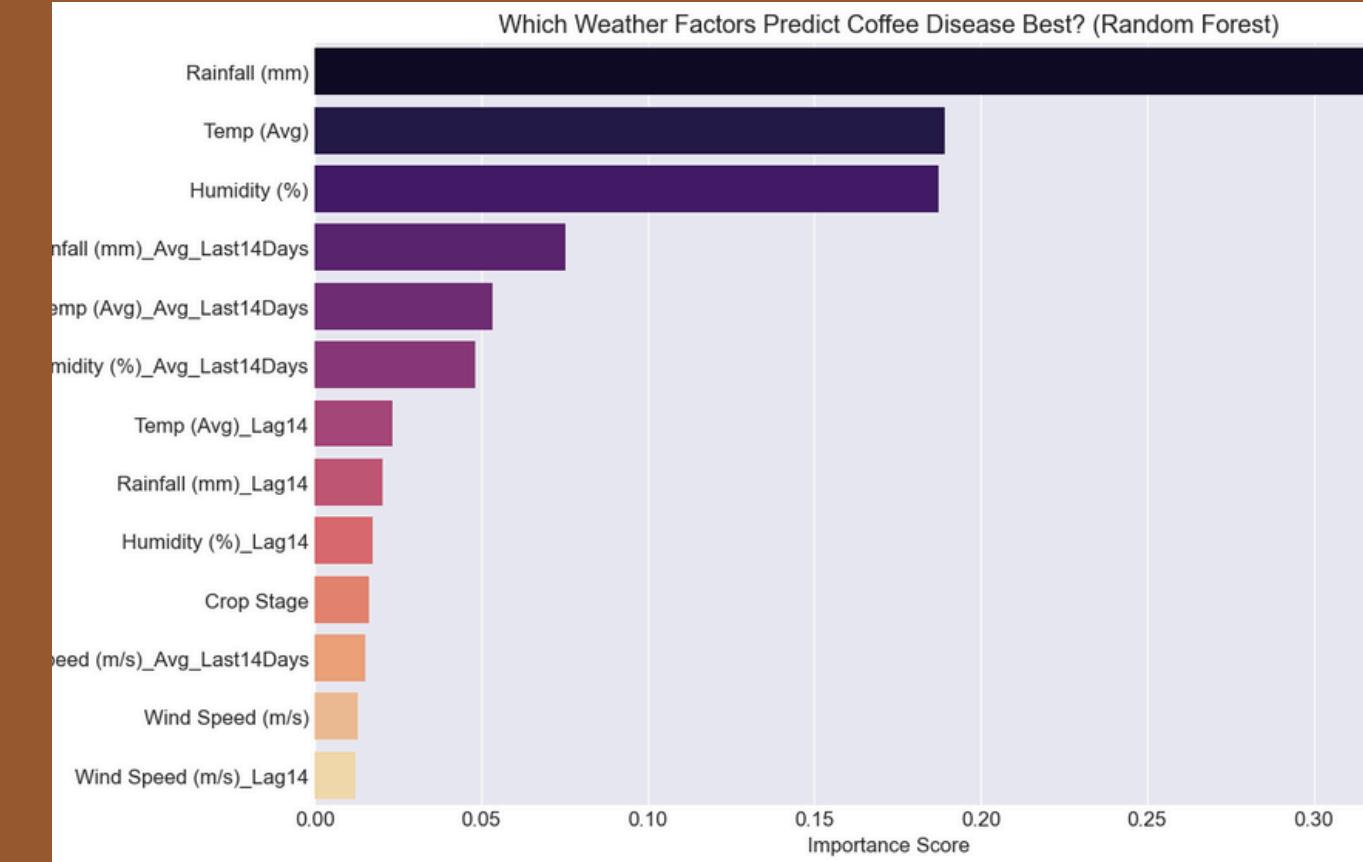
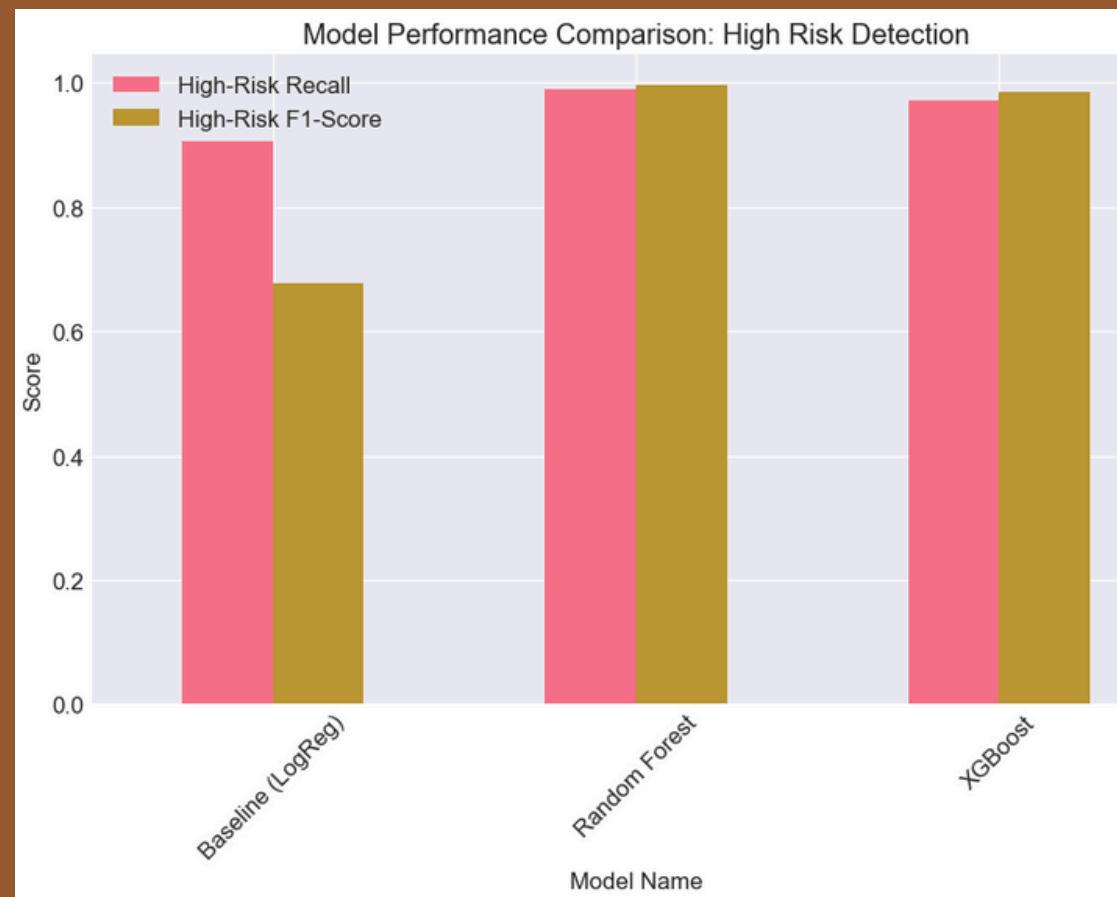
A simple graphic showing three different "tools" (icons for a decision tree, a group of trees, and a boosting arrow) being tested.



