FRAUD DETECTION : SECURING BANK TRANSACTIONS

Current Challenge: Bank A struggles with detecting evolving, sophisticated fraud patterns in banking. This results in financial losses and reduced customer trust due to missed fraud and inconvenience caused by flagging legitimate transactions.

Need: This calls for a robust, data-driven approach to enhance fraud detection while minimising disruptions to legitimate transactions.

What are the key patterns that indicate fraudulent activity?

BUSINFO 704

Predictive Business Analytics
Quarter 3, 2024

GROUP 22

Jingjing (jin233)
Rajitha (rren378)
)Henry (jche955)
Stephanie (hcha948)

REFERENCES

01. DATA OVERVIEW

500K
Transaction Records (synthetic data)

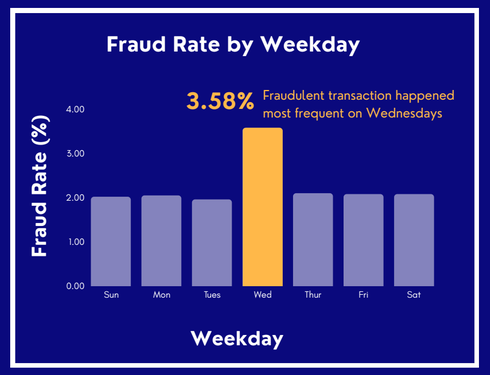
13
Variables - Transaction Amount, Type, Date, Location, Joint Flag, Balance, Agent and related IDs.

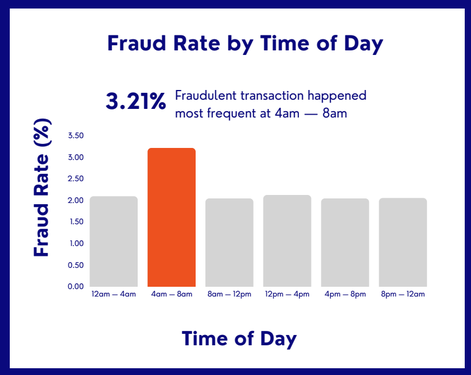
72 years old on average (for fraud transactions)

2% of 500k transactions are fraudulent

33% fraudulent transactions occur during withdrawals.

100 the median fraudulent transaction amount





FRAUDULENT TRANSACTIONS ARE...

- Typically smaller amounts (average $422, median $100), blending in with daily transactions
- Occur in early morning, between 4-5am, when customer activity is low
- Occur on Wednesdays

02. METHODOLOGY

Variable Selection

- Age
- Transaction Type
- Transaction Amount
- Transaction Hour
- Transaction Weekday

Feature Engineering

- Mutate Fraud to Level 1, 0
- Transaction Date to Weekday
- Transaction Date to Hour & time periods

Modeling of Classifiers

- Log Regression, LightGBM, XGBoost
- 80% train - 20% test data

Missing Data: Mutate all NAs to Unknown,
Mutate MerchantID Unknowns to Level 1, else 0, as 73% of the MerchantID is NA

03. RESULTS & FINDINGS

XGBOOST MODEL

XGBoost trained on a SMOTE oversampled dataset

Sensitivity 70%

Area Under The Curve 86%

Key Predictor Ranking

1. Age
2. Transaction Amount
3. Withdrawals

XGBOOST MATRIX

Predicted \ Actual	Fraud	Not Fraud
Fraud	1.6%	**17.6%**
Not Fraud	**0.7%**	80.1%

Target Class: Fraud

Strong Discriminatory Power
Highest True Positives
Lowest False Negatives

70% of the fraudulent transaction is correctly classified

04. RECOMMENDATIONS

ADAPTIVE ALERT SYSTEM

3 months

Adaptive risk thresholds: Adjust time and age specific thresholds, especially during high-risk hours (4 - 8 am).

- flag and delay processing transaction till later
- Withdrawals - push notifications for additional confirmation

TRANSACTION PROFILING

3-6 months

Profile Customer Transactions: Compare current transactions against confirmed legitimate patterns.

Whitelist: Develop a trusted merchant/transaction type list to reduce unnecessary alerts.

ADVANCE MODEL TRAINING

6-12 months

Segmented Models: Train models separately based on customers (age).

Re-calibrate Model: Adjust the fraud detection model to reflect shifts in customer behaviour over time.

Enhances fraud risk management

Reduces inconveniences for false positives

Improves detection accuracy