Fraud Risk Scoring System

Project: End-to-End Fraud Risk Scoring System

Prepared by:

Helena W.

Founder & Principal Consultant

Medical AI & Healthcare Data Science

Physician, MBBS | Data Science & Explainable AI

GlobalAdSnap

Prepared Date: January 2026

1. Purpose & Context

This report presents a **regulator-safe, end-to-end fraud risk scoring framework** designed for use in **financial services, healthcare claims, insurance payments, and fintech environments**. The objective is to demonstrate how modern machine learning can be deployed responsibly under **high regulatory scrutiny**, while delivering measurable business value.

This system is intentionally designed as **decision support**, not automated enforcement, aligning with regulatory expectations for material financial decisions.

2. Executive Summary

Organizations processing high-volume transactions face a persistent trade-off between:

- Preventing fraud losses
- Controlling false positives and operational cost
- Maintaining regulatory compliance and auditability

This project demonstrates a solution that:

- Scores transactions using a **class-weighted LightGBM model**
- Optimizes thresholds based on **business cost**, not accuracy alone
- Enforces **human-in-the-loop decision control**
- Implements **formal Model Risk Management (SR 11-7–style)** governance

Key Outcomes

- **81.3% reduction in expected fraud loss** compared to baseline
- $330K+ **simulated** annual savings under stated assumptions
- Stable model performance under temporal validation
- No evidence of population drift (PSI = 0.014)

Deployment Decision: Approved for **Phase 0 (Shadow Mode)** only.

3. Business Problem Definition

Challenges

- Extreme class imbalance (~3–4% fraud)
- High cost of false negatives (missed fraud)
- Limited manual review capacity
- Regulatory exposure from automated decisions

Business Objective

1. Maximize fraud dollars prevented
2. Control analyst workload
3. Prevent automated adverse actions
4. Maintain explainability and auditability

4. Solution Overview

Component	Design Choice
Model	LightGBM (class-weighted)
Output	Continuous fraud risk probability
Decision Role	Decision support only
Automation	Semi-automated (human approval required for all adverse actions)
Optimization	Cost-based thresholding
Governance	SR 11-7–aligned framework

The model produces **risk scores**, not decisions. All enforcement actions require human approval.

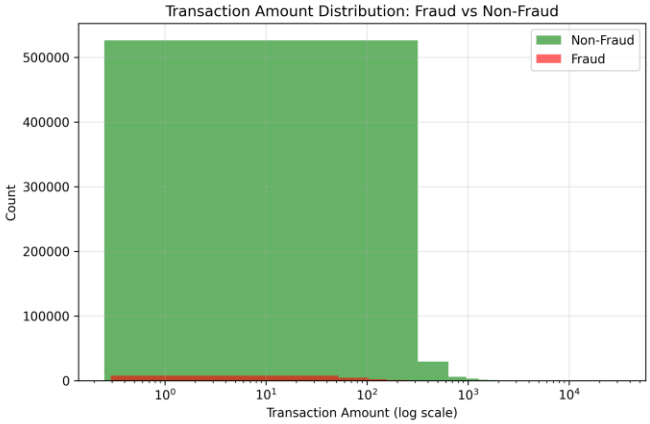
5. Data & Risk Insights

Exploratory analysis identified meaningful, business-relevant fraud patterns:

- Fraud risk increases non-linearly with transaction amount
- Temporal risk patterns by hour of day
- Elevated risk for missing identity attributes
- Product- and card-type heterogeneity

All sensitive features were used **only within the model**, never as hard-coded rules.

