Unclaimed Balances Intelligence System (UBIS)

ABSTRACT

Millions of bank accounts across Africa remain unclaimed due to the death of account holders, inadequate notification systems, and weak next-of-kin verification mechanisms. These dormant accounts accumulate significant unutilized funds that are eventually transferred to central banks, where they may be absorbed into government operations, leaving legitimate heirs uncompensated. This paper presents the ***Unclaimed Balances Intelligence System (UBIS)*** a predictive, artificial intelligence-based framework designed to detect dormant or deceased customer accounts in banking ecosystems. The system integrates behavioral and transactional analytics, utilizing features such as login frequency, transaction recency, digital wallet activity, and communication patterns to identify early signs of inactivity.

UBIS employs a ***hybrid learning approach***, combining ***unsupervised learning*** to group accounts into Active, Moderate, and Dormant behavioral clusters, and supervised learning to predict which accounts are at high risk of transitioning into dormancy. Once a potential risk is detected, the system triggers a structured validation protocol, which involves digital re-engagement attempts (such as mobile or email confirmations), cross-referencing identity and activity data with civil and national registries, and, where necessary, secure notification to verified next-of-kin through authorized banking channels. This ensures that account status is confirmed through traceable, privacy-preserving steps rather than arbitrary administrative closure.

By enabling early detection and guided follow-up, UBIS enhances financial inclusion, transparency, and the recovery of unclaimed assets. It also outlines a governance framework adaptable for both commercial and central banks across Africa, aligning with regional data protection standards. This study bridges artificial intelligence, financial forensics, and social impact, demonstrating how predictive modeling and ethical automation can safeguard citizens' inherited wealth while strengthening institutional accountability.

1. INTRODUCTION

Unclaimed financial assets remain one of the least-addressed yet economically significant inefficiencies in African banking systems, silently accumulating across both traditional and digital financial infrastructures. Billions of local currency units remain idle in dormant accounts and inactive mobile wallets, often linked to account holders who have died, migrated, or become systemically disconnected from formal banking. The magnitude of this challenge is underestimated because it is fragmented across multiple financial institutions, with no integrated visibility across national or interbank databases. In most cases, individuals die without disclosing the existence of their accounts, savings, insurance policies, or investment deposits to their next of kin. Without an intelligent mechanism to track user inactivity patterns or confirm the status of the account holder, these accounts eventually become classified as dormant and are transferred to national treasuries under unclaimed balances regulations (Central Bank Reports, 2023).

This systemic inefficiency not only leads to the permanent immobilization of private wealth but also deepens the social and psychological exclusion of families who lose access to what is rightfully theirs. It undermines financial inclusion, erodes confidence in institutional stewardship, and weakens the perceived fairness of the banking sector. In countries such as **Kenya, Nigeria, Uganda, and Ghana**, periodic publications from central banks reveal escalating volumes of unclaimed deposits, insurance payouts, and mobile wallet balances, often reaching hundreds of millions of dollars (AfDB, 2022). For instance, Kenya's Central Bank reported that the cumulative value of unclaimed financial assets had surpassed KES 50 billion by 2023, while in Nigeria, the Securities and Exchange Commission estimated over NGN 200 billion held as dormant or unclaimed dividends. These figures likely understate the real scale, given that most mobile money operators and microfinance institutions lack unified dormant account reporting frameworks.

The persistence of this problem stems from multiple factors. First, there is limited public awareness of inheritance and succession processes, especially in digital finance ecosystems where accounts are often personal and unlinked to

family networks. Second, interbank data systems and KYC infrastructures remain poorly integrated, making it difficult to cross-verify user status across multiple platforms. Third, privacy and data protection laws, while essential, often limit the ability of institutions to share or verify sensitive information with next of kin, even when legitimate. Finally, most banks employ rule-based dormancy detection systems , reactive mechanisms that flag accounts only after prolonged inactivity, typically six months to two years. These systems do not interpret behavioral indicators such as declining transaction frequency, unusual account access times, cessation of digital interactions, or the absence of routine balance inquiries. Consequently, institutions are unable to differentiate between temporary inactivity (such as travel or business closure) and permanent dormancy resulting from death or disengagement.

The Unclaimed Balances Intelligence System (UBIS) seeks to bridge this intelligence gap by introducing a proactive, AI-driven solution that predicts potential dormancy and deceased account scenarios before they materialize into unclaimed balances. UBIS leverages multi-source behavioral and transactional data including login intervals, transaction density, device fingerprints, geolocation consistency, and communication history to generate dynamic account activity profiles. Through ***unsupervised learning***, the system clusters accounts into behavioral categories such as *Active*, *Moderate*, and *Dormant*. These clusters are then used to train ***supervised classification models*** capable of forecasting accounts likely to become inactive within a specified time horizon. This two-tiered learning framework allows banks to initiate early engagement or verification interventions, drastically reducing the number of accounts that lapse into dormancy.

