INVESTOR PITCH PRESENTATION

PERSONAL FINANCE TRACKER

Empowering Users with Personalized Financial Insights and Actionable Recommendations.

PRESENTED BY

Group 1

DATE

February 2026

Data Extraction • Rule-Based Engine • Predictive Modeling

The Problem: Financial Blind Spots

Why traditional financial statements fail the modern user.

Categorization Chaos

Difficulty in categorizing diverse and irregular transactions.

Users struggle to group spending into meaningful buckets manually.

Missing Personalization

Traditional banking lacks personalized financial advice.

Statements provide raw data but no "next steps" tailored to individual goals.

Opaque Spending Habits

Inability to easily differentiate between essential needs and discretionary "wants," making budget optimization nearly impossible.

Static Hindsight

Limited tools for forward-looking financial planning. Statements tell you what happened, not what is likely to happen next.

"Complex transaction data remains a burden rather than an asset."

THE CLARITY GAP IN PERSONAL FINANCE

OUR MISSION

Turning Data into Intelligence

The Solution: Automated Data Extraction Pipeline

Transforming raw, password-protected M-Pesa PDF statements into high-fidelity, analysis-ready data structures.

Raw Input
Password-protected M-Pesa Statements

Extraction Engine
Multi-stage Tabula parsing & CSV cleanup

Analysis Ready
Structured data with engineered features

- **Scalable Data Handling**
  Successfully processed ~2,715 rows and 14 columns from a single statement with 100% accuracy.
- **Feature Engineering**
  Adds temporal (payday indicators) and behavioral (essential/discretionary flags) features.
- **Automated Normalization**
  Removes duplicates, standardizes column names, and formats dates for seamless analysis.
- **Robust Security**
  Handles encrypted PDFs securely within the local environment before conversion.

Transaction Identification & Categorization

A multi-stage intelligence layer that transforms raw strings into structured financial data.

Regex Identification

Scans transaction strings using strict priority patterns to identify types: **PayBill, Till, Pochi, M-Pesa Fees, and Fuliza Loans.**

Keyword Mapping

Assigns initial categories like **Health Care, Utilities, Betting, and Education** by matching merchant names against a curated keyword database.

Smart Logic Rules

Refines "Send Money" transfers based on **recurrence and amount thresholds** to distinguish between family support and merchant payments.

TRANSACTION TYPES

PayBill • Till • Pochi • Airtime • Fuliza • Bank Transfer • Cash Out

KEY CATEGORIES

Shopping • Food & Dining • Transport • Personal Care • Savings • Bills

LOGIC INDICATORS

Temporal (Payday) • Behavioral (Essential) • Frequency (Recurring)

Hybrid Merchant Learning System

A "Human-in-the-Loop" architecture that combines automated database matching with interactive user intelligence.

1. Automated Lookup

System scans the **Personal Merchant Database** for existing matches. Known merchants are auto-labeled instantly.

2. Interactive Query

For unknown merchants, the UI prompts the user for **one-time categorization**. No merchant remains unclassified.

3. System Learning

User feedback is **persisted back** to the database. The system gets smarter with every transaction analyzed.

Final Result

Refined **merchant_subcategory** column and 100% Analysis-Ready data for the engine.

Rationale for a Rule-Based Engine

Why deterministic logic outperforms ML in early-stage financial data analysis.

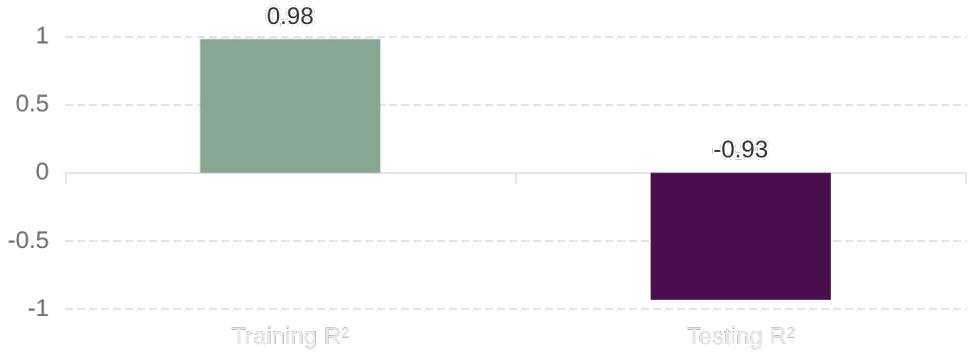
0.01
AUTOCORRELATION
Spending is event-driven, not routine-driven. No momentum for models to learn.

