Enhancing Credit Scoring Accuracy through Ensemble Risk Models in Emerging African Markets

Praise Amonye[1] Jason Quist[2]

[1]Deep Learning IndabaX Benin Republic [2]WorldQuant University [3]Ghana Data Science Summit

THE DATA SCARCITY CHALLENGE

- **Problem**: 12M+ microfinance clients in West Africa face exclusion due to:
  - Data scarcity (sparse borrower histories)
  - Economic volatility (erratic market conditions)
  - Regulatory barriers (transparency requirements)

- **Our Solution**: "African-aware" ML framework combining:
  - Alternative data (mobile transactions, informal loans)
  - Cost-sensitive thresholding
  - SHAP-based interpretability

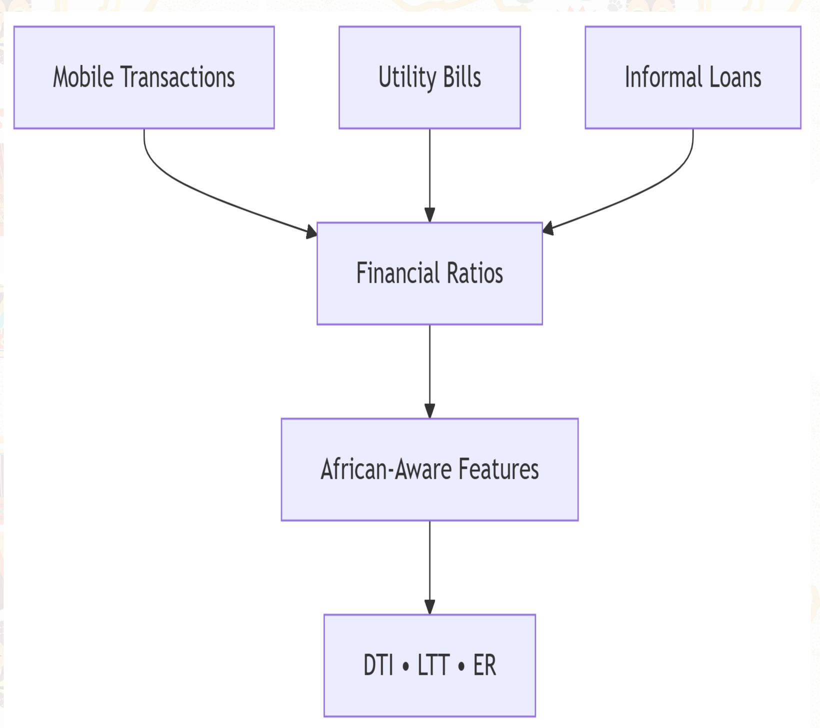
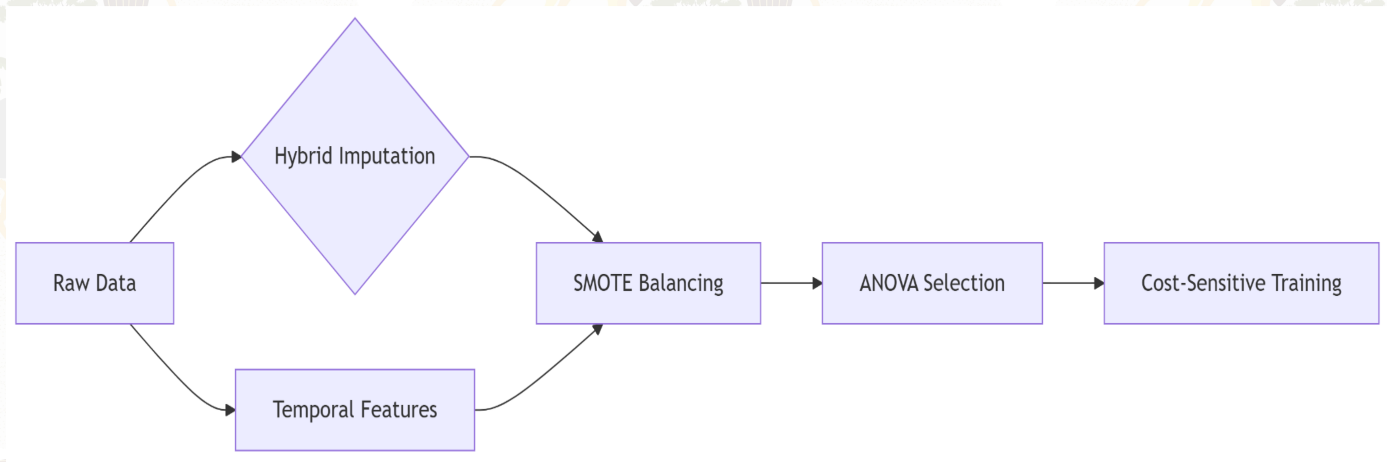