Figure 1: Fraud Risk by Transaction Amount

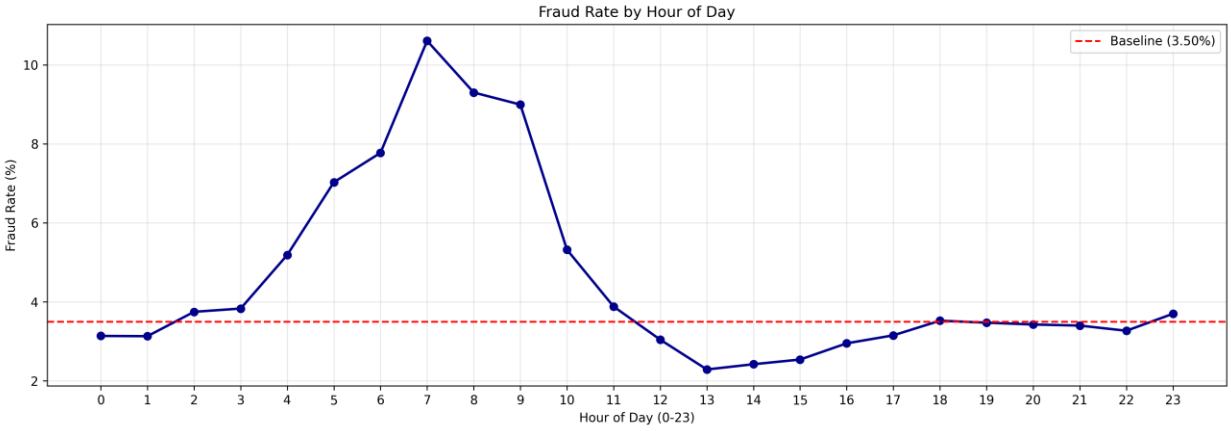


Transaction Amount Statistics

Type	Mean	Median	Std
Non-Fraud	134.51	68.5	239.4
Fraud	149.24	75.0	232.21

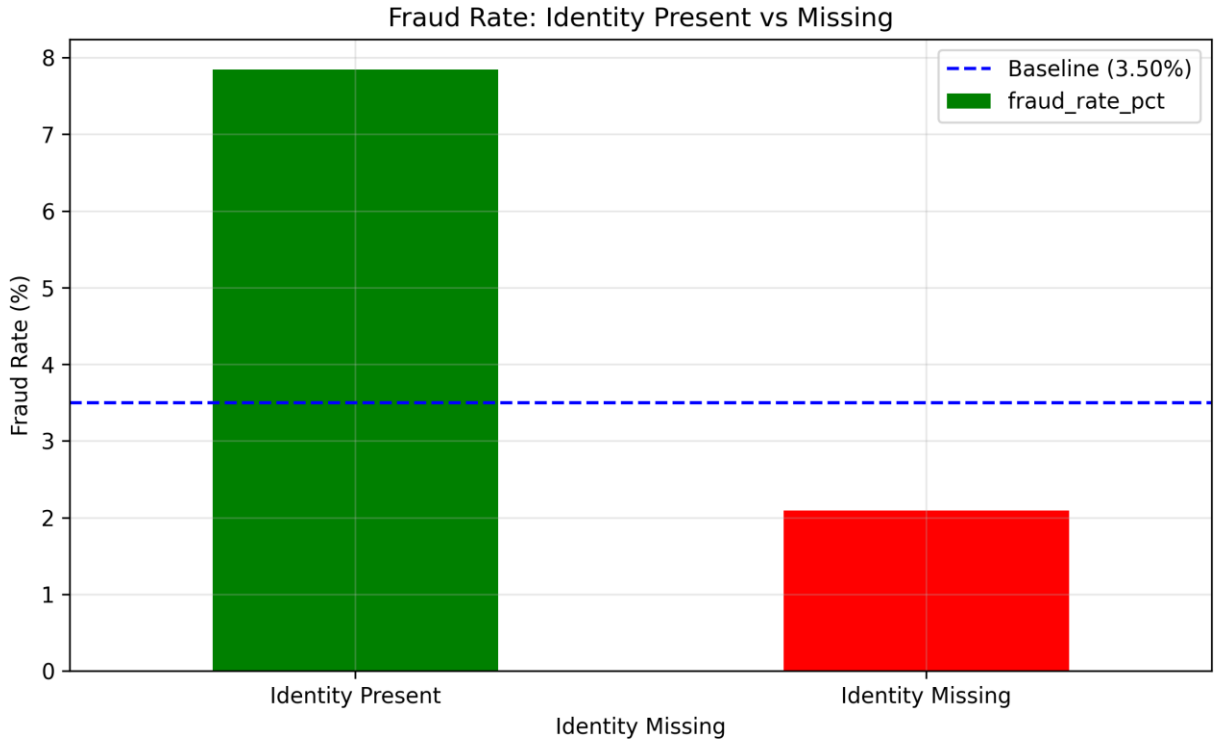
Interpretation:
Fraud risk increases non-linearly with transaction amount, supporting higher risk sensitivity for large-value transactions without imposing hard rules.

