Advanced Machine Learning for Nigerian Banking Fraud Detection using NIBSS Dataset

Team 17

Data Conduits

Swarup E (swarupe@iisc.ac.in)

Novoneel C (cnovoneel@iisc.ac.in)

Sarthak S (sarthak1@iisc.ac.in)

Rakshit R (rrakshit@iisc.ac.in)

Problem Definition

Background

The Nigerian banking industry reported

- ₦17.67 billion in fraud losses in 2023,
- 23% increase from 2022, across over 95,000 cases.
- Mobile, social-engineering-based and pin-swap frauds dominate, highlighting the growing sophistication of cyber-criminals.
- Traditional rule-based systems struggle with evolving fraud patterns and high false positive rates.

Importance

- Fraud detection accuracy directly impacts customer trust, regulatory compliance, and financial stability.
- With Lagos accounting for 48% of fraud cases, there's urgent need for detection systems that can adapt to Nigerian banking patterns, reduce manual intervention, and provide real-time protection

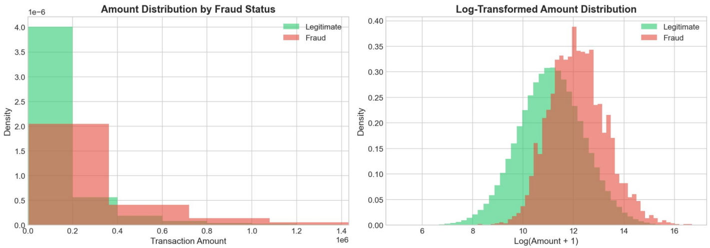
Data and EDA

Data Characteristics

- NIBSS transaction dataset, which presented a classic, albeit extreme, case of class imbalance
- Out of 1M transactions from 10k customers, only about 3000 were confirmed as fraudulent—a prevalence rate of just **0.3%**.

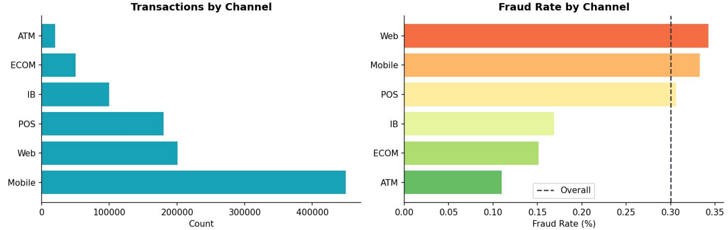
EDA

- Legitimate transactions peaked during standard business hours (8 AM to 5 PM), fraud attempts high during the early morning hours (1 AM to 4 AM)
- Digital vulnerabilities: Mobile and Web channels showed a disproportionately higher rate of fraud compared to ATM or POS transactions
- Fraudulent transactions tended to involve higher amounts
- Fraud tended to happen more frequently in the larger cities of Nigeria like Lagos and Abuja



Fraud distribution by amount

Fraud distribution by time



Fraud distribution by channel

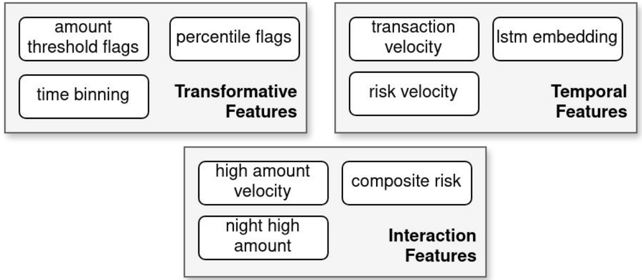
Feature Engineering

- Broadly performed feature engineering of three types – Transformative, Temporal and Interaction

- Transformative feature engineering –
  - Categorize existing columns such as amount and time
  - Percentiles and thresholds based labels

- Temporal feature engineering –
  - Aggregate transaction timestamps for users to come up with temporal features
  - Use LSTM to come up with vector embedding features to form a stacked model later

- Interaction feature engineering –
  - Combine multiple features to form new features
  - Composite risk score base aggregating risk weights from the transaction channel, time of day, and location

