



(REVIEW ARTICLE)



Leveraging NLP and graph-based machine learning for behavioral and transactional big data analytics in the Nigerian retail banking sector

Abdulrahman Balogun *

Computer Science Department

ORCID ID - <https://orcid.org/0009-0004-5585-3059>

World Journal of Advanced Research and Reviews, 2026, 30(03), 963-973

Publication history: Received on 01 May 2026; revised on 10 June 2026; accepted on 12 June 2026

Article DOI: <https://doi.org/10.30574/wjarr.2026.30.3.1651>

Abstract

The rapid digitalisation of Nigerian retail banking has generated unprecedented volumes of behavioural and transactional data, exposing the inadequacy of conventional rule-based and single-modality machine learning systems. Existing approaches—applied exclusively to structured tabular data—fail to exploit unstructured behavioural signals in customer communications or complex relational patterns in transaction networks. This paper proposes, implements, and empirically validates the NLP-GNN Integrated Analytics Framework for Nigerian Banking (NIAF-NB), the first integrated framework to combine Natural Language Processing (NLP) and Graph-Based Machine Learning (GBML) for multimodal analytics in Nigerian retail banking. The framework comprises three core components: (1) Nigerian FinBERT, a BERT-based language model adapted for Nigerian banking text, including Nigerian Pidgin English; (2) a Heterogeneous Graph Attention Network (HAN) modelling the transaction ecosystem as a multi-type relational graph; and (3) a Joint Cross-Attention (JCA) fusion mechanism that integrates NLP-derived semantic representations with GBML-derived topological representations. Evaluated on an anonymised dataset of approximately 1.53 billion transactions and 102 million customer interactions from three tier-1 Nigerian commercial banks over a 12-month period, NIAF-NB significantly outperforms all baselines across three tasks: customer churn prediction (AUC-ROC 0.967), credit risk assessment (AUC-ROC 0.893; 0.887 for thin-file customers), and Anti-Money Laundering (AML) detection (Precision 0.71, vs. 0.08 for existing rule-based systems). Ablation studies confirm the incremental contribution of each component. Fairness evaluation demonstrates equitable outcomes across demographic subgroups. The framework presents a deployable blueprint for Nigerian banks seeking to move from fragmented data silos to integrated, multimodal intelligence.

Keywords: Natural Language Processing; Graph Neural Networks; Nigerian Retail Banking; Financial Inclusion; Anti-Money Laundering; Customer Churn; Credit Risk; Multimodal Machine Learning; FinBERT; Heterogeneous Graph Attention Network; Big Data Analytics; Ajo/Esusu.

1. Introduction

The Nigerian retail banking sector stands at a remarkable inflection point. Over a single decade, the number of electronic payment transactions grew from approximately 1.5 billion in 2018 to over 9.7 billion in 2023—a more than six-fold increase—while their total value expanded from ₦89 trillion to over ₦387 trillion (NIBSS, 2023). The Bank Verification Number (BVN) system (2014), the CBN Cashless Policy (2012, expanded subsequently), USSD banking, and a vibrant FinTech ecosystem collectively transformed the sector, generating a data ecosystem of unprecedented richness and variety. Every digital interaction—transfer, chatbot query, POS payment, failed login—creates a data record that, if properly analysed, can illuminate customer health, risk, and behaviour.

* Corresponding author: Abdulrahman Balogun Email: abnbalo@gmail.com

This data ecosystem is inherently dual-natured. Structured transactional data captures the who, what, when, and how of financial activity and is amenable to conventional machine learning. Unstructured behavioural data—chatbot transcripts, transaction narrations, call centre notes—captures emotional states, intent, and context that transactional records cannot. A customer who types "abeg I wan close my account, una service too bad" (Nigerian Pidgin: "please I want to close my account, your service is terrible") is signalling churn with an explicitness no declining transaction metric can match. Yet the vast majority of Nigerian banks continue to rely on rule-based systems and single-modality models that process only structured data, leaving this informational wealth entirely untapped.

Compounding this, the linguistic landscape of Nigerian banking is uniquely complex. Customers communicate in Standard English, Nigerian Pidgin English (spoken by an estimated 75 million people as a second language), and various local languages. Transaction narrations such as "ajo contribution week 7" or "esusu pool payment" carry specific cultural meaning invisible to models trained on Western English corpora. Informal savings groups—Ajo (Yoruba) and Esusu (Igbo), rotating savings and credit associations (ROSCAs)—generate circular transaction patterns that superficially resemble money-laundering smurfing and produce unacceptable false-positive rates in generic detection systems.

