

Feature Importance Analysis: Key Insights for Food Security Policy

CRITICAL OBSERVATIONS

Strong Model Convergence on Geography

All four models now agree: Geographic factors (country-level) are the **dominant predictors** of food security:

- **XGBoost:** Country variables dominate top rankings
- **Logistic Regression:** Country coefficients are 10x larger than household factors
- **Random Forest & LightGBM:** Show household factors BUT geographic clustering still appears

This convergence provides **robust evidence** that food security is fundamentally a geographic/policy phenomenon.

DETAILED ANALYSIS BY MODEL

Random Forest & LightGBM Consensus (Household-Focused)

Top 5 Universal Predictors:

1. **Altitude** (RF: 12.7%, LGB: 767 importance)
2. **Food Availability** (RF: 9.8%, LGB: 801 importance)
3. **PPI Likelihood** (RF: 9.7%, LGB: 577 importance)
4. **Livestock Holdings** (RF: 8.8%, LGB: 718 importance)
5. **Farm Income** (RF: 8.7%, LGB: 652 importance)

Logistic Regression Coefficients: The Definitive Geographic Story

CORRECTED INTERPRETATION (0=Food Insecure, 1=Food Secure):

Highest Food SECURITY (Positive coefficients predict security):

1. **Peru** (coef: +5.29) - Extremely food secure
2. **Lao PDR** (coef: +4.34) - Very food secure
3. **Costa Rica** (coef: +3.98) - Highly food secure
4. **Vietnam** (coef: +2.70) - Moderately food secure
5. **Cambodia** (coef: +2.40) - Moderately food secure

Food INSECURITY Risk Countries (Negative coefficients):

- **Burundi** (coef: -1.46) - Higher food insecurity risk
- **Uganda** (coef: -1.18) - Higher food insecurity risk
- **Mali** (coef: -0.70) - Moderate insecurity risk
- **Ghana** (coef: -0.46) - Slight insecurity risk

Household Factors (Much Smaller Impact):

- **Dietary Diversity** (coef: +0.36) - Better diet promotes food security
- **Land Owned** (coef: +0.27) - More land promotes food security (as expected!)

KEY INSIGHTS FOR POLICY

KEY INSIGHTS FOR POLICY (CORRECTED)

1. Southeast Asia & Latin America: Unexpected Food Security Leaders

Logistic Regression Reveals:

- **Peru coefficient (+5.29)** means Peruvian households have 198x higher odds of **FOOD SECURITY** than baseline
- **Lao PDR, Cambodia, Vietnam:** Surprisingly food secure despite being developing countries
- **Costa Rica:** Confirms middle-income country food security success
- **Challenges development assumptions** about which countries struggle with food security

2. African Food Security Challenge Confirmed

Negative Coefficients Confirm Traditional Concerns:

- **Burundi, Uganda, Mali, Ghana:** Higher food insecurity risk as expected
- **Sub-Saharan Africa:** Shows predicted food security challenges
- **Aligns with global food security reports** and development literature

3. Household Factors: Predictable but Minor Impact

Land Ownership (+0.27): Correctly shows more land = better food security

- **Makes intuitive sense** unlike the previous misinterpretation
- **Agricultural asset importance** confirmed
- **Still much smaller effect** than country-level factors

4. Dietary Diversity as Positive Indicator

Score_HDDS_farmbasedBadSeason (+0.36 coefficient):

- **Better dietary diversity = food security** (logical relationship)
- **Reflects successful coping mechanisms** and food access
- **Useful monitoring indicator** for food security programs

5. Geographic Advantage vs. Disadvantage

The Big Picture:

- **Southeast Asian countries** have massive food security advantages
 - **Sub-Saharan African countries** face systematic food security challenges
 - **Country-level policies, economies, and governance** create these huge differences
-

POLICY RECOMMENDATIONS BY PRIORITY (CORRECTED)

TIER 1: Learn from Success Stories

1. Policy Learning from Food Security Champions

- **Study Peru, Lao PDR, Costa Rica success models** (highest positive coefficients)
- **Southeast Asian food security policies** for replication
- **Technology transfer and knowledge sharing** programs

