Project Proposal

M-Pesa Financial Analyzer: An Intelligent Personal Finance Management Platform

Executive Summary

M-Pesa is the dominant mobile money service in Kenya, processing billions of transactions annually. While users have access to their transaction history via PDF statements, extracting meaningful financial insights from these raw documents is incredibly difficult.

The M-Pesa Financial Analyzer aims to bridge this gap by providing an automated, secure, and intelligent platform that transforms raw M-Pesa statements into actionable financial insights, interactive visualizations, and personalized savings recommendations.

Problem Statement

Despite the reliance on mobile money, most users lack visibility into their spending habits.

Data Obscurity: Official M-Pesa PDF statements are tedious to read and manually categorize.

Lack of Insights: Users struggle to track burn rates, identify unnecessary subscriptions, or separate essential living expenses from discretionary spending.

Static Tools: Existing budget apps require manual entry, which introduces friction and leads to high abandonment rates.

Proposed Solution

A web-based financial intelligence platform that allows users to securely upload their M-Pesa statements and instantly receive a comprehensive financial dashboard.

The application leverages a hybrid data pipeline combining Rule-Based heuristics (for deterministic categorization like utilities and transfers) and Machine Learning (for pattern recognition and transaction forecasting).

Key Features & Capabilities

- Intelligent Analytics Pipeline

High-Fidelity PDF Extraction: Decrypts and parses official M-Pesa password-protected statements with extreme accuracy using Pandas and custom Regex pipelines.

Explainable AI (XAI): A robust recommendation engine provides transparent financial advice on budgeting, savings, and behavioral patterns.

Smart Money Rules: Allows users to configure custom logic for "Send Money" transactions (e.g., classifying frequent transfers under KES 500 as "Friends & Family").

- Continuous Merchant Learning (Memory System)

Interactive Training: The system automatically flags "Unknown" PayBill or Till numbers. Users categorize them once via a dynamic UI.

Persistent Cloud Memory: Categorizations are saved securely to Google Firestore. Future uploads automatically recognize these merchants, creating a seamless, "zero-touch" categorization experience over time.

- Real-Time Dynamic Dashboard

Visual Analytics: Interactive Chart.js visualizations detailing monthly spending vs. income, category breakdowns, and payday burn rates.

Actionable Insights: Highlights specific savings opportunities based on user spending velocity.

- Enterprise-Grade Security

Multi-Mode Authentication: Secure sign-in supporting both Google OAuth and Email/Password flows via Firebase Auth.

Ephemeral Processing: PDF files are processed in-memory and scrubbed immediately after analysis, ensuring total data privacy.

Target Audience

Everyday M-Pesa Users: Individuals looking to take control of their budget and cut down on unnecessary spending.

Freelancers & Small Business Owners: Users who mix personal and business transactions on a single number and need help separating expenditures.

Technology Stack

Frontend: Vanilla JavaScript (ES6+), CSS3 (Glassmorphism design), HTML5, Chart.js 4.4

Backend: FastAPI (Python 3.10+)

Data Science/ML: Pandas, NumPy, Scikit-Learn (Ridge Regression, Gradient Boosting, Random Forest)

Infrastructure & Auth: Firebase Authentication, Google Firestore, Docker

Deployment: Hugging Face Spaces (Containerized Cloud Hosting)

Expected Outcomes & Impact

By the end of the development phase, the platform will successfully:

Parse and categorize at least 95% of standard M-Pesa transactions.

Provide users with an intuitive, visually appealing dashboard that clearly illustrates their financial health.

Demonstrate a working continuous learning system where the app becomes progressively smarter at identifying merchants based on user input.

This capstone project will serve as a strong demonstration of full-stack engineering, secure cloud architecture, automated data pipelines, and applied machine learning.