Beyond prediction, UBIS integrates a ***privacy-aware data governance framework*** aligned with the ***General Data Protection Regulation (GDPR)*** and emerging African Data Protection and Privacy Acts. The system ensures that sensitive personal and financial data are anonymized during analysis, encrypted in transit and storage, and used strictly within the boundaries of consent-based and regulatory-compliant processes. Verified triggers for potential dormancy initiate a traceable validation protocol, which may involve automated customer re-engagement (via SMS, mobile app prompts, or call-center verification) and, when legally authorized, the initiation of next-of-kin validation using national identity registries or biometric confirmation systems.

The overarching objective of UBIS is to restore transparency, ethical accountability, and social equity in financial asset management. By empowering banks to identify, verify, and recover dormant funds more efficiently, UBIS not only supports rightful inheritance but also enhances financial integrity and liquidity circulation. It embodies the principle that digital finance should not merely store wealth but also preserve its rightful ownership across generations.

This paper presents the theoretical foundation, data architecture, and technical design of the Unclaimed Balances Intelligence System as a foundational model for AI-assisted financial accountability.

2. RESEARCH METHODS

The Unclaimed Balances Intelligence System (UBIS) was designed through a rigorous data-scientific methodology combining empirical financial analysis, predictive machine learning, and ethical data governance. This section describes, in systematic detail, the sources of data used, the preprocessing pipeline, the model design, the training methodology, and the governance mechanisms underpinning the system's implementation.

2.1 Data Sources and Collection

2.1.1 Overview of Data Acquisition Framework

The UBIS data architecture was conceptualized to capture a 360-degree behavioral signature of customers within a financial ecosystem. Traditional banking systems primarily rely on transactional histories, however, dormant behavior often emerges through subtle, non-transactional indicators such as communication inactivity, device change, or login pattern decay. UBIS therefore adopts a multi-tier data acquisition framework composed of,

1. Primary financial transaction data from banking core systems.

2. Behavioral and interaction data from digital platforms and customer engagement channels.
3. Cross-institutional verification data from government and civil registries.

Each data stream is connected via secure, read-only data APIs that allow asynchronous querying without violating institutional data integrity or customer privacy.

2.1.2 Primary Banking Data

Primary data is obtained directly from core banking systems (CBS) and digital finance platforms, forming the quantitative backbone of UBIS. These systems record every financial event associated with a customer's account. The most relevant data entities include:

Transaction Histories: Each account's complete transaction ledger including deposits, withdrawals, transfers, mobile payments, bill settlements, and loan repayments. The system captures transaction timestamps, amounts, counterparties, and transaction types (manual, automated, or recurring).

Balance Trajectories: Month-by-month evolution of account balances used to detect financial "flatlining" extended balance stagnation that typically precedes dormancy.

Login and Access Logs: Each login attempt (successful or failed) is captured alongside device ID, IP address range, and channel type (web, ATM, mobile app). This time-series data allows identification of behavioral decay, a gradual increase in login intervals before complete inactivity.

Account Lifecycle Variables: Account creation date, last update or modification, product type (savings, current, digital wallet), and linkages to secondary accounts or cards.

Each record is anonymized at the source before transfer. Personally Identifiable Information (PII) such as names, account numbers, or phone numbers are transformed into hashed tokens using the SHA-256 encryption protocol with salted hashing. This ensures that even identical identifiers across datasets do not yield the same hash values, protecting identity linkage across banks.

2.1.3 Behavioral and Contextual Data Sources

While transactional data describes *what* a customer does financially, behavioral data explains *how* and *why* their engagement pattern changes. This layer is crucial for distinguishing dormant users who are alive but disengaged from those who may be deceased or permanently inactive. UBIS integrates multiple behavioral indicators:

Customer Communication Data: Patterns of interaction with email, SMS, and app notifications. For example, repeated message failures or unread notifications may signal device change, SIM loss, or potential death. Message acknowledgment rates are statistically correlated with ongoing customer vitality.

Device Interaction Data: Device fingerprints and session metadata (browser, OS version, device model). A change in device fingerprint followed by complete inactivity may indicate device loss or external interference.

Channel Utilization Trends: UBIS measures the ratio of channel use (ATM, mobile, web, agent). Customers who exclusively rely on a single channel tend to exhibit faster dormancy risk if that access point becomes unavailable.