1.73
VARIATION COEFF.
Extremely irregular daily spend swings make averages meaningless targets.

~480
AVAILABLE ROWS
Stable ML requires 1,000+ rows. Current volume leads to severe overfitting.

58.2%
TRUE SAVINGS RATE
Rules correctly distinguish savings from spend; naive models fail this logic.

The Overfitting Challenge: Training vs. Testing



1 **Accuracy & Reliability**
Rules capture fundamental financial truths (like savings vs. spending) that models currently lack the data to discern.

2 **Immediate Explainability**
Users receive advice they can trust because the logic is transparent and tied to verified data categories.

3 **Robust Foundation**
The engine serves as a baseline that future ML models will eventually learn to optimize, not replace.

Core Architecture

The 5-Module Recommendation Engine Orchestration

SOURCE DATA
2,715 Transactions

SYSTEM OUTPUT
Ranked Insights & Health Score

Budgeting

Generates dynamic monthly targets per category based on historical consumption averages.

Savings Ops

Identifies discretionary "want" cuts and analyzes M-Pesa fee leakage for immediate optimization.

Behavioral

Highlights spending spikes during payday weeks (27.8% of spend) and weekend patterns.

Predictions

Analyzes current 30-day trends and burn rates to provide balance alerts and trend warnings.

Comparative

Benchmarks essential vs. discretionary spending ratios and assesses the 58.2% true savings rate.

ENGINE LOGIC

All modules process data simultaneously to produce a unified recommendations.json file.

1,104	58.2%	Ranked
CONSUMPTION ROWS ANALYZED	VERIFIED SAVINGS RATE	OUTPUT PRIORITIZATION

Delivering Actionable Financial Recommendations

The engine prioritizes high-impact actions based on verified spending patterns and real-time balance data.

CRITICAL ALERT

M-Pesa Balance Alert

Current balance (KES 776) covers ~0 days at current spending rate (KES 1,626/day).

ACTION: Top up M-Pesa or reduce daily spend immediately to prevent failed transactions.

SAVINGS OPPORTUNITY

Reduce Discretionary Spend

You spend KES 5,427/month on "wants". A 30% reduction targets KES 3,799/month.

IMPACT: Save KES 1,628/month (approx. KES 19,536 per year).

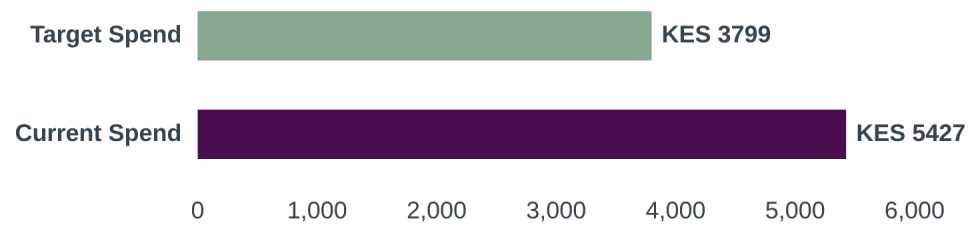
BUDGET TARGET

Construction Tracking

Largest consumption category at KES 9,734/month (19.7% of total consumption).

ACTION: Set a project end-date and monthly cap to prevent silent overshooting.

Monthly Savings Potential (KES)



Explainable Logic

Recommendations are derived from 2,715 verified transactions over 24 months, ensuring high confidence.