Figure 2: Temporal Fraud Risk by Hour of Day



Interpretation:
Distinct time-of-day fraud patterns justify temporal feature inclusion and inform operational staffing for manual review teams.

Figure 3: Fraud Risk by Identity Data Availability



Interpretation:
Transactions with missing identity attributes exhibit materially higher fraud risk, highlighting data completeness as an operational dependency.

6. Model Performance & Validation

Primary Metric

- **Precision–Recall AUC (PR-AUC): 0.548** (cross-validated, class-weighted LightGBM)

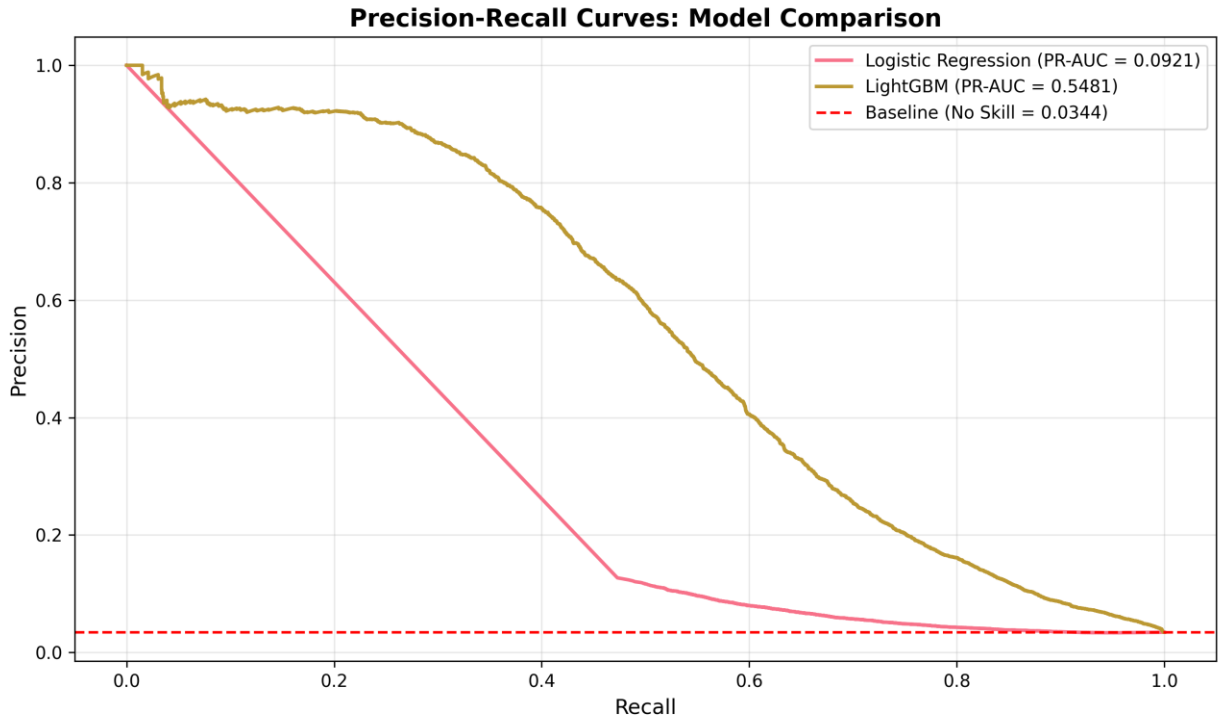
PR-AUC was selected due to extreme class imbalance; ROC-AUC was deemed insufficient for business decisioning.