This paper addresses the gap by proposing the NIAF-NB framework, which integrates NLP and Graph-Based Machine Learning (GBML) into a unified multimodal architecture. The central premise is that language and network topology are complementary: NLP reveals what customers communicate, GBML reveals how capital flows and how entities are connected. Together, they produce a richer analytical picture than either alone. The paper makes five principal contributions: (1) the NIAF-NB framework itself; (2) Nigerian FinBERT, the first language model adapted for Nigerian banking text; (3) the Joint Cross-Attention fusion mechanism; (4) the Ajo/Esusu graph schema for modelling informal savings groups; and (5) the Behavioural-Topological Complementarity Principle, which theorises why multimodal integration produces synergistic gains.

2. Literature review

2.1. Digital Transformation of Nigerian Banking

The 2004–2005 banking consolidation exercise reduced Nigeria's licensed commercial banks from 89 to 25 and created a smaller number of better-capitalised institutions capable of large-scale technology investment. The subsequent introduction of USSD banking extended digital financial services to feature-phone users without internet access, generating a layer of structured transactional data previously non-existent for millions of Nigerians. The Cashless Policy (2012) incentivised electronic payments; the BVN system (2014) created a biometric-linked customer node, effectively founding the graph that the GBML module exploits; and the National Financial Inclusion Strategy (revised 2022) established targets that motivate alternative credit scoring. Nigeria's FATF grey-listing in February 2023 added regulatory urgency to AML improvement (FATF, 2023). Despite these advances, approximately 38% of Nigerian adults remain financially excluded (World Bank, 2022), and false-positive rates in existing AML systems reach 90–92%.

2.2. NLP in Financial Services

Financial NLP evolved from Loughran-McDonald dictionary-based sentiment scoring (Loughran & McDonald, 2011) through word embedding models (Mikolov et al., 2013; Pennington et al., 2014) to the Transformer-based FinBERT (Araci, 2019; Yang et al., 2020). BloombergGPT (Wu et al., 2023) scaled the paradigm to 50 billion parameters but at prohibitive inference cost for production banking pipelines. Research demonstrates that NLP-derived sentiment from customer interactions is a leading indicator of churn (Li, 2019), and that borrower language predicts default risk in peer-to-peer lending (Ge et al., 2017). However, all published financial NLP work assumes Standard English; Nigerian Pidgin English—with its distinctive vocabulary, code-switching, and cultural idioms—has received no attention in the financial NLP literature.

2.3. Graph-Based Machine Learning in Banking

Graph Neural Networks (GNNs) emerged from the foundational work of Scarselli et al. (2009), Graph Convolutional Networks (Kipf & Welling, 2017), GraphSAGE (Hamilton et al., 2017), and Graph Attention Networks (Veličković et al., 2018). Heterogeneous Graph Attention Networks (HAN; Wang et al., 2019) extended the paradigm to graphs with multiple node and edge types—critical for banking, where customers, accounts, merchants, and devices interact through semantically distinct relationships. Weber et al. (2019) demonstrated GCN superiority over Random Forests for Bitcoin AML detection; Liu et al. (2021) extended this to credit card fraud; Yao et al. (2022) applied heterogeneous GNNs to credit risk assessment. Social contagion research (Dasgupta et al., 2008) established that a customer's churn

probability is significantly elevated when network neighbours churn—a finding directly motivating the graph-based churn module.

2.4. Multimodal Integration and Research Gap

Multimodal financial analytics has demonstrated that combining text with structured data consistently outperforms unimodal approaches for stock prediction (Xu & Cohen, 2018; Sawhney et al., 2020). PwC (2023) highlighted "Graph LLMs" as a frontier for banking, but this remains conceptual. Four critical gaps persist in the literature: (i) no published framework integrates Transformer-based NLP with HGNNs for retail banking analytics; (ii) no principled attention-based fusion mechanism for semantic and topological embeddings exists; (iii) the Nigerian—and broader African—banking context has been entirely neglected; and (iv) no study evaluates a unified framework across multiple banking tasks simultaneously. NIAF-NB directly addresses all four gaps.

3. Theoretical framework

3.1. Three complementary theoretical traditions justify the integrated design of NIAF-NB.

Behavioural Finance Theory (Kahneman & Tversky, 1979; Thaler, 1985; Taffler & Tuckett, 2010) establishes that financial decisions are driven by emotional and cognitive states—loss aversion, mental accounting, emotional finance. These states are expressed in customer communications before they manifest in transaction records. This provides the theoretical justification for the NLP module: NLP-derived sentiment and intent signals are leading indicators of churn, credit distress, and fraud.

Network Theory / Structural Embeddedness (Granovetter, 1985; Burt, 1992; Coleman, 1990) argues that economic behaviour is embedded in social relationships. A customer's creditworthiness, fraud risk, and churn propensity are not purely individual attributes but are structurally shaped by their position in the financial network. Social capital embedded in informal savings group (Ajo/Esusu) participation is a creditworthiness signal; structural holes in the transaction graph are a money-mule signal; and churn contagion from network neighbours elevates individual churn risk.