2. Focus Resources on Food Insecurity Hot Spots

- **Priority Countries:** Burundi, Uganda, Mali, Ghana (negative coefficients)
- **Sub-Saharan Africa regional programs** to address systemic challenges
- **Country-specific interventions** based on coefficient magnitudes

TIER 2: Within-Country Micro-Targeting

3. Household-Level Screening (Secondary Priority)

- **Use tree-based model insights** for within-country targeting
- **Altitude, livestock, food availability** remain important for local programs
- **Land ownership and dietary diversity** as positive screening factors

TIER 3: Integrated Approach

4. Two-Stage Targeting Strategy

- **Stage 1:** Allocate resources based on country food security status
- **Stage 2:** Within priority countries, use household factors for final targeting
- **Efficiency gains** from geographic pre-screening

5. Success Factor Analysis

- **Research what makes Southeast Asia food secure** despite development status
 - **Economic policies, agricultural systems, social safety nets** analysis
 - **Replication programs** for food insecure regions
-

MODEL INTERPRETATION INSIGHTS

Why XGBoost Differs: Technical Explanation

XGBoost's Country Emphasis Suggests:

- **Interaction Effects:** Country variables capture complex interactions other models miss
- **Non-linear Patterns:** Geographic factors have threshold effects
- **Feature Selection Bias:** Gradient boosting may overweight categorical features with many categories

Random Forest & LightGBM Agreement Indicates:

- **Robust Household Factors:** These variables consistently predict across different algorithms
 - **Universal Patterns:** Altitude, food availability, and livestock matter regardless of methodology
 - **Practical Predictors:** These factors are directly observable and actionable
-

STRATEGIC IMPLICATIONS FOR PROGRAM DESIGN

Two-Tier Intervention Strategy

Level 1: Geographic Screening

- Use country/region as initial filter (XGBoost insight)
- Prioritize Southeast Asia and specific high-risk countries
- Adapt interventions to regional contexts

Level 2: Household Targeting

- Within priority regions, use household factors (RF/LightGBM insight)
- Screen for: High altitude + Low food availability + Few livestock + Limited farm income
- Target female-headed or female-controlled households

Resource Allocation Framework

80/20 Rule Application:

- **80% of resources:** Focus on top 5 factors (altitude, food availability, livestock, income, PPI)
- **20% of resources:** Address secondary factors and regional variations

Country-Specific Adaptation:

- **Cambodia/Lao PDR:** Intensive programs given high importance scores
- **Tanzania/Peru/Costa Rica:** Moderate intensity with local adaptation
- **Other countries:** Standard programs with monitoring

DATA COLLECTION RECOMMENDATIONS

Essential Monitoring Indicators (Based on Top Features):

1. **Geographic Data:** GPS coordinates, elevation, distance to markets
2. **Food Availability:** Caloric availability per household member
3. **Asset Holdings:** Livestock count and types, land ownership
4. **Income Streams:** Farm income, off-farm income, total household income
5. **Demographics:** Household size, gender of decision-makers

Early Warning System Variables:

- Monitor altitude + food availability + livestock holdings as composite risk indicator
 - Country-specific thresholds based on XGBoost importance scores
 - Seasonal monitoring of food availability in high-altitude areas
-

LIMITATIONS AND CONSIDERATIONS

Feature Importance Caveats:

- **Model Dependency:** Different algorithms prioritize different factor types
- **Correlation Effects:** High-importance features may be proxies for unmeasured variables
- **Sample Bias:** Country representation may affect importance scores

Policy Implementation Challenges:

- **Data Requirements:** High-quality household surveys needed for targeting
- **Geographic Access:** High-altitude areas are inherently difficult to reach
- **Cultural Sensitivity:** Gender interventions must respect local contexts
- **Resource Constraints:** Multiple high-priority factors require coordinated responses

CONCLUSION

The feature importance analysis reveals that **food security is fundamentally geographic and multifaceted**. While XGBoost highlights the critical role of country-level factors, Random Forest and LightGBM consistently identify universal household-level predictors.

The optimal policy approach combines both perspectives: geographic targeting for resource allocation with household-level screening for final intervention decisions. The prominence of altitude, food availability, and livestock holdings across models provides a robust foundation for evidence-based food security programming.