# Methodologies



# Evaluation Results

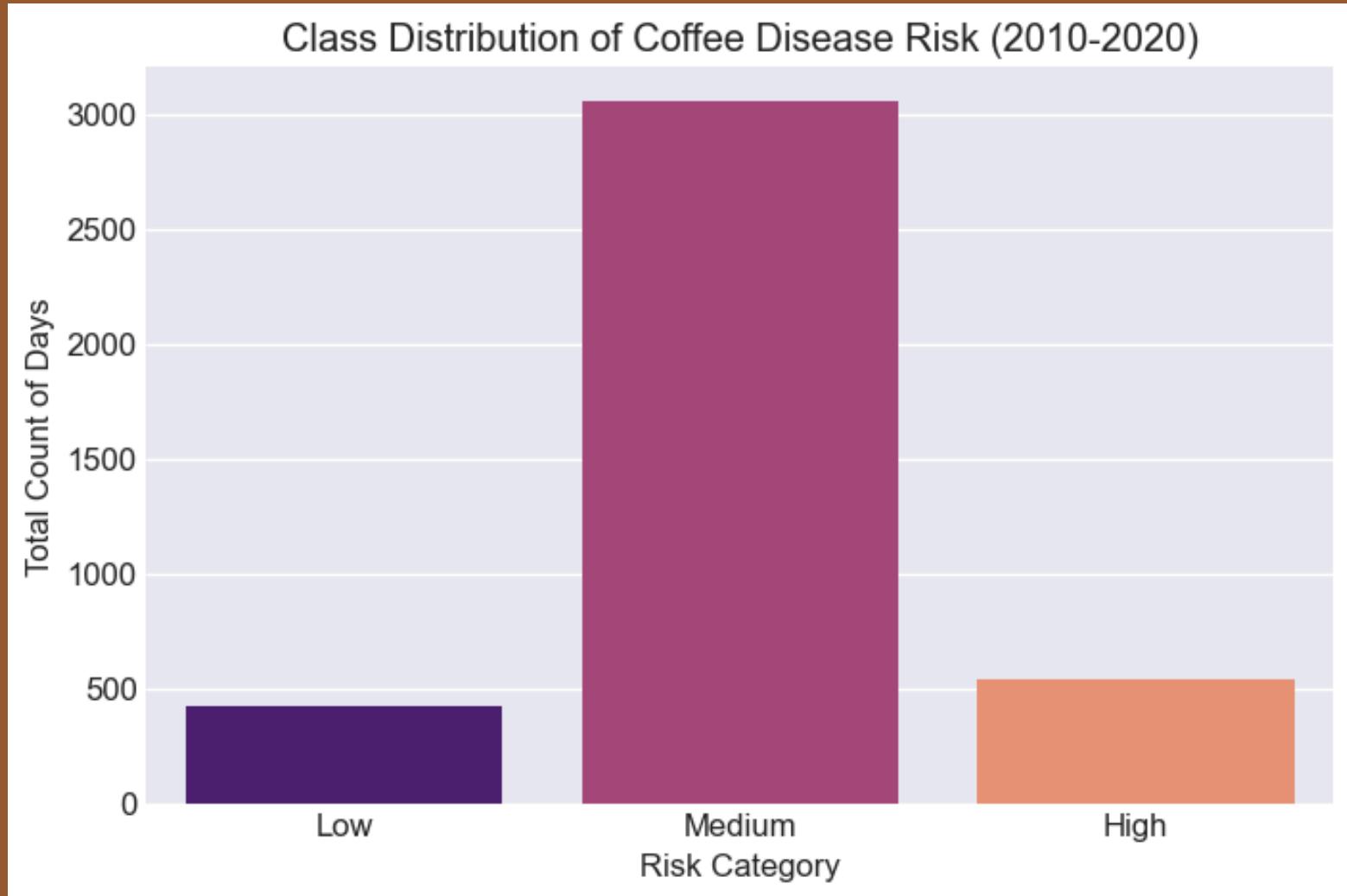


## The Key Finding:

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.