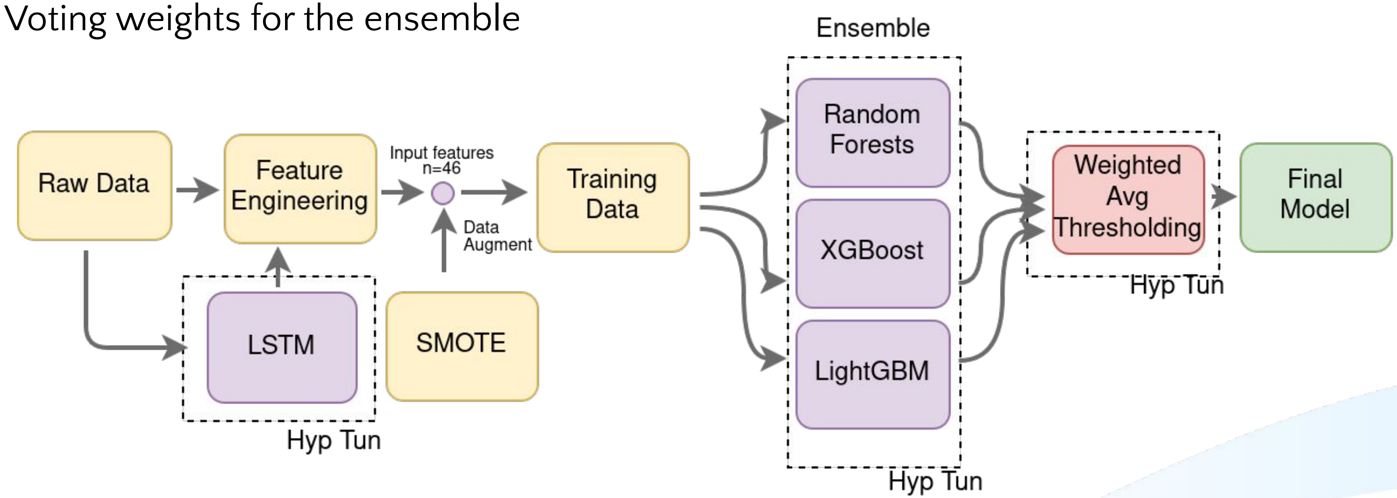


Feature Engineering Components

Feature Engineering

Methodology

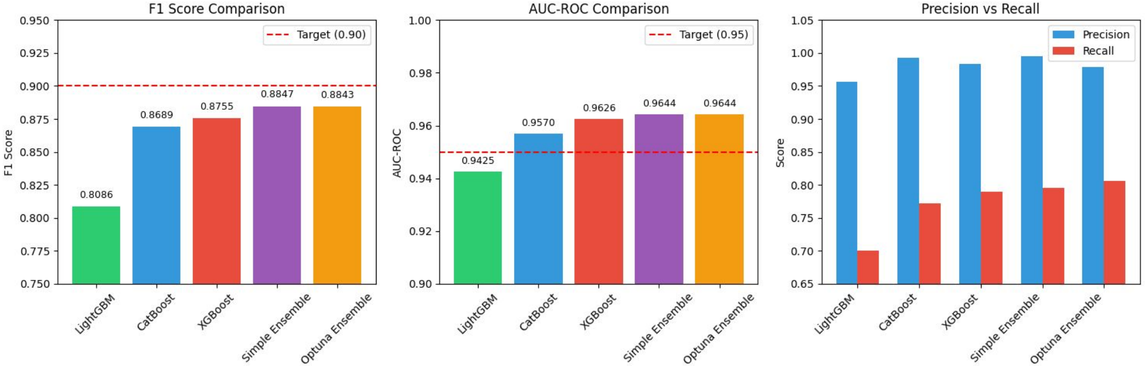
- Extensive feature engineering with both Composite and temporal features using LSTMs
- Oversampling of minority fraud rate of 0.3% to improve model attention
- Ensemble of 3 models to reduce overfitting on the large amount of non fraudulent transactions and to capture complex non-linear patterns
- Extensive Hyperparameter Tuning with Optuna to find -
  - Optimum LR, Thresholds, Loss weights (Focal Loss)
  - Voting weights for the ensemble