Temporal Validation

Metric	Value
Train PR-AUC	0.502
Future PR-AUC	0.427
Overfitting Gap	15.0%
PSI	0.014 (Stable)

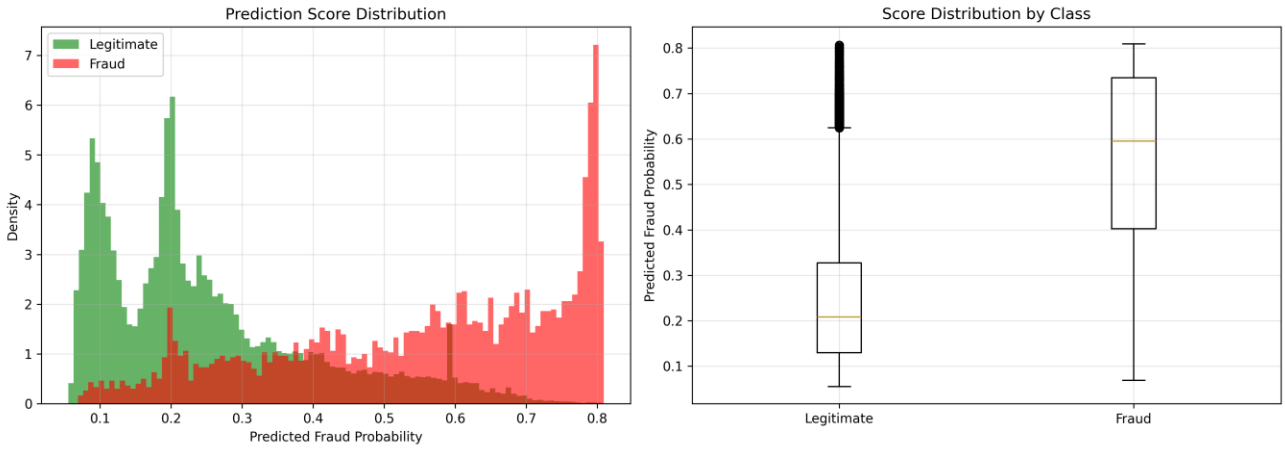
Performance degradation was within acceptable bounds for fraud use cases.

Figure 4: Precision–Recall Performance Under Class Imbalance



Interpretation:
The class-weighted LightGBM model outperforms baseline models across recall levels, validating PR-AUC as the appropriate performance metric.

Figure 5: Risk Score Separation: Fraud vs Legitimate Transactions



Interpretation:
Clear score separation confirms the model's ability to rank fraud risk effectively, enabling threshold-based governance rather than binary automation.

7. Business Threshold Optimization

Thresholds were optimized using a **cost-based decision framework**, incorporating:

- Fraud loss per event
- Manual review cost
- Analyst capacity constraints

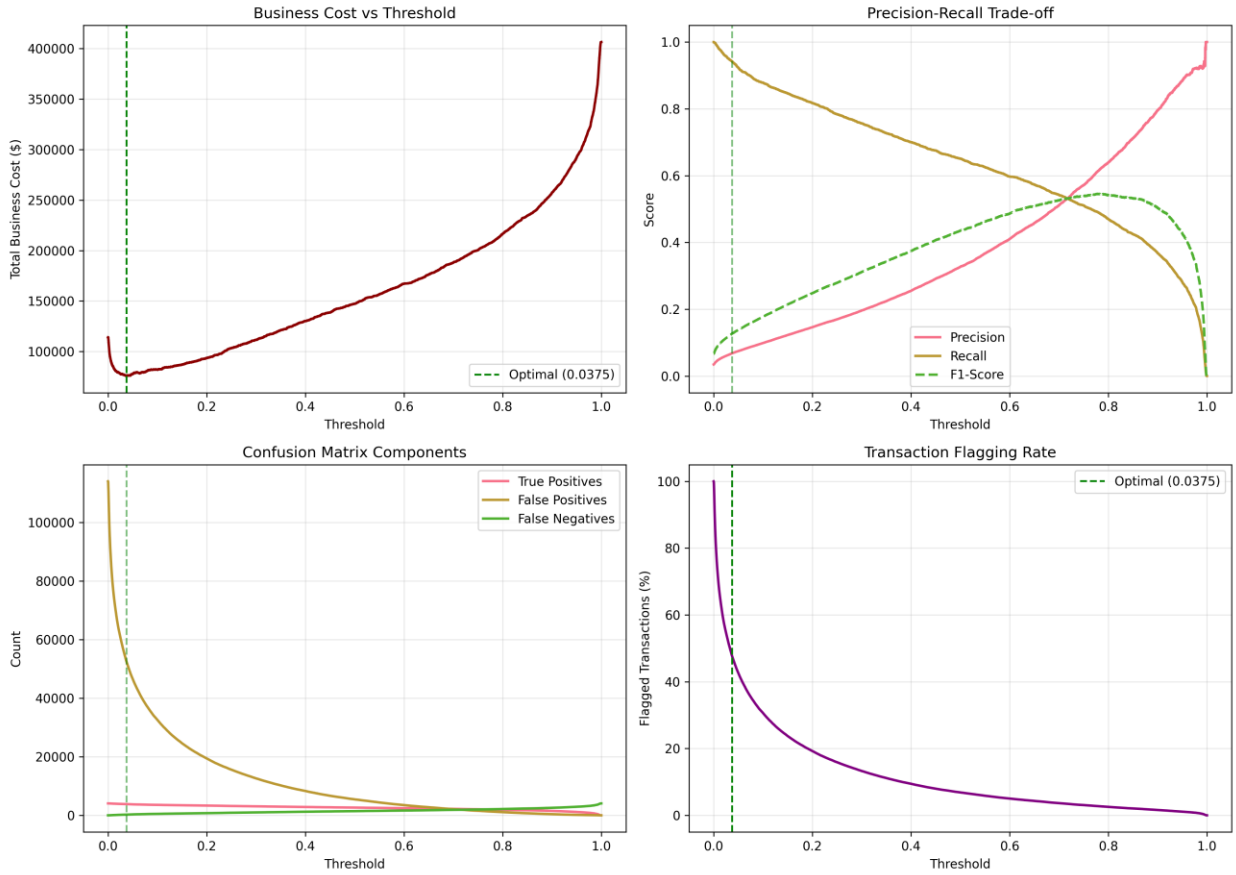
Optimal Threshold (Simulated)