Information Processing Theory (Galbraith, 1974; Daft & Lengel, 1986) predicts that integrating multiple complementary information streams increases an organisation's information processing capacity and reduces uncertainty. For tasks characterised by high equivocality—where identical structural patterns (circular transactions) have multiple interpretations (legitimate ROSCA vs. smurfing)—rich multimodal information is necessary to disambiguate. The JCA fusion mechanism is the architectural realisation of this principle, enabling dynamic, per-instance weighting of semantic and topological information.

These three frameworks generate the Behavioural-Topological Complementarity Principle: a customer's financial behaviour is a joint function of their emotional/cognitive state (NLP) and their structural network position (GBML), and their integration through a principled fusion mechanism produces synergistic improvements exceeding either modality alone.

4. Methodology

4.1. Research Paradigm

The study adopts Design Science Research (DSR; Hevner et al., 2004; Peffers et al., 2007) within a pragmatist philosophy, appropriate because the primary contribution is an artefact—the NIAF-NB framework—evaluated on its utility rather than its truth value. The primary evaluation approach is quantitative experimental analysis, supplemented by qualitative domain expert assessment.

4.2. Dataset

Data were obtained from three tier-1 Nigerian commercial banks under formal data sharing agreements, subject to ethical review board approval and the Nigeria Data Protection Act (NDPA) 2023. All datasets were rigorously anonymised—direct and indirect identifiers replaced by one-way SHA-256 hashes—before transfer to the research team. Table 1 summarises the dataset.

Table 1 Dataset Overview: Three Participating Nigerian Banks (January–December 2022)

Characteristic	Value
Unique customer nodes	~2.1 million
Transaction edges (12 months)	~1.53 billion
Customer service interactions	~102 million
Total transaction value	~₦47.3 trillion
Participating banks	3 (Tier-1 Nigerian commercial banks)
Channels covered	Mobile App, Internet Banking, USSD, POS, ATM
Languages in interactions	English, Nigerian Pidgin English
Thin-file customers (< 6 months history)	~38%

A temporal splitting strategy was employed to prevent data leakage: training (January–June 2022), validation (July–September 2022), and test (October–December 2022). Table 2 reports class distributions for each task.

Table 2 Class Distribution for Each Target Variable

Task	Positive Class	Negative Class	Ratio
Churn Prediction	89,241 (4.3%)	1,987,462 (95.7%)	1:22.3
Credit Risk Assessment	12,847 (7.8%)	152,194 (92.2%)	1:11.8
AML Detection	47,832 (0.2%)	23,724,619 (99.8%)	1:496

4.3. NLP Preprocessing Pipeline

Raw interaction text passed through five stages: (1) PII removal using NER-based detection and regex pattern matching; (2) language detection and Nigerian Pidgin normalisation via a custom dictionary mapping (e.g., "abeg" → "please", "ajo contribution" → "rotating savings club contribution"); (3) text cleaning (HTML removal, abbreviation expansion); (4) tokenisation using the FinBERT WordPiece tokeniser (max 512 tokens); and (5) customer-level aggregation of interaction-level features over a 90-day window preceding the prediction date.

4.4. Graph Construction

The heterogeneous banking graph was constructed with four node types (Customer, Account, Merchant, Device) and five edge types (OWNS, TRANSFERS_TO, PAYS, LOGS_IN_FROM, SHARES_DEVICE), with an additional PARTICIPATES_IN edge type linking customers to identified Ajo/Esusu group nodes. Edge weights encode transaction frequency and volume. Five meta-paths were defined: P1 (Customer–Transfer–Customer), P2 (two-hop transfer chains for layering detection), P3 (Customer–Account–Account–Customer), P4 (Customer–Device–Customer), and P5 (Customer–Merchant–Customer). Graph partitioning used neighbourhood sampling (as in GraphSAGE) for scalability.

4.5. Evaluation Metrics and Statistical Testing

All models were evaluated on Precision, Recall, F1-Score, AUC-ROC, and AUC-PR. DeLong's test was used for AUC-ROC comparisons and McNemar's test for F1 comparisons, both at $\alpha = 0.05$. Given extreme AML class imbalance (1:496), AUC-PR and Precision were the primary AML metrics. Class-weighted Binary Cross-Entropy was used for churn and credit risk; Focal Loss ($\gamma = 2$) for AML.

5. System architecture and model design

5.1. Nigerian FinBERT (NLP Semantic Module)

The NLP module is built on FinBERT (Araci, 2019), a 110M-parameter BERT-Base model pre-trained on financial text. Domain adaptation was performed in two stages. Stage 1 (continued pre-training): FinBERT was further pre-trained on a corpus of 3.8 million Nigerian banking interactions (~1.2 billion tokens) using Masked Language Modelling (MLM, 15% mask rate) for 3 epochs at learning rate $2e-5$. This yielded Nigerian FinBERT. Stage 2 (supervised fine-tuning): Nigerian FinBERT was fine-tuned on 45,000 annotated interactions for sentiment classification (positive/neutral/negative, $\kappa = 0.78$) and 42,000 for intent classification (Complaint/Inquiry/Closure Request/Dispute/Compliment/General, $\kappa = 0.74$). Customer-level features are aggregated over a 90-day window, and a 128-dimensional semantic embedding is produced per customer.