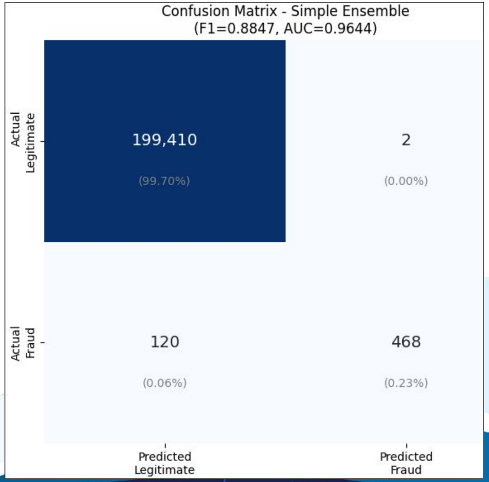
End to End model architecture

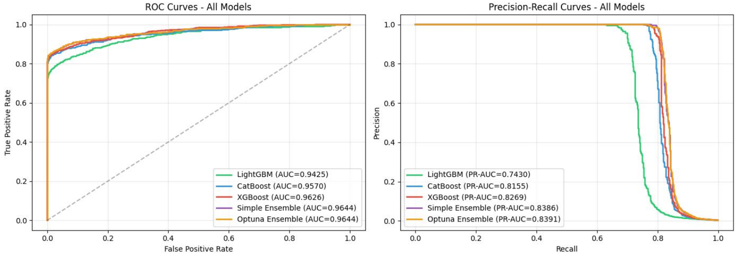
Preliminary Results

- Model score comparison plot shows how the ensemble of models consistently scores better than the individual models on F1, AUC-ROC and Precision vs Recall
- ROC and Precision vs Recall curves show that the model is performing well
- ROC Curve is closer to top left corne high true positive gain
- Precision vs Recall is close to top ri corner showing that most of the positives that are picked are also precise
- Sharp dropoff of precision at highe recall values indicate imbalance distribution
- Ensemble precision at 0.96



Model score comparisons



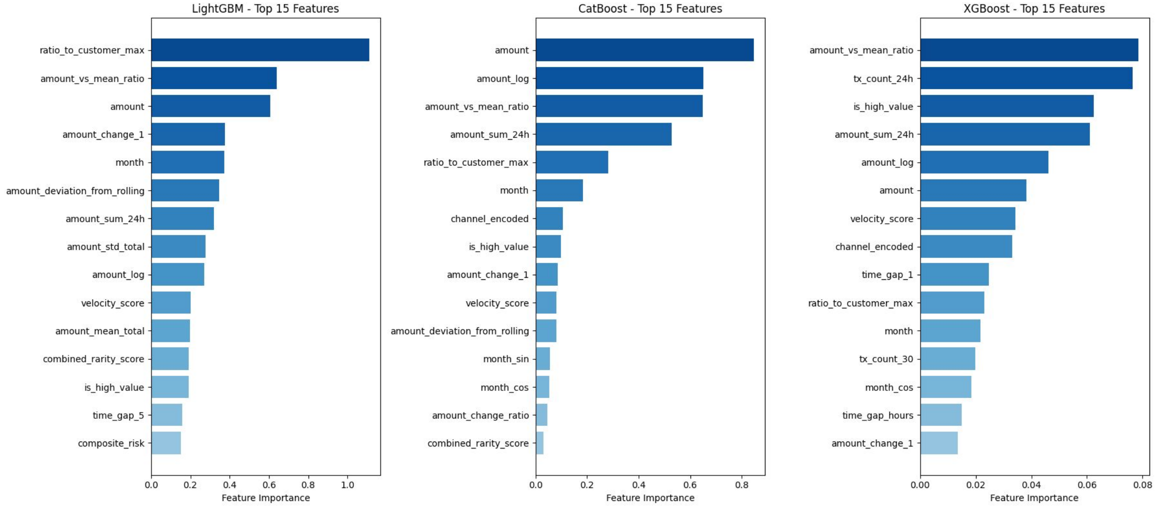


ROC and P-R Curve

Feature Importance

- Our choice for an ensemble of models is justified in the SHAP plot shown below

- Different models result in more importance for different features –
  - LightGBM has ratio_to_customer max as most important feature
  - LightGBM also looks at deviation features such as amount_deviation_from_rolling
  - CatBoost and XGBoost gives more importance to temporal features
  - XGBoost considers lag indicators when compared with the others



