

# Portfolio Risk Narrative (Business-Focused)

## Executive Summary

A fraud detection model was developed to assess transaction-level fraud risk using both latent anomaly signals and engineered behavioral features. The model demonstrates that fraud risk within the portfolio is driven by:

- Strong latent anomaly components (PCA-derived features)
- Temporal transaction behavior (hour of transaction)
- Transaction magnitude and spending deviation patterns

The model captures both structural anomalies and behavioral irregularities, making it robust for real-world fraud detection.

---

## 1. Portfolio Risk Drivers

### A. Core Structural Risk (Latent Components)

SHAP analysis shows that the strongest fraud signals originate from PCA-transformed variables:

- V14
- V4
- V12
- V1
- V3

These represent hidden transaction structure patterns such as:

- Abnormal merchant interactions
- Network-based transaction irregularities
- Behavioral shifts relative to typical customer activity

These components form the **primary fraud detection backbone** of the portfolio.

---

### B. Temporal Risk Exposure

The feature `Hour` is among the top SHAP contributors.

This indicates:

- Fraud likelihood varies significantly by transaction time.

- Certain hours carry higher fraud risk.
- Time-of-day patterns are meaningful portfolio risk amplifiers.

Implication:

The portfolio exhibits time-dependent vulnerability, which could be leveraged for dynamic risk thresholding.

---

## C. Transaction Magnitude & Behavioral Deviation

Key behavioral features:

- Amount\_log
- Amount\_Deviation
- Raw Amount (in top 15)

Insights:

- Large transaction magnitude increases fraud probability.
- Transactions deviating significantly from historical behavior increase risk.
- Fraud is not only about large amounts — it is about abnormality relative to baseline behavior.

Implication:

Fraud risk in the portfolio is strongly linked to **behavioral deviation rather than absolute size alone**.

---

## 2 Risk Concentration Insights

SHAP distribution shows:

- No single feature fully dominates prediction contribution.
- Risk is distributed across multiple structural and behavioral signals.
- Model dependency is diversified rather than concentrated.

This reduces model fragility and suggests stable fraud detection performance.

---

## 3 Portfolio-Level Interpretation

Fraud in this portfolio appears to be driven by:

1. Structural anomalies embedded in transaction feature space
2. Time-based behavioral irregularities
3. Spending pattern deviation

This indicates the fraud pattern is sophisticated and multi-dimensional rather than rule-based or simplistic.

---

## 4 Strategic Recommendations

Based on model insights:

- Implement time-sensitive risk thresholds.
  - Monitor high-deviation spending behaviors more aggressively.
  - Combine anomaly detection with behavioral profiling.
  - Consider adaptive thresholds during high-risk hours.
- 

## 5 Overall Conclusion

The model successfully captures:

- Deep anomaly structure
- Behavioral irregularities
- Temporal risk dynamics

This creates a well-balanced fraud detection framework suitable for portfolio monitoring and risk segmentation.

---

# Technical Data Science Report

This section documents the complete modeling pipeline and methodology.

---

## 1 Data Understanding

- Binary classification problem (Fraud vs Non-Fraud)
- Highly imbalanced dataset
- Features include PCA components (V1–V28), transaction amount, and timestamp-derived variables

Objective:

Predict fraud probability and interpret model behavior.

---

## 2 Data Cleaning

Steps performed:

- Checked for missing values
- Verified class distribution imbalance
- Ensured no data leakage between training and testing sets
- Converted timestamp into interpretable time-based features (Hour)

No major data corruption detected.

---

## 3 Preprocessing

### A. Class Imbalance Handling

Applied:

- SMOTE (Synthetic Minority Oversampling Technique)

Purpose:

- Address severe fraud class imbalance
  - Prevent model bias toward majority class
- 

### B. Feature Scaling

- PCA features already scaled
- Applied log transformation to Amount → Amount\_log

- Reduced skewness in transaction magnitude
- 

## 4 Feature Engineering

Created behavioral variables:

- `Amount_log` → reduces skew
- `Amount_Deviation` → deviation from baseline transaction behavior
- `Hour` → temporal behavior
- `Night_Transaction` (binary indicator)

These features introduce behavioral fraud detection capability beyond raw anomaly signals.

---

## 5 Modeling

Model Used:

- XGBoost Classifier

Why XGBoost:

- Handles non-linear relationships
- Robust to multicollinearity
- Works well with imbalanced data
- Strong performance in fraud detection tasks

Hyperparameters:

- `n_estimators = 200`
  - `max_depth = 5`
  - `learning_rate = 0.1`
  - `eval_metric = logloss`
- 

## 6 Model Evaluation

Evaluated using:

- AUC-ROC
- Confusion matrix
- Risk segmentation analysis

Observed:

- Strong discrimination between fraud and non-fraud
  - Risk score distribution shows meaningful separation
  - No excessive overfitting detected
- 

## 7 Feature Importance Analysis

### A. Gain-Based Importance

Initially showed heavy dominance of V14 (~58%).

Limitation:

Gain can exaggerate early split features.

---

### B. SHAP-Based Importance (Preferred)

Computed mean absolute SHAP values:

Top Contributors:

- V14
- V4
- V12
- Hour
- V1
- V3
- Amount\_log
- Amount\_Deviation

SHAP confirms:

- Multiple structural features drive fraud prediction.
  - Behavioral and temporal features contribute materially.
  - Model is not dependent on a single feature.
- 

## 8 Model Interpretability

SHAP analysis provides:

- Global feature ranking
- Local transaction-level explanations
- Contribution magnitude per feature

This ensures regulatory transparency and explainability readiness.

---

## 9 Risk Segmentation

Final risk scores were bucketed into segments.

Observations:

- Majority of transactions cluster in low-risk bins.
- Fraud concentration increases sharply in upper risk intervals.
- Risk score distribution aligns with fraud probability output.

This supports threshold-based fraud decision systems.

---

## 10 Model Stability Assessment

- SHAP distribution confirms distributed contribution.
  - Behavioral features reinforce anomaly signals.
  - No artificial dominance from engineered variables.
  - Model generalizes beyond single-dimensional anomaly detection.
- 

## ⑩ Final Technical Conclusion

The modeling framework:

- Properly handles imbalance
- Incorporates behavioral intelligence
- Uses robust gradient boosting
- Provides interpretable SHAP-based explanation
- Produces stable multi-factor fraud detection

This is a production-aligned fraud modeling architecture.