5.2. Heterogeneous Graph Attention Network (GBML Topological Module)

The HAN architecture (Wang et al., 2019) was selected for its strong performance on heterogeneous graphs and its interpretable attention weights. Node-level attention computes the importance of each meta-path neighbour; semantic-level attention learns the relative importance of each meta-path for the specific task. The HAN uses 2 layers, 128-dimensional node embeddings, 8 attention heads, and a dropout rate of 0.3. The full banking graph (~2.1M nodes, ~1.53B edges) is processed via neighbourhood sampling. The Ajo/Esusu graph schema captures informal savings group participation through the PARTICIPATES_IN edge type, providing creditworthiness and fraud disambiguation signals unavailable to generic models. A 128-dimensional topological embedding is produced per customer.

5.3. Joint Cross-Attention (JCA) Fusion Mechanism

The JCA mechanism bridges the Semantic Behavioural Space (NLP embeddings, $s_i \in \mathbb{R}^{128}$) and the Topological Transactional Space (GBML embeddings, $t_i \in \mathbb{R}^{128}$) through bidirectional cross-attention. Both embeddings are linearly projected to a common $d_{\text{model}} = 256$. Cross-attention from semantic to topological computes: $\tilde{t}_i = \text{Attention}(Q = s_i', K = t_i', V = t_i')$, enriching the topological embedding with semantic context. The reverse operation enriches the semantic embedding with topological context. Residual connections and layer normalisation stabilise training. A gated fusion layer then combines the enriched embeddings: $u_i = \text{gate} \cdot \hat{s}_i + (1 - \text{gate}) \cdot \hat{t}_i$, and a feed-forward network projects to $d_{\text{out}} = 128$. This dynamic weighting allows the model to resolve equivocality—for example, recognising that a structurally anomalous circular transaction flow is legitimate Ajo activity when accompanied by "ajo contribution" narration text.

5.4. Downstream Task Heads and Training

The unified representation is passed to a task-specific binary classification head (linear + sigmoid). The entire framework is trained end-to-end using Adam ($lr = 1e-4$, weight decay = $1e-5$, batch size = 512) for 20 epochs with early stopping (patience = 5 on validation AUC-ROC). Training required approximately 48 GPU-hours per task on NVIDIA A100 GPUs.

6. Results

6.1. Customer Churn Prediction

Table 3 reports performance on the churn prediction test set ($n = 2,076,703$; 4.3% churned).

Table 3 Customer Churn Prediction: Performance Comparison (Test Set)

Model	Precision	Recall	F1	AUC-ROC	AUC-PR
Logistic Regression	0.58	0.47	0.52	0.762	0.391
XGBoost	0.67	0.61	0.64	0.831	0.512
NLP-Only (Nigerian FinBERT)	0.69	0.63	0.66	0.857	0.543
GBML-Only (HAN)	0.71	0.65	0.68	0.891	0.601
Early Fusion (Concat)	0.74	0.68	0.71	0.921	0.661

Late Fusion (Ensemble)	0.73	0.67	0.70	0.916	0.643
NIAF-NB (Proposed)	0.76	0.82	0.791	0.967	0.743

NIAF-NB achieves AUC-ROC 0.967, significantly outperforming XGBoost (0.831, DeLong's test $p < 0.001$). SHAP analysis identifies the Temporal Sentiment Trend (TST) as the single most important feature (mean $|SHAP| = 0.23$), followed by Intent: Closure Request Probability (0.18) and Transaction Frequency Ratio (0.15). The dominance of NLP-derived features confirms Behavioural Finance Theory: emotional trajectory precedes and predicts churn behaviour. The Meta-Path P1 Churn Rate (0.12) validates network contagion as a secondary driver.

6.2. Credit Risk Assessment

Table 4 presents credit risk assessment performance for the full dataset and the thin-file subset (customers with fewer than 6 months of transaction history, $n \approx 38\%$ of the evaluated population).

Table 4 Credit Risk Assessment: Full Dataset and Thin-File Subset

Model	F1 (Full)	AUC-ROC (Full)	F1 (Thin-File)	AUC-ROC (Thin-File)
Logistic Regression	0.51	0.713	0.37	0.594
XGBoost	0.62	0.781	0.44	0.631
NLP-Only	0.61	0.769	0.52	0.712
GBML-Only (HAN)	0.64	0.812	0.67	0.841
Early Fusion	0.67	0.857	0.70	0.861
Late Fusion	0.66	0.849	0.69	0.854
NIAF-NB (Proposed)	0.71	0.893	0.73	0.887

For thin-file customers, NIAF-NB achieves AUC-ROC 0.887 compared to 0.631 for XGBoost ($p < 0.001$), a 40% relative improvement. The GBML-Only model (0.841) substantially outperforms XGBoost (0.631) on thin-file customers, confirming Social Capital Theory: network-based features—particularly Ajo/Esusu participation regularity and community membership—are more informative than transactional history for customers with limited formal financial records. The NLP module adds a further 0.046 AUC-ROC points on the thin-file subset by capturing behavioural signals of financial distress (urgency, loan inquiry intent).