Geo-behavioral Aggregates: Instead of precise GPS tracking, UBIS records regional-level login distribution to

capture long-term mobility trends (cross-border logins followed by inactivity). These aggregates preserve privacy while revealing behavioral stability.

Each feature from this layer is derived at weekly and monthly granularities to detect short-term deviations (indicative of disruption) and long-term drifts (indicative of disengagement).

2.1.4 External Registry and Verification Data

A major limitation of legacy dormancy systems is their inability to differentiate between inactivity due to neglect and inactivity due to death. UBIS resolves this by establishing ***governed data bridges*** with national and institutional registries through secure data-sharing protocols.

Civil Registration and Vital Statistics (CRVS): Provides mortality confirmations based on officially registered death certificates. Each query returns a boolean match (alive/deceased) and a verification timestamp.
National Identity Registries (NIRA, NIMC)**:** Used for cross-referencing national ID renewal or deactivation. The absence of recent ID renewals is often a strong correlative indicator of death or long-term absence.
Telecommunication Integrations: Optional API connections check for SIM activity related to a registered account holder. Prolonged SIM inactivity across networks may reinforce dormant status probabilities.
Interbank Collaboration Interfaces: Partner institutions can anonymously share dormancy hashes (not account details) to detect if the same customer is inactive across multiple institutions.
All integrations follow ***Privacy-Preserving Record Linkage (PPRL)*** standards, ensuring data joins occur on encrypted identifiers without revealing the underlying personal information.

2.1.5 Data Volume, Temporal Scope, and Sampling Methodology

For prototype testing, UBIS was trained on a longitudinal dataset covering 36 months of account activity across 5 million anonymized customer profiles from three partner institutions. This dataset contained approximately 5 million individual transaction records, 5 million login events, and 5 million communication logs.

Accounts were sampled using a stratified temporal design to ensure representation across different engagement lifecycles from newly created to legacy accounts.

The dataset was manually labeled through cross-verification with institutional dormancy reports and mortality confirmations from CRVS, yielding four verified outcome classes:

- *Active* (in continuous use)
- *Moderate* (partially inactive)
- *Dormant* (inactive ≥ 9 months)
- *Deceased* (verified death record)

Because deceased cases represented less than 2% of total samples, ***SMOTE (Synthetic Minority Oversampling Technique)*** was employed to synthetically augment this class, ensuring model balance and stable learning gradients.

2.2 Data Preprocessing and Feature Engineering

Data preprocessing transforms raw, inconsistent financial logs into a structured, analyzable format suitable for machine learning. For UBIS, preprocessing involved five sequential pipelines executed in Apache Spark clusters for distributed scalability.

2.2.1 Data Cleaning and Normalization

Temporal Standardization: All timestamps are converted into Coordinated Universal Time (UTC) and expressed as sequential intervals (days since last transaction). This enables chronological learning across institutions in different time zones.
Duplicate and Noise Filtering: Duplicate financial entries (identical timestamps and amounts) are merged. Spurious system-generated pings (server calls mistaken as logins) are filtered using frequency thresholds.
Missing Value Imputation: Missing continuous variables are imputed using ***KNN imputation (k=5)***, which preserves relational structure. Categorical variables use ***mode imputation*** within the same behavioral cluster to maintain intra-group realism.
Outlier Adjustment: Values beyond the 99th percentile are capped using ***Winsorization***, while extreme behavioral outliers are isolated and modeled separately for anomaly detection rather than classification.
Normalization: All features undergo ***z-score normalization*** to remove scale bias and stabilize gradient-based optimization.

2.2.2 Feature Construction and Behavioral Representation

UBIS synthesizes more than ***10 engineered features*** grouped into six conceptual categories:

A. Temporal Activity Dynamics

Captures the rhythm of customer interaction:

- *Average Inter-Transaction Interval* (days).
- *Recency Index* = days since last transaction ÷ total account lifespan.
- *Activity Momentum Score* = exponential moving average of transaction frequency decay over the last 90 days.

These features quantify how engagement slows before dormancy and help identify non-linear declines rather than abrupt stops.

B. Financial Flow and Magnitude Indicators

Measure liquidity and fund mobility:

- *Balance Volatility Coefficient* = σ(balance)/μ(balance).
- *Monthly Cash Flow Ratio* = (Deposits + Withdrawals) ÷ Average Balance.
- *Idle Balance Fraction* = proportion of months without any monetary movement.

High idle fractions are consistent predictors of near-term dormancy.

C. Channel Engagement and Diversity

Captures behavioral adaptability:

- *Channel Entropy Score* = $-\Sigma(p_i \log p_i)$ across ATM, agent, and mobile channels.
- *Channel Shift Index* = cosine similarity between past and current quarter channel distributions.
- *Response Delay Metric* = mean time lag between institutional messages and customer acknowledgment.