Action-Oriented

Every insight is paired with a specific target (e.g., 20% reduction in Subscriptions) to guide user behavior.

Predictive Modeling: A Hybrid Learning Journey

A two-layer architecture ensuring immediate accuracy while the AI matures.

STAGE 1: TODAY

Rule-Based Engine

- **Instant Logic:** Applies verified category rules and spending averages.
- **Pattern Detection:** Flags fee leakage and payday spikes immediately.
- **Explainability:** Delivers clear, ranked recommendations based on data facts.

Current Status: 100% Operational

STAGE 2: FUTURE

ML Prediction Layer

- **Deep Learning:** Trains in background on accumulating 1,000+ daily rows.
- **Seasonality:** Learns complex annual and monthly spending cycles.
- **Forecasting:** Predicts next-month spend to pre-emptively guide users.

Target Deployment: Sept 2028

The Goal: ML-powered predictions will feed into the proven rule engine for higher-calibration recommendations.

Current ML Status and Technical Challenges

Phase: Training & Data Accumulation

A transparent look at model performance and deployment readiness.

The Overfitting Challenge

Complex models (Gradient Boosting, Random Forest) are memorizing noise rather than learning patterns due to highly irregular spending behaviors.

AUTOCORRELATION	SPEND VARIATION
0.01	1.73 CV

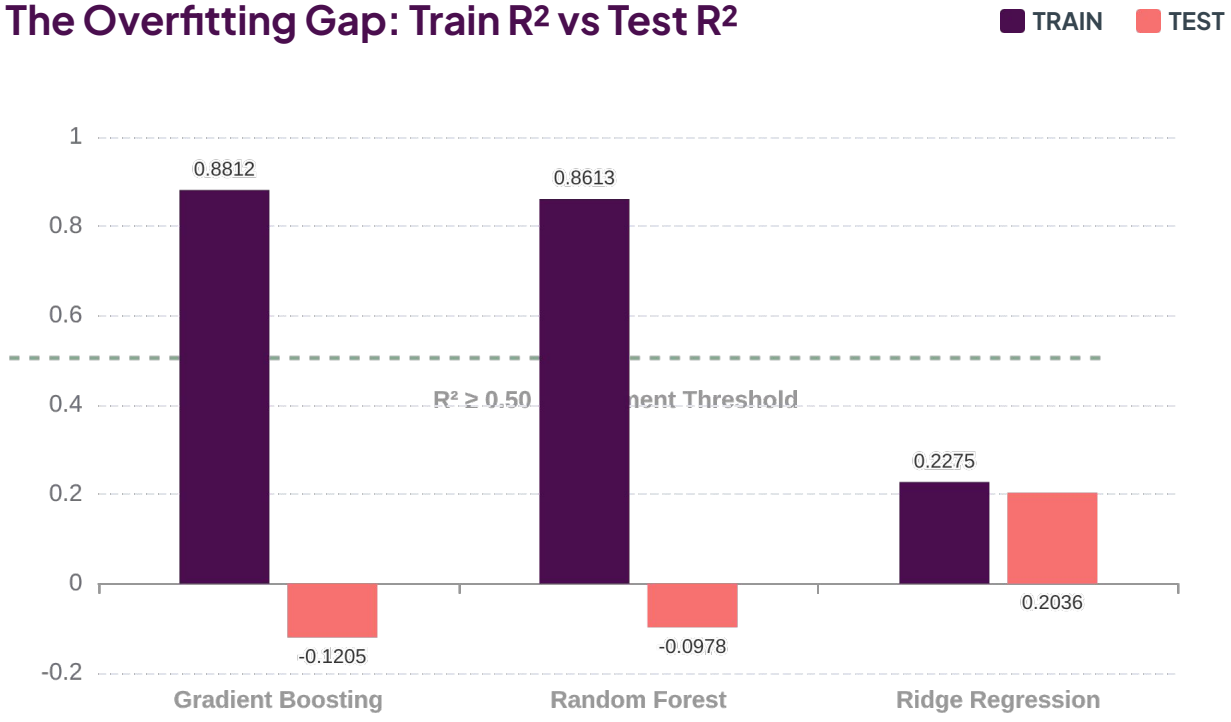
Data Requirements

Current Clean Rows **450**

Deployment Goal **1,000+**

**Estimated readiness: September 2028 based on current data velocity.*

The Overfitting Gap: Train R^2 vs Test R^2



Ridge Regression selected as current baseline due to positive generalization (+0.20 Test R^2).