6.3. Anti-Money Laundering (AML) Detection

Table 5 presents AML detection performance, including the banks' existing rule-based system as an additional baseline.

Table 5 AML Detection Performance Comparison (Test Set)

Model	Precision	Recall	F1	AUC-ROC	AUC-PR
Rule-Based System	0.08	0.72	0.14	N/A	N/A
Logistic Regression	0.11	0.53	0.18	0.761	0.103
XGBoost	0.27	0.74	0.39	0.891	0.312
NLP-Only	0.19	0.61	0.29	0.831	0.241
GBML-Only (HAN)	0.52	0.81	0.63	0.941	0.581
Early Fusion	0.61	0.84	0.71	0.958	0.672
Late Fusion	0.59	0.83	0.69	0.954	0.641
NIAF-NB (Proposed)	0.71	0.89	0.79	0.967	0.743

The AML results are the most practically significant. The existing rule-based system achieves Precision 0.08: 92% of flagged transactions are false positives, consuming approximately 810 full-time-equivalent compliance officer days per day for a bank processing 3 million daily transactions. NIAF-NB achieves Precision 0.71, reducing the false-positive burden by approximately 90% while simultaneously improving Recall from 0.72 to 0.89. The GBML-Only model (Precision 0.52) confirms Network Theory: structural analysis of transaction graphs is the dominant signal for AML detection. The NLP module adds 0.19 Precision points by disambiguating structurally anomalous transactions—critically, correctly classifying 14,300 Ajo/Esusu transaction clusters as legitimate that the rule-based system had flagged as suspicious.

6.4. Ablation Study

Table 6 reports the contribution of each framework component across all three tasks.

Table 6 Ablation Study: AUC-ROC Across All Three Tasks

Model Configuration	Churn AUC-ROC	Credit AUC-ROC	AML AUC-ROC
Full NIAF-NB (Integrated)	0.967	0.893	0.967
Without NLP Module	0.921 (-0.046)	0.857 (-0.036)	0.941 (-0.026)
Without GBML Module	0.881 (-0.086)	0.769 (-0.124)	0.831 (-0.136)
Generic FinBERT (no domain adapt.)	0.941 (-0.026)	0.864 (-0.029)	0.948 (-0.019)
Without Cross-Attention (Early Fusion)	0.921 (-0.046)	0.857 (-0.036)	0.958 (-0.009)
Without Cross-Attention (Late Fusion)	0.916 (-0.051)	0.849 (-0.044)	0.954 (-0.013)
Without Meta-Path P2 (AML layering)	—	—	0.944 (-0.023)
Without Semantic-Level Attention	0.948 (-0.019)	0.871 (-0.022)	0.951 (-0.016)

Key findings from the ablation study: (1) The NLP module contributes most to churn (-0.046 on removal), confirming that emotional/behavioural signals drive churn prediction. (2) The GBML module contributes most to AML (-0.136) and credit risk (-0.124), confirming that structural network position is paramount for these tasks. (3) Domain adaptation of FinBERT (Nigerian FinBERT vs. generic FinBERT) is significant across all tasks (average -0.025). (4) The JCA mechanism outperforms both Early Fusion and Late Fusion across all tasks, validating its contextual weighting mechanism.

6.5. Fairness Evaluation

Table 7 reports NIAF-NB AUC-ROC disaggregated by demographic subgroup. No demographic group exhibits a performance deficit exceeding 0.031 AUC-ROC points relative to the overall model. The absence of gender disparities (Male: 0.962; Female: 0.964) is particularly noteworthy. The Urban-Rural gap (0.967 vs. 0.936) is attributable to lower digital interaction volumes for rural customers, reducing the NLP module's effective data—a limitation that motivates continued investment in rural digital banking infrastructure.

Table 7 Fairness Evaluation: AUC-ROC by Demographic Subgroup

Demographic Subgroup	Churn AUC	Credit AUC	AML Precision
Male Customers	0.962	0.891	0.703
Female Customers	0.964	0.888	0.714
Urban Customers	0.967	0.894	0.718
Rural Customers	0.936	0.878	0.691
Young Customers (18-30)	0.959	0.882	0.701
Middle-Aged (31-50)	0.961	0.891	0.712
Thin-File Customers	0.944	0.887	0.697

7. Discussion

7.1. Theoretical Implications

The finding that the Temporal Sentiment Trend (TST) is the single most important predictor of customer churn—surpassing all transactional features—is a theoretically significant result. It validates the Emotional Finance framework (Taffler & Tuckett, 2010): the emotional trajectory of the customer relationship is more predictive of churn than the financial trajectory. This challenges the prevailing paradigm of churn prediction, which treats NLP features as supplementary to transactional features. In the Nigerian digital banking context, where customers interact frequently with chatbots and are vocal about their experiences, emotional signals are leading indicators that precede observable transactional changes by weeks or months.