- Threshold: **0.0375**
- Recall: **94.2%**
- Precision: **6.8%**
- Transactions flagged: ~47.6%

While aggressive, this threshold maximized net business value under stated assumptions.

Final threshold approval is a business governance decision and must be revalidated under real production volumes.

Figure 6: Operational Impact of Threshold Selection



Interpretation:
Lower thresholds materially increase recall but also analyst workload, making threshold selection a business governance decision rather than a modeling choice.

8. Human-in-the-Loop Controls

Function	Automation Level
Risk Scoring	Automated
Case Prioritization	Automated
Fraud Decision	Human Investigator
Blocking / Reversal	Human Only

Analyst overrides are logged and monitored to detect model or process issues.

9. Deployment Readiness

Phased Rollout Strategy

Phase	Traffic	Description
Phase 0	0%	Shadow mode (logging only)
Phase 1	10%	Conservative threshold
Phase 2	50%	Balanced threshold
Phase 3	100%	Full production

Current Approval: Phase 0 only.

10. Monitoring Framework

Key Metrics

- Model performance: PR-AUC
- Drift detection: PSI
- Operational latency (p95)
- Business impact: fraud dollars prevented

Predefined alert thresholds, escalation paths, and rollback procedures are documented.

11. Model Risk & Governance

- Overall Risk Level: **Medium–High**
- Materiality: **High ($6M+ exposure)**
- Governance Framework: **SR 11-7–aligned**

Eight material model risks were identified, including data drift, proxy bias, and over-reliance risk. All risks are mitigated or controlled via documented procedures.

12. Regulatory Alignment

Framework	Status
SR 11-7	Compliant
PCI-DSS	Compliant
SOX	Compliant
GDPR / CCPA	Partial (explainability implemented)
FCRA	Not applicable

13. Limitations & Disclaimer

This system:

- Is **not production-ready**
- Has not undergone independent third-party validation
- Must not be used for automated adverse actions

Intended Use

✓ Fraud risk scoring
✓ Analyst decision support
✓ Case prioritization

Prohibited Use

✗ Automated transaction blocking
✗ Account termination
✗ Credit scoring or lending decisions
✗ Clinical or medical decision-making

14. Conclusion

This project demonstrates how machine learning can be applied responsibly in regulated environments to deliver **measurable business value without compromising governance, compliance, or human oversight**.

The design reflects **Principal Consultant–level delivery standards**, suitable for organizations operating under regulatory scrutiny.