ROADMAP TO MACHINE LEARNING DEPLOYMENT

Strategic data accumulation to reach the deployment threshold of 1,000 clean daily rows.

Key Milestones

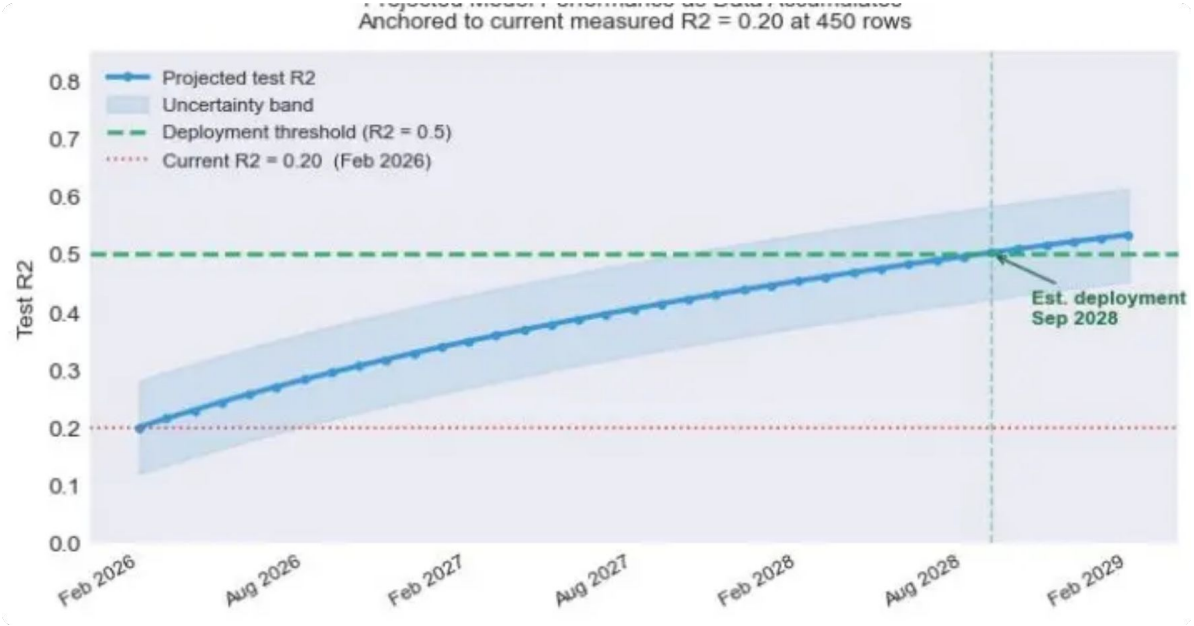
FEB 2026 (NOW)
450 Clean Rows
Core Rule-Based Engine Active

AUG 2026 (+6m)
~570 Clean Rows
Catching Seasonal Patterns

FEB 2027 (+12m)
~690 Clean Rows
Second Full Year Learnable

SEP 2028 (+31m)
1,000+ Clean Rows
Deployment Threshold ($R^2 > 0.50$)

PROJECTED MODEL PERFORMANCE



ML layer activates as a 30-day forward projection to the trend chart once the threshold is met.

Until deployment, the **Rule-Based Engine** remains fully operational and explainable.

