# Comprehensive Guide: Automated Fraud Rule Generation for Credit Card Issuers

## 1. Introduction

Credit card fraud is a persistent challenge for issuers, impacting financial institutions, customers, and merchants alike. Fraudsters constantly evolve their techniques, making it essential for issuers to deploy automated, data-driven fraud detection systems. This document provides a structured approach to fraud rule generation, focusing on identifying fraudulent transactions, optimizing rule effectiveness, leveraging AI for real-time decision-making, and balancing security with customer experience.

## 2. Identifying Fraud Rules :

Identifying the fraudulent activities and applying fraud rules over them is the first layer and we can call it as “Problem Discovery Layer”. We have broadly classified them under the below three categories:

### 2.1 Issuer-Specified Rules (Manual Definitions)

Issuers most of the time knows what exact rules they want they have the specific parameters based on which they want to generate and apply the fraud rules.

Examples:  
- Transactions exceeding $10,000 require real-time multi-factor authentication (MFA).  
- Decline transactions from high-risk countries unless a travel notice is registered by the cardholder.  
- Transactions with a merchant category code (MCC) related to cryptocurrency or gambling are automatically flagged for additional review.

### 2.2 Emerging Fraud Hotspots & Dynamic Risk Adjustments

Fraud trends shift over time, necessitating continuous monitoring and real-time rule updates to address emerging threats. Our system should be able to identify an over-time or growing hotspot and generate the rule to help the issuer identify the fraud at the very beginning itself.

A screenshot of a phone

AI-generated content may be incorrect.

Example:

If we will carefully analyse the above chart spending in the small city like “Bhilai” has increased suddenly means there could be some fraud happening in that region so our system should be able to identify the pattern at the very beginning to avoid the fraud .

### 2.3 Behavioral & Reference Cohort-Based Rule Generation

Comparing a cardholder’s spending patterns against reference groups (cohorts) allows issuers to detect deviations indicating potential fraud. We can call it as Snapshot view which provides a comparative analysis between treatment and control groups.

Examples:  
- If a cardholder typically spends $2,000 per month but suddenly charges $15,000 in a single day, flag it for manual review.  
- Compare a user’s location history with past transactions: if they have never made transactions outside their home country and suddenly attempt purchases in multiple foreign cities within hours, trigger an authentication request.

A graph of a bar chart

AI-generated content may be incorrect.

## 3. Defining & Optimizing Fraud Rules

Fraud detection rules should be optimized by balancing efficiency and relevance

### 3.1 Pruning Ineffective & High False Positive Rules

Excessively strict fraud rules can generate false positives, leading to unnecessary transaction declines and customer dissatisfaction. Regular audits help refine these rules for better accuracy.

Examples:  
- Old Rule: Decline all transactions from non-domestic locations by default.  
- Optimized Rule: Decline transactions from non-domestic locations only if the device and IP address are also new, or if multiple declined attempts were made in a short timeframe.

### 3.2 Adaptive Rule Enhancement for New Fraud Techniques

Fraudsters continuously change tactics. Adaptive fraud detection frameworks replace outdated rules with ones tailored to emerging risks.

Examples:  
- Old Rule: Flag transactions above $3,000 from a new location.  
- New Rule: Flag transactions above $3,000 from a new location \*\*only if\*\*:  
 - The transaction is card-not-present (CNP), AND  
 - The device has not been previously associated with the card, AND  
 - The transaction occurs within minutes of another flagged transaction.

## 4. Evaluating Rule Effectiveness

Fraud detection rules should be measured using key performance indicators (KPIs) to assess effectiveness and minimize unnecessary declines.

1. **Default Metrics for Evaluation**
   * **False Positive Rate (FPR):** Measures legitimate transactions incorrectly flagged as fraud.
     + Example: A customer attempts to purchase a laptop online, but the system incorrectly flags it as fraud, requiring additional verification or declining the transaction.
   * **False Negative Rate (FNR):** Identifies fraudulent transactions that bypass detection.
     + Example: A fraudster successfully makes several small transactions that mimic a cardholder’s spending pattern, allowing them to remain undetected.
   * **Fraud Capture Rate:** Percentage of detected fraud cases versus total fraud attempts.
     + Example: Out of 1000 fraud attempts, the system correctly identifies and blocks 900, yielding a fraud capture rate of 90%.
   * **Transaction Approval Rate:** Ensures that legitimate transactions are not excessively declined.
     + Example: A bank optimizes fraud rules to minimize unnecessary declines, ensuring that legitimate cardholders face minimal transaction friction.
   * **Customer Complaint Rate:** Tracks dissatisfaction due to false positives.
     + Example: If a bank receives numerous complaints about valid transactions being blocked, adjustments to fraud rules may be necessary to improve customer experience.
2. **Stability of Boolean Expressions Over Time & Segments:** Fraud rules defined using Boolean logic should remain stable across different time periods and customer segments to ensure consistency.
   * Example: Rules based on seasonal spending behaviors should be adaptive but stable across customer demographics.
3. **Sensitivity to FPR & Other Metrics** Fraud rules should be stress-tested for sensitivity to changes in values underlying a Boolean expression.
   * Example: If a rule flags all transactions above $1,000, how does increasing or decreasing the threshold affect the false positive rate?

**5. Integrating GEN -AI:**

Gen-AI can enhance fraud detection in multiple ways:

1. **Orchestrator** : AI can serve as an orchestrator, dynamically adjusting fraud rules based on real-time transaction data.
2. **Human-Readable Report:**  AI can generate explainable fraud detection reports, summarizing flagged transactions and providing justifications.
3. **Feature Engineering for Fraud Detection:** AI can engineer features that are not explicitly present in transactional data but enhance fraud detection, such as:
   1. Behavioral risk scores.
   2. Device and geolocation-based anomaly tracking.
   3. Temporal patterns in customer spending.