FRAUD PATTERN DETECTION FOR A FINTECH PLATFORM USING MINITAB DATA MINING

1. Overview

Client:

A mid-sized fintech company offering peer-to-peer payments, wallet services, and micro-lending products in the United States

Objective:

To detect statistical patterns and anomalies in transaction data using Minitab's data mining capabilities, supporting fraud detection and lowering false alert rates.

2. Background

The client's internal fraud detection system generated an overwhelming number of false positives, creating friction for legitimate users and operational overload for the compliance team. They sought a statistical partner to help uncover more accurate fraud patterns using existing transactional datasets. Minitab was chosen for its robust visualization, segmentation, and anomaly detection tools.

3. Data Summary

Dataset Scope:

Random sample of 520,000 transactions from the last 9 months (2023)

Dataset Fields:

Variable	Type	Description
Transaction_ID	Identifier	Unique transaction reference
Amount_USD	Continuous	Transaction value in USD
Time_of_Day	Categorical	Binned: Morning, Afternoon, Evening, Night
Location_Match_Flag	Binary	1 = Location matches user history, 0 = New location
Device_Type	Categorical	Desktop, Mobile, Tablet
Transaction_Type	Categorical	Transfer, Withdrawal, Bill Payment, Merchant Purchase

Prior_Fraud_Flag	Binary	1 = User previously flagged, 0 = Clean
Fraud_Label	Binary	1 = Confirmed fraud, 0 = Not fraud

4. Methodology

Software Used:

Minitab 21

Key Goals:

- Isolate statistically significant fraud indicators
- Segment transactions with anomalous traits
- Support feature design for machine learning pipeline

Steps in Minitab:

1. **Exploratory Data Analysis (EDA):**
    - Descriptive statistics and frequency distributions
    - Boxplots of transaction amounts for fraud vs. clean labels
2. **Logistic Regression:**
    - *Stat > Regression > Binary Logistic Regression*
    - Modeled fraud label as a function of amount, time, device, and location behavior
3. **Chi-Square Test:**
    - *Stat > Tables > Cross Tabulation and Chi-Square*
    - Assessed relationships between device type, transaction type, and fraud flag
4. **Cluster Analysis (Supplementary Insight):**
    - *Stat > Multivariate > Cluster Observations*
    - Segmented transactions based on behavior profiles to identify high-risk groups
5. **Residual and ROC Analysis:**
    - Evaluated regression model performance and feature influence

5. Findings

Significant Predictors of Fraud (p < 0.01):

Feature	Odds Ratio	Interpretation
New Location Use	3.1×	Transactions from unfamiliar locations were 3 times more likely fraudulent
Prior Fraud Flag	5.6×	Strongest indicator for fraud reoccurrence
Nighttime Transactions	2.3×	Increased likelihood of fraud compared to daytime
High Value (> \$750)	1.9×	Outlier amounts more susceptible to fraud attempts

Model Performance:

- **AUC (ROC Curve):** 0.87
- **Classification Accuracy:** 82.6%
- **False Positive Rate Reduction:** 21% compared to previous heuristic system

6. Visual Outputs (Created in Minitab)

- **Boxplot:** Amount distribution by fraud label
- **ROC Curve:** Logistic model evaluation
- **Cluster Plot:** Transaction profile segmentation
- **Bar Chart:** Fraud frequency by time of day
- **Pareto Chart:** Feature importance ranking based on fraud frequency

7. Results & Implementation

- Based on the regression insights, **a new fraud scoring layer** was built into the client's rule engine
- Transactions with specific risk combinations (e.g., high value + new location + night) were flagged with **priority tiering**
- Customer trust score thresholds were updated using output from cluster analysis

- Post-deployment monitoring showed **a 21% drop in false positive alerts**, reducing review workload by ~30 hours/week

8. Recommendations

- Automate location history tagging at the device level for improved behavioral tracking
- Use Minitab control charts to monitor fraud rates by transaction type in real-time
- Integrate transaction clustering as a dynamic segment for future risk scoring updates
- Continue data mining quarterly to update thresholds and detect fraud pattern evolution

9. Future Scope

- Extend logistic regression model with interaction terms and nonlinear predictors
- Create monthly anomaly dashboards for compliance and fraud teams
- Integrate clustering-based anomaly alerts with live transaction feeds
- Explore ensemble learning integration (Minitab → Python pipeline for advanced ML)

10. Strategic Value

- Transformed legacy heuristic-based fraud detection with statistical modeling
- Provided evidence-based thresholds for dynamic rule generation
- Improved compliance efficiency while preserving legitimate user experience