

DOCUMENTATION

Step 1 (Data Generation):

This updated data generation creates a realistic dataset that reflects actual Ghana mobile money market patterns:

- **MTN Market Dominance:** 72% market share matching real-world data
- **Urban-Rural Split:** 70% urban, 30% rural customers reflecting population distribution
- **Realistic Transaction Patterns:** Higher transaction values and volumes in urban areas
- **Regional Economic Variations:** Based on actual Ghana economic data
- **Default Risk Patterns:** Slightly higher defaults in urban areas due to more credit usage

Step 2 (Data Exploration):

This simplified exploration creates clear, stakeholder-friendly visualizations:

Key Visualizations Created:

1. **Market Share Pie Chart** - Shows MTN's 72% dominance in simple terms
2. **Urban vs Rural Comparison** - Three clear charts showing customer distribution, default rates, and transaction values
3. **Risk Factors Analysis** - Direct comparison of key behaviors between good customers and defaulters
4. **Regional Performance** - Focused on top 8 regions for clarity, showing transaction values and default rates

Stakeholder Benefits:

- **Easy to understand charts without technical jargon**
- **Clear comparisons between different customer segments**
- **Focus on actionable business insights**
- **Professional styling suitable for presentations**
- **Direct answers to key business questions about market structure and risk**

Step 3 (Feature Engineering):

This step transforms the raw data into features that are meaningful for both machine learning models and business stakeholders.

Key Feature Engineering:

1. **Financial Health Score** - Combines multiple positive financial behaviors into one easy-to-understand score
2. **Digital Engagement Score** - Measures how actively customers use mobile money services
3. **Risk Flags** - Simple binary indicators for high-risk behaviors (easy for stakeholders to understand)
4. **Customer Value Score** - Estimates long-term customer value
5. **Urban Advantage** - Captures the benefits of urban location

Data Preparation:

- **Clear Feature Names:** Renamed technical terms to business-friendly language
- **Categorical Encoding:** Converted text categories (like region, gender) to numerical format
- **Data Splitting:** 70% for training, 30% for testing while maintaining default rate balance
- **Feature Scaling:** Standardized numerical features for better model performance

Stakeholder Benefits:

- Features have clear business interpretations
- Risk flags provide simple rules for manual review
- Scores are on intuitive scales (0-100 or 0-1)
- Maintains transparency in how decisions are made

Step 4 (Model Training):

This step trains and evaluates machine learning models for credit scoring, with a focus on business impact and interpretability.

Models Trained:

1. **Logistic Regression** - Provides interpretable coefficients and probability scores
2. **Random Forest** - Captures complex patterns and provides feature importance

Key Performance Metrics:

- **AUC-ROC:** Measures overall model performance (0.83 target from research)
- **Accuracy:** Overall prediction correctness
- **Precision/Recall:** Balance between identifying defaults and approving good customers
- **Business Impact:** Approval rates and default rates at different risk thresholds

Visualizations Created:

1. **ROC Curves** - Shows model performance across all thresholds
2. **Feature Importance** - Identifies which factors most influence credit decisions
3. **Confusion Matrix** - Details model predictions vs actual outcomes

Business Insights:

- Recommended risk threshold for optimal approval/default balance
- Comparison against traditional credit scoring methods
- Clear metrics on how many good customers would be approved vs risky ones

The Random Forest model achieved strong performance (AUC ~0.83) matching the research targets, with the ability to approve significantly more customers while maintaining acceptable default rates.

Step 5 (Business Deployment):

This step implements the credit scoring model for real-world business application and analyzes the financial impact.

Key Components:

1. **Prediction Engine**
 - Function to score new applicants
 - Converts probabilities to easy-to-understand credit scores (0-100)
 - Provides clear risk categories and recommendations
2. **Business Impact Analysis**

- Simulates different approval thresholds
- Compares against traditional credit scoring
- Calculates approval rates and default rates
- Identifies optimal risk threshold

3. **Implementation Roadmap**

- 4-phase deployment strategy matching research paper
- Gradual scaling from pilot to national implementation
- Clear timelines and milestones

4. **Stakeholder Visualizations**

- Approval vs default trade-off analysis
- Applicant distribution by credit score
- Comparison of traditional vs FinTech methods
- Regional impact assessment

Business Outcomes:

- **42% more loan approvals** while maintaining acceptable default rates
- **Clear credit decision framework** for loan officers
- **Implementation roadmap** for phased deployment
- **API-ready model** for integration with mobile money platforms

Step 6 (Monitoring and Retraining):

This step establishes a comprehensive monitoring and retraining system to ensure long-term model performance.

Key Monitoring Components:

1. **Performance Tracking**

- AUC scores, accuracy, and default rates over time
- Automated alerts when performance drops below thresholds
- Historical performance tracking

2. **Data Drift Detection**

- Monitors changes in feature distributions
- Alerts when data patterns change significantly
- Identifies which features are drifting

3. **Automated Retraining Assessment**

- Performance-based triggers ($AUC < 0.78$)
- Data drift triggers
- Time-based triggers (every 6 months as per research)
- Clear retraining recommendations

4. **Visual Monitoring Dashboard**

- Performance trends over time
- Data drift analysis
- Retraining schedule
- Alert summary

Retraining Process:

- **Assessment:** Automatic evaluation of retraining need
- **Validation:** New model performance validation
- **Versioning:** Proper model version control
- **Deployment:** Safe deployment of improved models

Business Benefits:

- **Proactive risk management** through continuous monitoring
- **Maintained model performance** through regular retraining
- **Early detection** of changing market conditions
- **Regulatory compliance** through documented model governance

The system ensures the credit scoring model remains effective as customer behaviors and economic conditions evolve, maintaining the 42% improvement in financial inclusion while controlling default risks.

Credit Scoring Application.

What The App Does

This is a fair credit scoring system that helps determine if someone in Ghana should get a loan, using only their mobile money transaction behavior - no personal information like age, gender, or location is considered.

Key Features

Fair and Unbiased

- No discrimination based on personal characteristics
- Everyone assessed by the same rules
- Uses only transaction patterns, not who you are

Transparent Scoring

- Clear explanation of how your score is calculated
- Shows exactly what factors affect your result
- Provides suggestions to improve your score

Easy to Use

- Simple sliders to input your transaction habits
- Instant results with clear recommendations
- Visual display of your risk level

How It Works

The system looks at three main things:

1. Transaction Amount - How much money moves through your account monthly
2. Transaction Frequency - How often you make transactions
3. Bill Payments - How consistently you pay bills on time

Plus Loan Affordability

- Checks if the loan amount is reasonable for your transaction level
- Calculates monthly payments you can handle
- Ensures you don't take on too much debt

Results You Get

Credit Score (0-100)

- 75-100: Low Risk - Loan Approved

- 50-74: Medium Risk - Approved with Conditions
- 0-49: High Risk - Loan Rejected

Plus You See

- Why you got that score
- What factors helped or hurt your application
- Specific tips to improve your score
- Full loan payment breakdown

Benefits

For Applicants

- Fair chance for everyone
- Clear understanding of decisions
- Privacy protection - no personal data needed

For Lenders

- Reliable risk assessment (79% accurate)
- More approvals while managing risk
- Compliance with fair lending practices

Perfect For

- Mobile money users in Ghana's informal sector
- People without traditional credit history
- Anyone wanting a fair, transparent loan assessment