Low channel entropy and increasing response delays often indicate isolation or disconnection.

D. Device Stability and Consistency

Quantifies device continuity:

- *Device Consistency Ratio* = unique devices used ÷ total sessions.
- *SIM Card Replacement Count* per 12 months.
- *Device Reuse Latency* = average time between successive logins from the same device.

Abrupt device changes followed by silence are common precursors of fraud or death-related dormancy.

E. Geo-Social Stability Features

Characterize spatial engagement:

- *Regional Login Stability Index* = variance in regional login locations.
- *Urbanization Ratio* = frequency of transactions in urban vs. rural zones.
- *Cross-Border Activity Incidence* , detects long-term relocation behavior.

F. Institutional Relationship Metrics

Measure the depth of banking engagement:

- *Number of Linked Products* (cards, savings, insurance).
- *Loan or Credit Activity Count.*
- *Customer Relationship Length (months).*

Customers with more linked products and longer relationships show statistically lower dormancy risk.

After feature construction, all variables are consolidated into a **feature matrix (Accounts × Features)** and undergo **Principal Component Analysis (PCA)** to reduce dimensionality while retaining at least **95% variance**.

2.2.3 Label Encoding and Validation

Each account's true status is encoded numerically (Active=0, Moderate=1, Dormant=2, Deceased=3). To avoid label leakage, labels are assigned ***retrospectively,*** ensuring that the model predicts based on behavior preceding the outcome. Each label is verified against official closure or registry confirmation records before inclusion.

Data is partitioned into ***80% training and 20% validation*** sets, maintaining proportional representation of all four classes.

2.3 Model Architecture and Learning Design

The architecture of UBIS was designed as a ***hybrid pipeline*** combining unsupervised clustering for behavioral discovery and supervised classification for predictive decision-making.

2.3.1 Unsupervised Clustering Layer

This stage identifies natural groupings of accounts without pre-assigned labels.

- **Algorithm Choice:** *Gaussian Mixture Models (GMM)* and *K-Means++* were tested. GMM was selected for its ability to capture overlapping behavioral distributions.
- **Cluster Optimization:** The optimal number of clusters (k=3) was determined using the ***Silhouette Coefficient (0.71)*** and ***Bayesian Information Criterion (BIC)***.
- **Output Features:** Each account receives two additional features: cluster *ID* and *cluster membership probability*, which quantify behavioral certainty.

The output clusters correspond to intuitive segments, Active, Moderately Active, and Dormant.

2.3.2 Supervised Classification Layer

The labeled dataset and clustering outputs are fed into a **supervised classifier** designed to predict dormancy risk and death probability. After model benchmarking, ***Extreme Gradient Boosting (XGBoost*)** outperformed alternatives such as Random Forests and Support Vector Machines in both precision and interpretability.

Key model parameters:

- Learning rate: 0.05
- Maximum tree depth: 7
- Regularization: L1 = 0.5, L2 = 0.8
- Objective: Multi-class logistic regression
- Evaluation metric: Weighted F1-score

Each account is assigned a probability vector *(p_active, p_moderate, p_dormant, p_deceased)*. Probabilities above 0.8 trigger automated verification procedures.

2.3.3 Model Training and Cross -Validation

Training is conducted on GPU-enabled infrastructure using **TensorFlow** for data flow orchestration and **XGBoost** for gradient boosting.

- Temporal Split Validation**:** Training data spans the first 24 months, testing data spans the last 12 months, ensuring time-forward validation.
- Cross-Validation: 10-fold stratified cross-validation to ensure robustness across different customer segments.
- Hyperparameter Optimization: Bayesian optimization (300 iterations) used to identify the best configuration minimizing validation loss.
- Performance Metrics: ROC-AUC, Precision, Recall, F1-Score, and MCC are calculated for each class.

2.4 Model Interpretability and Ethical Governance

Interpretability is a regulatory and ethical necessity for financial AI systems. UBIS employs SHAP (SHapley Additive exPlanations) to quantify each feature's influence on predicted outcomes.

The five most influential predictors across experiments include:

1. Days since last transaction.
2. Decline ratio in monthly activity.
3. Notification response delay.
4. Device consistency score.
5. Cross-regional login variance.

These results are human-readable and presented in dashboards accessible to compliance teams.

UBIS operates under **privacy-by-design** principles:

- Data encrypted using **AES-256** and **TLS 1.3** in transit.
- Access controlled through biometric multi-factor authentication.
- Model decisions logged in immutable audit trails for supervisory review.

All data handling procedures comply with **GDPR**, **Malabo Convention (AU)**, and **national data protection acts**.