Fraud Detection Model

Executive Summary

Poundbank, a major London-based financial institution, has observed a significant decline in the accuracy of its machine learning–based fraud detection system. Such degradation directly increases the risk of **financial loss, regulatory scrutiny, reputational damage,** and **customer attrition**.

Our analysis applies advanced **model monitoring and drift detection**—specifically using the `nannyml` library—to compare **historical (reference)** and **recent (production)** transaction datasets.
The investigation revealed **statistically significant data drift** across all monitored input features, with **time_since_login_min** and **transaction_amount** showing the largest shifts. These changes substantially reduce model reliability, making retraining and enhanced monitoring critical.

This report is designed for **both technical stakeholders** (data scientists, ML engineers) and **business decision-makers** (risk/compliance managers, executives).

1. Business Context

- **Challenge:** Fraudulent behavior evolves rapidly—attackers adapt once they understand detection models. Older models become misaligned with real transaction patterns.
- **Impact:** Declining model performance means more fraudulent transactions may bypass detection systems, causing direct monetary loss and damaging customer confidence.
- **Goal:** Identify **how and why** the model's predictive power is declining, and provide a path to restore and maintain accuracy.

2. Data Summary

Two labeled datasets were analyzed:

Dataset	Description
reference.csv	Historical transactions used for testing ("gold standard").
analysis.csv	Recent production transactions observed by the deployed model.

Key Features:

Feature	Description
`timestamp`	Date-time of transaction.
`time_since_login_min`	Minutes since the user last logged in.
`transaction_amount`	Amount in GBP transferred.
`transaction_type`	PAYMENT, CASH-IN, CASH-OUT, etc.
`is_first_transaction`	Boolean: whether this is the customer's 1st transaction.
`user_tenure_months`	Age of customer's account in months.
`is_fraud`	1 if fraud detected, otherwise 0.
`predicted_fraud_proba`	Model-predicted probability of fraud.
`predicted_fraud`	Model's binary classification output.

3. Methodology

Step 1: Data Validation

- Verified column integrity, types, and completeness between reference and production datasets.

Step 2: Drift Detection

- Utilized `nannyml` to calculate:
  - **Pearson correlation coefficients** between historical and production feature distributions.
  - **P-values** to determine statistical significance.
  - **Drift rank** to prioritize features according to their impact.

Step 3: Interpretation

- Mapped detected drift to business impact on fraud detection reliability.

4. Findings

Drift Summary Table:

Feature	Correlation	P-Value	Drift Detected	Rank
`time_since_login_min`	0.953	1.05e-09	☑ True	1
`transaction_amount`	0.626	5.43e-03	☑ True	2
`is_first_transaction`	0.054	0.83	☑ True	3
`user_tenure_months`	-0.101	0.69	☑ True	4
`transaction_type`	-0.187	0.46	☑ True	5

Key Observations:

- **Top Risk Factors:**
  - Large shifts in `time_since_login_min` likely signify changes in customer login and payment behavior—could indicate different fraud origination patterns.
  - Significant variation in `transaction_amount` may reflect evolving fraud schemes aimed at different payment tiers.
- **All features show drift**, meaning the overall input profile seen by the model has materially changed.
- These shifts erode the validity of the model's learned decision boundaries, increasing **false negatives** (missed fraud) and **false positives** (blocking legitimate users).

5. Business Impact

- **Financial:** Higher undetected fraud attempts may result in **substantial monetary losses**.
- **Operational:** Rising false positives could waste fraud analyst capacity and irritate customers.
- **Compliance & Risk:** Regulatory scrutiny may increase if fraud incidents spike.
- **Customer Trust:** Perceived insecurity in transactions can prompt customers to switch to competitors.

6. Technical Impact

- **Model Obsolescence:** Old decision rules no longer match reality.
- **Data Pipeline Vulnerability:** Potential delays in detecting crucial feature drifts.
- **Monitoring Gaps:** Lack of real-time alerts delays corrective retraining.

7. Recommendations

Immediate Actions:

1. **Retrain the model** with latest production data, focusing on drifted features.
2. Set up **automated drift monitoring** using `nannyml` with monthly/weekly reports.
3. **Adjust feature engineering**:
    - Normalize for changing login patterns.
    - Introduce composite fraud indicators.
4. **Human-in-the-loop verification** for ambiguous transactions flagged with medium risk scores.

Long-Term Actions:

- Incorporate **adaptive learning** so the model incrementally updates with labeled fraud data.
- Collaborate with **fraud analysts, risk teams, and engineering** to align monitoring thresholds with business risk tolerance.
- Expand feature set with **external behavioral or geolocation signals**.

8. Conclusion

The detected feature drifts clearly explain why Poundbank's fraud model is losing accuracy. As fraud tactics continue to evolve, Poundbank's strategy must move from static ML models to a **monitor-adapt-retrain cycle**.

By implementing automated monitoring and scheduled retraining, Poundbank can **react faster, detect more fraud, and maintain customer trust**—turning this challenge into a competitive advantage.