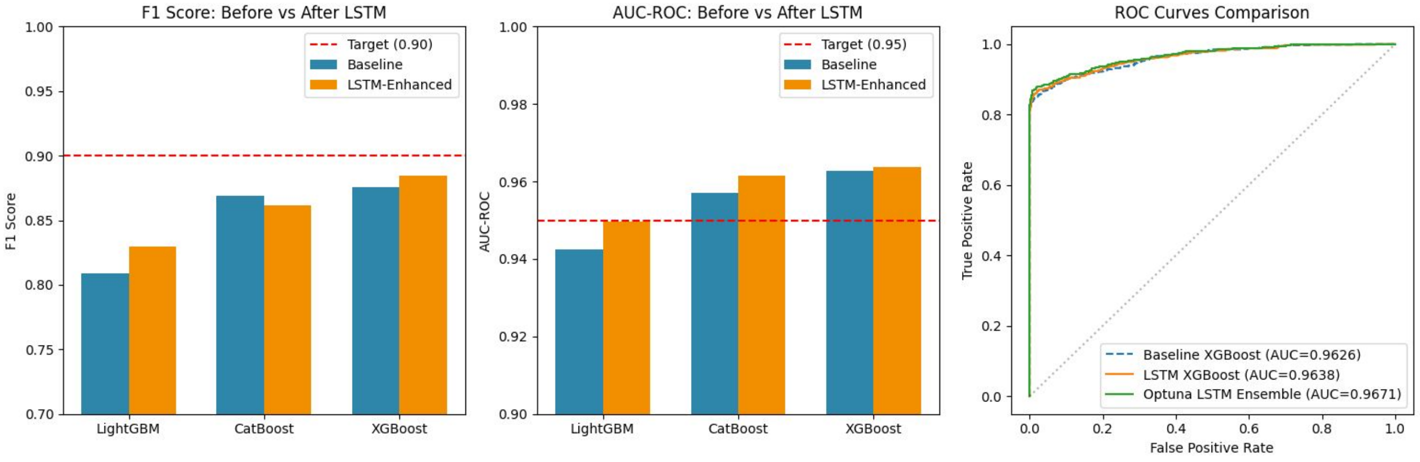
SHAP plot showing importance of features per model

Preliminary Results (without LSTM) cont…

Model	AUC	PR AUC	F1	Precision	Recall	FPR	Threshold
LightGBM	0.9425	0.743	0.8086	0.9559	0.7007	0.0001	0.3463
CatBoost	0.957	0.8155	0.8689	0.9934	0.7721	0	0.534
XGBoost	0.9626	0.8269	0.8755	0.9831	0.7891	0	0.4703
Simple Ensemble	0.9644	0.8386	0.8847	0.9957	0.7959	0	0.3943
Optuna Ensemble	0.9644	0.8391	0.8843	0.9793	0.8061	0.0001	0.3422

Results – Incremental improvements (with LSTM)