Empowering Users via Interactive UI

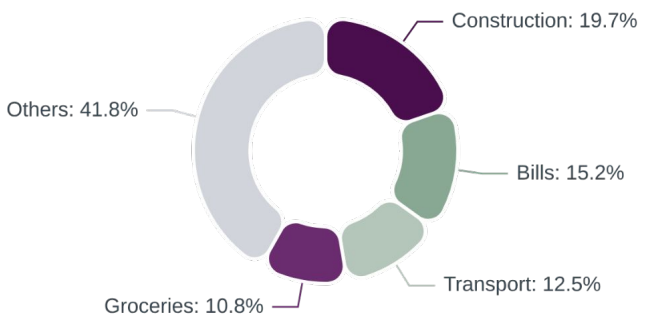
A seamless, browser-based dashboard designed for intuitive financial management.

Visual Analytics
4 dynamic charts tracking trends, categories, and savings potential.

Smart Money Rules
User-defined spending constraints applied before analysis runs.

Merchant Learner
Interactive interface to confirm or correct merchant auto-categorization.

Sample Analysis: Consumption Mix



The Seamless Journey

1. Upload password-protected M-Pesa PDF
2. Set Personal Spending Rules
3. Engine executes 6 Intelligence Modules
4. Verify & Train Merchant Classifier
5. Access Ranked Recommendation Cards
6. Deep-dive into Tool & Merchant Directory

UI THEME

M-Pesa Green + Plus Jakarta Sans

Quantifying Financial Health & Savings

Turning raw data into measurable financial outcomes

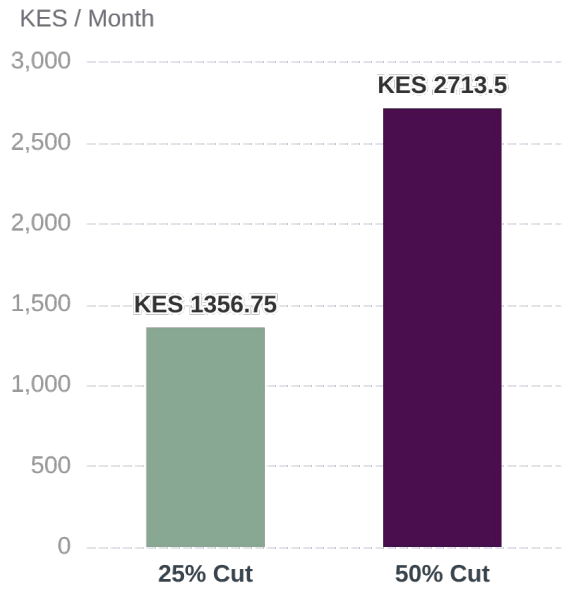
FINANCIAL HEALTH SCORE

2.73

Wealth Builder

"Your income is 2.7× your consumption spend, placing you in the top tier of financial stability."

Monthly Savings Potential



*Based on KES 5,427 monthly discretionary spend

Performance Highlights

58.2%
True Savings Rate

KES 1,357
Minimum monthly optimization

2.7×
Income to Consumption Ratio

Full System Architecture

"Privacy-centric financial intelligence platform with local-first processing and zero-storage policy for sensitive data."

Data Input Layer

M-Pesa PDF/CSV

Password-protected statements with 2,715+ transactions over 24 months

Feature Engineering

36 derived columns including temporal patterns and behavioral flags

Privacy Layer

Local-first processing with zero cloud storage of PII

Recommendation Engine

- budget_recommendations()
- savings_opportunities()
- behavioral_insights()
- spending_predictions()
- comparative_analysis()
- financial_health_score()

ML Layer (Training)

Ridge Regression ($R^2 = 0.2036$)
Deploys when $R^2 > 0.50$

User Interface

Interactive Dashboard

4 charts, 3 recommendation cards, period filters

8-Screen Journey

Upload → Rules → Processing → Learning → Analytics

JSON Output

Structured recommendations with priority ranking

Data Flow ML Training Active Components

Thank You

Thank you for your time and consideration. We look forward to discussing how we can revolutionize personal financial management together.