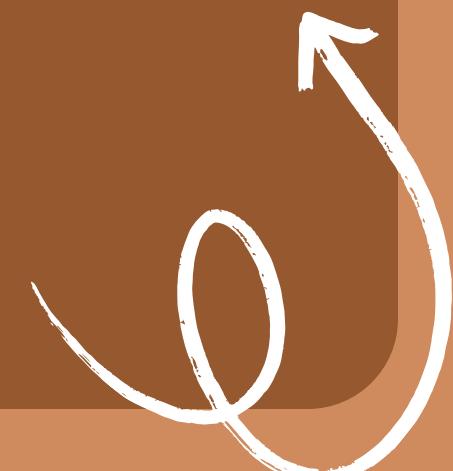
Your paragraph text

# Modeling

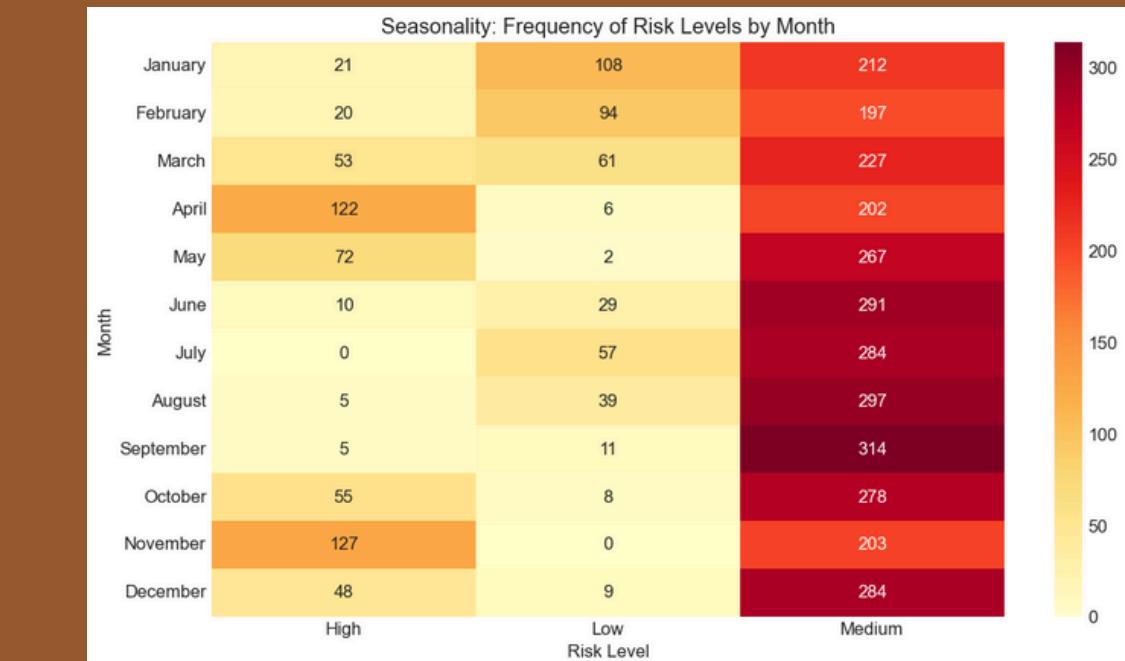
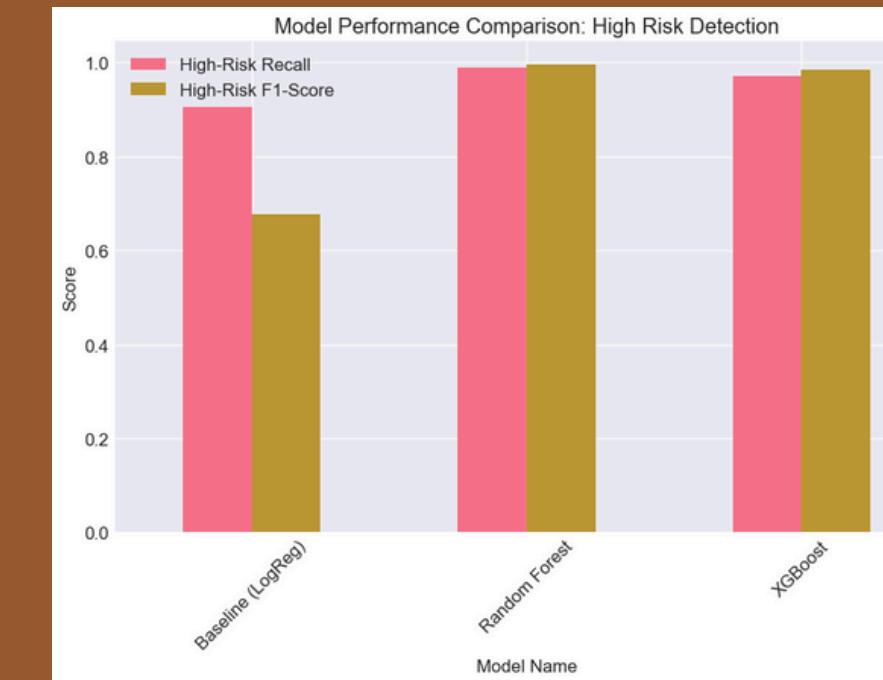
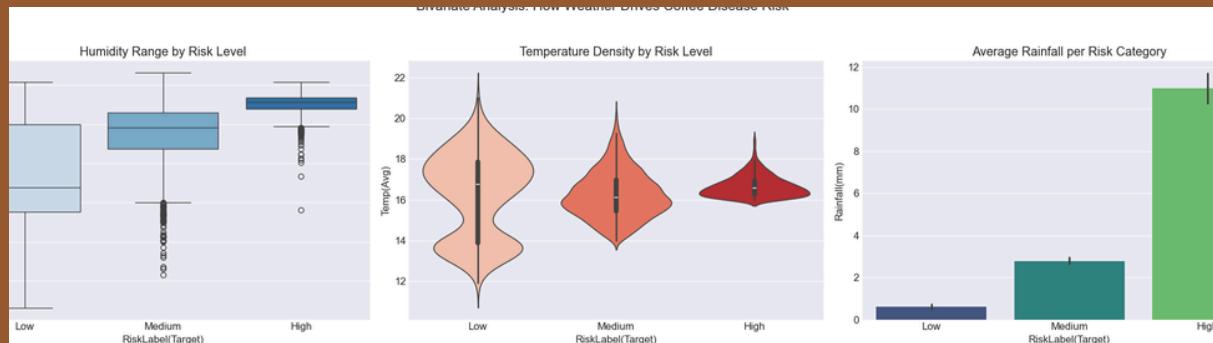
entails building three models(Logistic regression, Random Forest and XGBoost Classifier) using ML pipelines.



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

The screenshot shows a mobile-style dashboard for coffee disease risk prediction. On the right, a main panel displays the title "Coffee Disease Risk Prediction" with a magnifying glass icon, followed by the instruction "Enter the farm details below to assess the risk of Coffee Leaf Rust." Below this is a button labeled "Analyze Risk Level". On the left, a sidebar titled "Input Weather & Farm Data" contains five input fields with sliders: "Average Temperature (°C)" set at 22.00, "Humidity (%)" set at 70.00, "Rainfall (mm)" set at 5.00, "Wind Speed (m/s)" set at 2.00, and "Month (1-12)" set at 6. At the bottom of the sidebar is a "Crop Stage" section.

- **"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.

# Coffee Disease

Enter the farm details below to assess the risk of coffee disease.

Analyze Risk Level

# Thank You!

We appreciate your questions and engagement