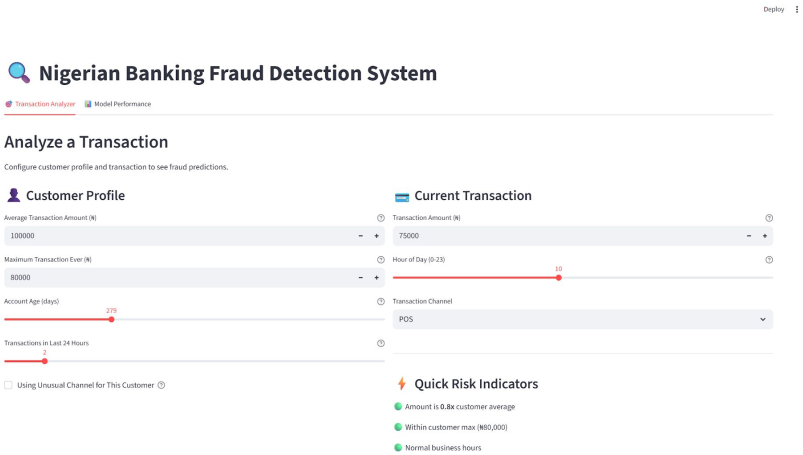
- Use of LSTM to create additional temporal features
- Increases the number of features from 61 to 125 with 64 additional temporal embeddings
- Idea is to capture hidden temporal transaction trends
- This results in a hybrid architecture
- We observe only marginal increase in overall performance of about 1% in F1 and AUC-ROC metrics
- We attribute this to the lack of temporal trends seen in the limited data

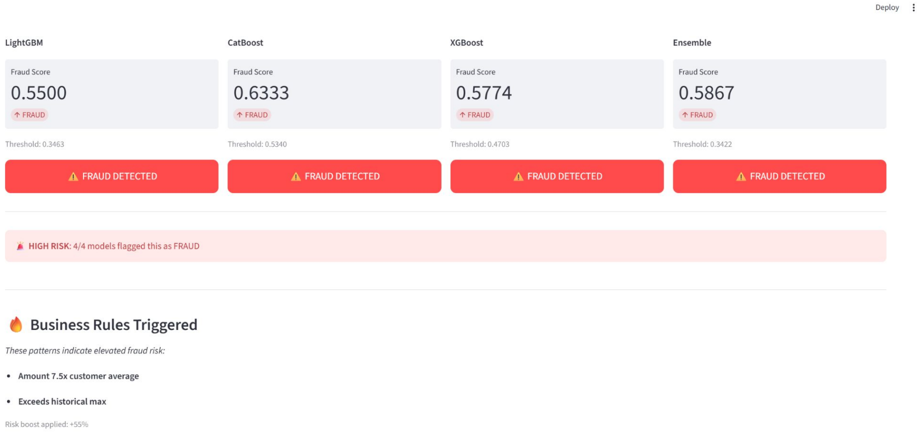


Results – Incremental improvements (with LSTM) cont…

Model	AUC	PR AUC	F1	Precision	Recall	FPR
Baseline LightGBM	0.9425	0.743	0.8086	0.9559	0.7007	0.0001
Baseline CatBoost	0.957	0.8155	0.8689	0.9934	0.7721	0
Baseline XGBoost	0.9626	0.8269	0.8755	0.9831	0.7891	0
LSTM LightGBM	0.9496	0.7786	0.8293	0.9725	0.7228	0.0001
LSTM CatBoost	0.9616	0.8164	0.8615	0.9912	0.7619	0
LSTM XGBoost	0.9638	0.8312	0.8847	0.9957	0.7959	0
Optuna LSTM Ensemble	0.9671	0.844	0.8883	0.9815	0.8112	0

Fraud Detection utility





Results and conclusions

- Ensemble methods outperform individual models by 1-2% in F1
- LSTM embeddings provide modest improvement (~1% F1 gain)
- Original hand-crafted features remain most important
- FPR constraint is easily achievable (actual 0.001% vs target 0.1%)
- This project successfully developed a fraud detection system for Nigerian banking transactions using ensemble machine learning with LSTM enhancement.
- The final model achieves:
  - F1 Score: 0.8847 (98.3% of target)
  - AUC-ROC: 0.9638 (exceeds 0.95 target)
  - FPR: 0.001% (100x better than 0.1% target)

Future work

- Transformer Models: Attention-based sequence modeling
- Graph Networks: Model customer-merchant relationships
- Online Learning: Adapt to emerging fraud patterns
- Cost-Sensitive Learning: Incorporate asymmetric business costs

Data Science Canvas

Project: Advanced Machine Learning for Nigerian Banking Fraud Detection using NIBSS Dataset

