# **METHODOLOGY.md**

## **Model Design**

### **1. Credit Scoring**

Ensemble of **Gradient Boosting** and **Logistic Regression** models trained on multi-source alternative data to generate a dynamic *Financial Readiness Score*.  
 **Key inputs:**

* M-Pesa transaction history
* Farm sales records and listing data
* Reliability ratings and guarantor credibility
* Local climate and soil data
* Farm size
* Crop yield

The ensemble approach balances interpretability (Logistic Regression) with performance (Gradient Boosting), improving both transparency and predictive accuracy.

### **2. Parametric Insurance Triggers**

Weather-indexed **parametric models** determine insurance payouts automatically using rainfall, temperature, and storm data from remote sensing and weather APIs.

**Trigger conditions:**

* Drought or excessive rainfall based on threshold deviations
* Hailstorm confirmation via satellite or weather event logs
* Pest epidemic alerts (e.g., armyworm) linked to local outbreak data

When a trigger is met, a **payout or recovery loan** is automatically issued via M-Pesa — eliminating assessor delays and minimizing catastrophic loss.

### **3. Carbon Scoring & Sustainability Analytics**

Rule-based and geospatial models estimate farm-level **carbon offset potential** based on emission factors, land use, and practices.

**Inputs:**

* Tillage type (no-till, minimal till, conventional)
* Fertilizer and agrochemical use
* Crop type and rotation cycles
* Farm transport and logistics data

The system supports future scalability into verified **carbon credit schemes**, enabling smallholders to earn from sustainable practices.

### **4. Farm Management & Advisory Engine**

An AI-powered **farm management module** provides record-keeping, crop cycle tracking, and personalized recommendations.

**Features:**

* Predictive crop calendar (planting, irrigation, harvest reminders)
* Integration of soil data and localized weather forecasts
* Record logs for inputs, expenses, yields, and pest management
* Offline-first SMS/USSD entry for non-smartphone users

### **5. Marketplace & Logistics Optimization**

A **matching algorithm** connects farmers to verified buyers and transport using clustering and trend analysis.

**Model Logic:**

* Geo-clustering for route optimization and aggregation
* Dynamic pricing based on historical transactions and demand
* Farmer reliability score incorporated for trust and traceability

This reduces post-harvest loss and improves market access and fair pricing.

## **Bias Mitigation & Gender Inclusion**

To address existing **gender gaps** and data bias against women farmers:

* Models are tested on **gender-segregated datasets** to ensure equitable credit access.
* Social collateral and guarantor systems prioritize inclusion of **women-led networks**.
* Advisory and finance prompts are phrased in accessible, inclusive language.

## **Evaluation Metrics**

* **Credit Scoring:** ROC-AUC, Precision/Recall, and Gini coefficient.
* **Insurance Triggers:** False positive/negative rate on event detection.
* **Advisory Accuracy:** Correlation of predicted vs. actual yield timelines.
* **Marketplace Efficiency:** Average transaction distance and time-to-sale.

## **Scalability and Integration**

* **APIs:** OpenWeather, FAO, Sentinel-2, M-Pesa, and National Crop Boards, Africa's Talking.
* **Future Expansion:** Integration of carbon credit registries, cooperative data sharing, and micro-insurance partnerships.