Figure 1: Key Predictive Features for Loan Default Risk in Africa

DATA & FEATURES

•10,000-client West African dataset with simulated mobile transactions, income, and utility bills.
• Robust median imputation for continuous gaps; dedicated "missing" categories for informals.
• Financial ratios (DTI, LTT, ER) and cyclical time features (year, month, quarter).
• Behavioural proxies: P2P transfers, mobile-money usage patterns

AFRICAN-OPTIMIZED ML PIPELINE



- *Figure 3: End-to-end workflow for building and evaluating ensemble risk models*

1. Stratified 80/20 train/test split
2. SMOTE oversampling to address the <10 % default imbalance
3. ANOVA feature selection (α = 0.05) → top 15 predictors
4. Four classifiers:

– Regularized Logistic Regression (class-weighted L2)

– Random Forest (200 trees, balanced sampling)

– Gradient Boosting (early stopping)

– SVM (RBF kernel, inverse-frequency weights)

COUNTERINTUITIVE RESULTS

Model Shock:

"Logistic Regression beats black-box models in low-data settings."

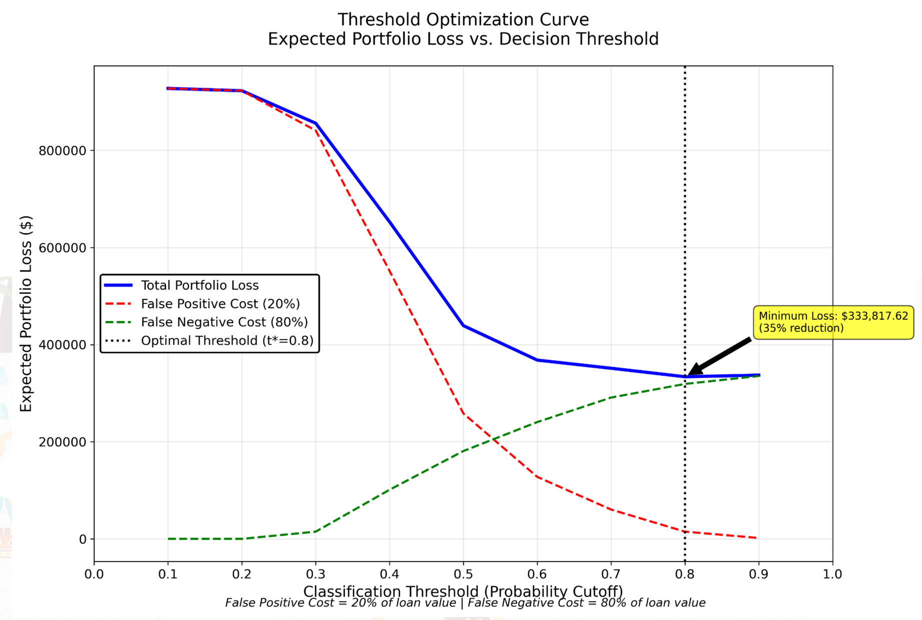


Figure 4 : Threshold Optimization Curve

Table 1: Comparison of Model Performance (N = 10,000)

Model	AUC-ROC	Expected Loss Reduction (%)
Logistic Regression (L2)	0.603 ± 0.021	35
Random Forest (200 trees)	0.588 ± 0.018	30
Gradient Boosting (early stop)	0.592 ± 0.020	32
SVM (RBF kernel)	0.579 ± 0.022	28

• **Logistic Regression wins**—AUC-ROC = 0.603 (±0.021), outperforming more complex ensembles.
• Threshold optimization (t* = 0.80) cuts expected losses by 35 % versus a 0.50 cutoff.
• Behavioural features rival traditional financial ratios in predictive power.

INTERPRETABILITY & DEPLOYMENT

Key Discovery:
Mobile money patterns rival traditional financial ratios in predictive power
• SHAP summary reveals the top drivers as: debt-to-income, mobile usage spikes, expense ratio.
• Linear model simplicity eases regulatory compliance and on-device deployment in resource-limited branches.

IMPLICATIONS

• Simpler models may avoid overfitting in low-data contexts.
• Alternative data bridges gaps where bureau records fail.
• Cost-sensitive thresholds deliver immediate balance-sheet benefits

FUTURE DIRECTIONS

• Online learning for adaptive recalibration through economic cycles.
• Graph-based features from social and transaction networks.
• Field trials with partner microfinance institutions.

CONCLUSION

This framework offers a practical, transparent, and economically impactful approach to credit scoring in emerging African markets—paving the way for broader financial inclusion.

REFERENCES (Selected)

• Mwangi et al., Journal of Financial Inclusion, 2024
• World Bank, Digital Financial Inclusion Report, 2023
• Mokheleli & Museba, J. of Info. Systems & Informatics, 2023