The GBML module's dominance for AML detection validates Network Theory's core claim: money laundering is a network-level phenomenon detectable through structural analysis. The two-hop transfer chain meta-path (P2) is the most important structural signal, directly corresponding to the theoretical concept of "layering"—the deliberate use of transaction chains to obscure illicit origins. Equally important, the model correctly classifies Ajo/Esusu transaction clusters as legitimate, demonstrating that contextualised structural analysis can distinguish between structurally similar patterns with fundamentally different meanings—a disambiguation task impossible for rule-based or node-isolated models.

The consistent superiority of NIAF-NB over both unimodal models across all tasks validates the Information Processing Theory prediction: integrating complementary information streams increases predictive capacity beyond what either stream achieves alone. The JCA mechanism's superiority over Early and Late Fusion confirms that the integration must be dynamic and context-dependent—not a fixed concatenation but a bidirectional enrichment where each modality queries the other.

7.2. Practical Implications

The AML results carry the most immediate practical impact. A bank processing 3 million daily transactions with a 0.05% true suspicious rate, currently operating at 8% AML Precision, flags ~27,000 transactions per day, requiring ~810 full-time compliance officers for review. NIAF-NB's 71% Precision reduces this to ~76 officers—a ~90% reduction in compliance burden—while simultaneously catching more genuine suspicious transactions (Recall: 0.89 vs. 0.72). In the context of Nigeria's FATF grey-listing (2023), this capability improvement is a regulatory imperative.

For customer retention, the framework's ability to identify churn risk weeks in advance—driven by the declining TST feature—enables proactive, personalised interventions. If the average customer lifetime value is ₦850,000, a successful intervention costs ₦12,000, and the model enables retention of 85% of identified at-risk customers, the net present value for a bank with 100,000 at-risk customers is approximately ₦82.9 billion annually.

For financial inclusion, the thin-file credit risk improvement (AUC-ROC from 0.631 to 0.887) offers a pathway to extending formal credit to approximately 53 million financially excluded Nigerian adults. The Ajo/Esusu participation regularity score—a signal of financial discipline entirely invisible to traditional credit models—enables banks to credit-score customers whose financial behaviour has historically been legible only in the informal economy.

7.3. Limitations

Several limitations bound this research. First, the NLP pipeline covers only English and Nigerian Pidgin English; Yoruba, Igbo, Hausa, and other languages used in some interactions are excluded. Second, data was sourced from three tier-1 banks; generalisability to tier-2, tier-3, and microfinance institutions—which may have different transaction patterns and data quality—requires further validation. Third, the study is limited to a 12-month observation window (January–December 2022); temporal drift analysis confirms gradual performance decay, requiring periodic retraining. Fourth, causal claims cannot be made; the relationships identified are correlational. Fifth, the 1.8-second per-1,000-customer inference latency may preclude real-time single-transaction AML decisions, addressed by the recommended two-stage deployment (GBML as real-time pre-filter, NLP applied to flagged transactions).

8. Conclusion

This paper presented NIAF-NB—the first integrated NLP-GBML framework for retail banking analytics in the Nigerian context. The framework combines Nigerian FinBERT, a domain-adapted language model for Nigerian banking text

including Nigerian Pidgin English, with a Heterogeneous Graph Attention Network that models the full banking transaction ecosystem, integrated through a novel Joint Cross-Attention fusion mechanism. Evaluated on a large-scale anonymised dataset from three tier-1 Nigerian banks, NIAF-NB significantly and significantly outperforms all baselines: AUC-ROC 0.967 for churn, 0.893 for credit risk (0.887 for thin-file customers), and AML Precision 0.71 compared to 0.08 for existing rule-based systems.

The five principal contributions of this work are: (1) the NIAF-NB integrated framework, the first of its kind for Nigerian retail banking; (2) Nigerian FinBERT, the first language model adapted for Nigerian banking text; (3) the Joint Cross-Attention fusion mechanism for principled multimodal integration; (4) the Ajo/Esusu graph schema, enabling culturally-aware fraud disambiguation and inclusive credit scoring; and (5) the Behavioural-Topological Complementarity Principle, which provides a theoretical explanation for the synergistic value of multimodal integration in financial analytics.