Team: Data Conduits

Problem Statement

Business Case & Value Added
Nigerian banks lost ₦17.67B (~$216M) in 2023, with fraud cases rising 23% YoY.
Traditional rule-based systems struggle with evolving fraud patterns and generate high false positives.
This project delivers a machine-learning–driven fraud detection system that:
i. Reduces fraud losses through early detection
ii. Minimizes false positives
iii. Automates low-risk decisions to reduce manual review workload
iv. Enables fraud teams to focus on high-risk, high-value cases
v. Improves compliance and customer trust with explainable predictions

Data Landscape
NIBSS Fraud Dataset (Kaggle) — synthetic data that is representative of transaction data reflecting Nigerian banking characteristics.
Synthetic data is chosen as banking data is sensitive, and personal.

Model Selection
Given the extreme class imbalance (0.3% fraud cases), the solution uses:

Gradient boosting models (LightGBM, XGBoost) for non-linear patterns

Ensemble stacking to leverage complementary feature importance

LSTM embeddings to capture short-term temporal transaction patterns

Optuna for hyperparameter and ensemble weight optimization

Precision–Recall threshold tuning to maintain extremely low FPR

Model Requirements
Performance: F1 ≥ 0.90 (achieved 0.8847), AUC-ROC ≥ 0.95, FPR < 0.1% (achieved 0.001%)

Explainability: SHAP-based interpretability for audit and compliance

Software & Libraries
Python
pandas, NumPy
scikit-learn
LightGBM, XGBoost, CatBoost
Optuna
PyTorch/Keras for LSTM
SHAP
Matplotlib/Seaborn

Skills
Data Engineering: Pipeline design for high-volume transaction data

Data Science: Feature engineering (temporal, velocity, interaction), ensemble modeling

ML Engineering: Optuna tuning, model serving, real-time scoring pipelines

Risk/Fraud Expertise: Interpret patterns, label verification, regulatory compliance

Execution & Evaluation

Model Evaluation
Performance is assessed using F1, AUC-ROC, PR AUC, and precision–recall balance due to extreme imbalance. The ensemble achieved:

AUC: 0.9638
F1: 0.8847

FPR: 0.001% (100× better than requirement)

Confusion matrix analysis guides threshold adjustment.
Real-time monitoring includes drift detection, data quality checks, and alerting when fraud distribution shifts.

Data Storytelling
Target Group Reqs
Clear, business-focused insights on fraud patterns affecting Nigerian payment channels.
Operationally relevant metrics (fraud hotspots, transaction patterns, risk levels by customer/merchant segment).
Actionable recommendations that regulators, banks, and fintech operators can implement immediately.
Simple, interpretable visuals that work for mixed stakeholders (analysts, compliance teams, executives).
Traceability and transparency in how fraud was detected, especially for regulatory review.

How?
Lead with key fraud insights (where, how, and why fraud occurs in the Nigerian context).
Use clean visuals (heatmaps, trend lines, risk scores) instead of technical plots.
Translate findings into operational actions: improved KYC checks, velocity limits, merchant risk tiers, model cutoffs.
Highlight local relevance: mobile money patterns, high-risk time windows, account takeover behaviors typical in Nigeria.

Data Collection & Preparation

Data Selection & Cleansing
Relevant features like:
Temporal: hour, day, rolling time gaps
Velocity: tx_count_24h, amount_sum_24h
Behavioral: ratio_to_customer_max, deviation features
Interaction: composite risk scores

Data cleansing included:
Handling missing timestamps
Outlier detection for extreme transaction amounts
Encoding categorical fields (channel, location, merchant)

Data Integration
Different data sources (transaction logs, customer profiles, merchant metadata, fraud investigation labels) are unified into a centralized feature store.

Data Collection
Real-time data from mobile banking, transfers, POS, ATM, web

Fraud labels from investigation teams

Need for sequence data per customer (for LSTM/transformer models)

Explorative Data Analysis
Fraud prevalence: 0.3%
Fraud concentrated 1 AM – 4 AM
High-risk channels: mobile, web
Fraud values skew higher than legitimate ones
Lagos & Abuja show highest fraud density
Temporal and velocity patterns drive strong predictive power