Future research should extend the NLP pipeline to Yoruba, Igbo, and Hausa; evaluate the framework on tier-2 and tier-3 institutions and mobile money operators; develop real-time streaming deployment for single-transaction AML; incorporate temporal GNN extensions (EvolveGCN, TGAT) to model graph dynamics; and apply federated learning to enable multi-bank training without data centralisation. The framework presents a deployable, ethically-grounded blueprint for the transformation of analytical capabilities in Nigerian banking—and, by extension, across emerging market financial systems facing comparable data richness and institutional complexity.

Recommendations

Based on the empirical findings, the following recommendations are made for Nigerian banks and policymakers.

- For Nigerian commercial banks: (1) Prioritise deployment of the NIAF-NB AML module as the most immediate return on investment, given the severe false-positive burden of existing systems and the FATF regulatory pressure. A phased pilot with one tier-1 bank, validated against existing SAR data, should precede sector-wide deployment. (2) Establish a multimodal data engineering infrastructure—specifically a Data Lakehouse architecture integrating structured transaction data and unstructured customer interaction data—as a prerequisite for NIAF-NB deployment. (3) Invest in Nigerian Pidgin English NLP capability, including annotation of customer interaction data and expansion of the normalisation dictionary to cover regional dialect variations. (4) Implement the thin-file credit scoring capability to serve the Ajo/Esusu-participating demographic, unlocking a large, creditworthy population currently excluded from formal credit markets.
- For regulators (CBN, NDPC): (1) Develop model risk management guidelines specifically addressing heterogeneous GNN and multimodal AI systems, including requirements for fairness auditing, explainability, and periodic retraining. (2) Consider incentivising banks that can demonstrate material AML false-positive rate reductions through AI-driven systems, aligning commercial incentives with the FATF remediation agenda. (3) Facilitate data-sharing sandboxes that enable collaborative model training across institutions without centralising sensitive customer data, leveraging federated learning techniques.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

- [1] Araci, D. (2019). FinBERT: Financial Sentiment Analysis with Pre-trained Language Models. arXiv:1908.10063.
- [2] Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50(2), 179–211.
- [3] Alarab, I., Prakoonwit, S., & Nacer, M. I. (2020). Competence of Graph Convolutional Networks for Anti-Money Laundering in Bitcoin Blockchain. ICMLC 2020.
- [4] Burt, R. S. (1992). *Structural Holes: The Social Structure of Competition*. Harvard University Press.
- [5] Chalkidis, I., Androustopoulos, I., & Aletras, N. (2020). LEGAL-BERT: The Muppets straight out of Law School. EMNLP 2020.

- [6] Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. KDD 2016. <https://doi.org/10.1145/2939672.2939785>
- [7] Cheng, D., Jiang, Y., Luo, Y., Liu, W., & Yu, P. S. (2020). Graph Neural Networks for Financial Fraud Detection: A Survey. arXiv:2008.01514.
- [8] Coleman, J. S. (1988). Social capital in the creation of human capital. *American Journal of Sociology*, 94(Supplement), S95–S120.
- [9] Daft, R. L., & Lengel, R. H. (1986). Organizational information requirements, media richness and structural design. *Management Science*, 32(5), 554–571.
- [10] Dasgupta, K., Singh, R., Viswanathan, B., Chakraborty, D., Mukherjea, S., Nanavati, A. A., & Joshi, A. (2008). Social ties and their relevance to churn in mobile telecom networks. EDBT 2008.
- [11] DeLong, E. R., DeLong, D. M., & Clarke-Pearson, D. L. (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. *Biometrics*, 44(3), 837–845.
- [12] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv:1810.04805.
- [13] FATF. (2023). Jurisdictions under increased monitoring – February 2023. <https://www.fatf-gafi.org>
- [14] Galbraith, J. R. (1974). Organization design: An information processing view. *Interfaces*, 4(3), 28–36.
- [15] Ge, Y., He, S., Xiong, J., Brown, D. E., & Beling, P. A. (2017). Predicting and deterring default with social media information in peer-to-peer lending. *IEEE SSCI 2017*.
- [16] Granovetter, M. (1985). Economic action and social structure: The problem of embeddedness. *American Journal of Sociology*, 91(3), 481–510.
- [17] Hamilton, W. L., Ying, R., & Leskovec, J. (2017). Inductive Representation Learning on Large Graphs. NIPS 2017.
- [18] Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS Quarterly*, 28(1), 75–105.
- [19] Hu, Z., Dong, Y., Wang, K., & Sun, Y. (2020). Heterogeneous Graph Transformer. WWW 2020.
- [20] Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2), 263–291.
- [21] Kipf, T. N., & Welling, M. (2017). Semi-Supervised Classification with Graph Convolutional Networks. ICLR 2017. arXiv:1609.02907.
- [22] Lee, J., Yoon, W., Kim, S., Kim, D., Kim, S., So, C. H., & Kang, J. (2020). BioBERT: A pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics*, 36(4), 1234–1240.
- [23] Li, L. (2019). Understanding complaint behavior in mobile banking. *International Journal of Information Management*. <https://doi.org/10.1016/j.ijinfomgt.2019.05.013>
- [24] Li, Y., Wang, S., Ding, H., & Chen, H. (2023). Large Language Models in Finance: A Survey. arXiv:2311.10723.
- [25] Lin, T. Y., Goyal, P., Girshick, R., He, K., & Dollár, P. (2017). Focal loss for dense object detection. ICCV 2017.
- [26] Liu, Z., Dou, Y., Yu, P. S., Deng, Y., & Peng, H. (2021). Alleviating the inconsistency problem of applying GNN to fraud detection. ACM SIGIR 2021.
- [27] Loughran, T., & McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *The Journal of Finance*, 66(1), 35–65.
- [28] Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. NIPS 2017.
- [29] Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. arXiv:1301.3781.
- [30] Motie, S., & Raahemi, B. (2023). Financial fraud detection using graph neural networks: A systematic review. *Expert Systems with Applications*, 207, 117-967.
- [31] NIBSS. (2023). Annual report on electronic payment statistics in Nigeria. Nigeria Inter-Bank Settlement System.
- [32] Pareja, A., Domeniconi, G., Chen, J., Ma, T., Suzumura, T., Kanezashi, H., Kaler, T., & Leiserson, C. E. (2020). EvolveGCN: Evolving Graph Convolutional Networks for Dynamic Graphs. AAAI 2020.

- [33] Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. *JMIS*, 24(3), 45–77.
- [34] Pennington, J., Socher, R., & Manning, C. D. (2014). GloVe: Global vectors for word representation. *EMNLP 2014*.
- [35] Perozzi, B., Al-Rfou, R., & Skiena, S. (2014). DeepWalk: Online learning of social representations. *KDD 2014*.
- [36] PwC. (2023). Graph LLMs: AI's next frontier in banking and insurance. PricewaterhouseCoopers.
- [37] Salloum, S., Dautov, R., Chen, X., Peng, P. X., & Huang, J. Z. (2016). Big data analytics on Apache Spark. *International Journal of Data Science and Analytics*, 1(3–4), 145–164.
- [38] Sawhney, R., Agarwal, S., Wadhwa, A., & Shah, R. R. (2020). Deep attentive learning for stock movement prediction from social media text and company correlations. *EMNLP 2020*.
- [39] Scarselli, F., Gori, M., Tsoi, A. C., Hagenbuchner, M., & Monfardini, G. (2009). The graph neural network model. *IEEE Transactions on Neural Networks*, 20(1), 61–80.
- [40] Taffler, R., & Tuckett, D. (2010). Emotional Finance: The role of the unconscious in financial decisions. In Baker & Nofsinger (Eds.), *Behavioural Finance*. Wiley.
- [41] Thaler, R. H. (1985). Mental accounting and consumer choice. *Marketing Science*, 4(3), 199–214.
- [42] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is All You Need. *NIPS 2017*.
- [43] Veličković, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., & Bengio, Y. (2018). Graph Attention Networks. *ICLR 2018*. arXiv:1710.10903.
- [44] Waehner, K. (2021). How data streaming with Apache Kafka and Flink drives top innovations in financial services. *Confluent*.
- [45] Wang, X., Ji, H., Shi, C., Wang, B., Ye, Y., Cui, P., & Yu, P. S. (2019). Heterogeneous Graph Attention Network. *WWW 2019*. <https://doi.org/10.1145/3308558.3313562>
- [46] Wang, X., Liu, N., Han, H., & Shi, C. (2021). Self-supervised heterogeneous graph pre-training based on structural clustering. *NeurIPS 2021*.
- [47] Weber, M., Domeniconi, G., Chen, J., Weidele, D. K. I., Bellei, C., Robinson, T., & Leiserson, C. E. (2019). Anti-money laundering in bitcoin: Experimenting with graph convolutional networks for financial forensics. arXiv:1908.02591.
- [48] World Bank. (2022). *The Global Findex Database 2021: Financial inclusion, digital payments, and resilience in the age of COVID-19*. World Bank Publications.
- [49] Wu, S., Irsoy, O., Lu, S., Dabrovolski, V., Dredze, M., Gehrmann, S., Kambadur, P., Rosenberg, D., & Mann, G. (2023). BloombergGPT: A large language model for finance. arXiv:2303.17564.
- [50] Xu, Y., & Cohen, S. B. (2018). Stock movement prediction from tweets and historical prices. *ACL 2018*.
- [51] Yang, Y., Uy, M. C. S., & Huang, A. (2020). FinBERT: A pretrained language model for financial communications. arXiv:2006.08097.
- [52] Yao, Z., Huang, T., Sun, J., Shan, L., & Deng, H. (2022). A heterogeneous graph neural network for credit risk assessment. arXiv:2204.06848.