International Journal of Research in Engineering and Science (IJRES)
ISSN (Online): 2320-9364, ISSN (Print): 2320-9356
www.ijres.org Volume 12 Issue 3 ǁ March 2024 ǁ PP. 164-168

Customer Data Fraud: A Rule Engine Perspective

Khirod Chandra Panda
Principal Engineer
khirodpanda4bank@gmail.com

Abstract *- Fraud in the realm of customer data poses a significant challenge. Customers interact with organizations of all sizes, seeking assurance that their data and digital footprint remain secure. Traditional rule engines, despite their complexity, have been instrumental in identifying and thwarting predefined patterns of fraud, offering a level of protection. However, as fraudsters continually adapt, tools for detecting their malicious intent evolve as well. Integrating machine learning and advanced AI into rule engines can enhance their effectiveness, particularly in scenarios where fraud patterns change over time. By combining supervised and unsupervised learning techniques with rule-based systems, organizations can improve their fraud detection capabilities.*

Keywords *–Business Rules Management System (BRMS), Business Rules Engine (BRE), Subject Master Expert (SME), Event Driven Architecture (EDA)*

Date of Submission: 05-03-2024 Date of acceptance: 18-03-2024

I. Introduction

When we talk about data, it usually means digital form of information stored in a way that can be moved or processed further. Customer data [1] is defined as the information a customer provides while interacting with a business entity via your website, mobile applications, surveys, social media, marketing campaigns or any other online and offline avenues. Simply defined, Data is collection of facts. Think of raw form of data. Given the technology power when the data is organized with context it becomes information that can be acted upon. Customer data is a cornerstone to a successful business strategy. All most all business entity or a Data-driven entity realize the importance of this and take action to ensure that they collect the necessary customer data points. By understanding a customer profile, organization can find out the specific demographic groups that are interested in their product, so can tailor marketing campaigns to them more effectively and hence Increase customer retention improve customer experience and fine-tune business strategy over time. organization collects a myriad of customer data points throughout the buyer's journey. Through consumer behavior and predictive analytics, companies regularly capture, store and analyze large amounts of quantitative and qualitative data on their consumer base every day. Some companies have built an entire business model around consumer data, whether they sell personal information to a third party or create targeted ads to promote their products and services.

The volume of these data points is vast, and for ease of understanding, it can be segregated in 4 different categories.

Personally identifiable Information (PII). This category includes personally identifiable information such as Social Security numbers and gender, as well as nonperson ally identifiable information, including your IP address, web browser cookies and device IDs (which both your laptop and mobile device have).

Engagement data. This type of data details how consumers interact with a business's website, mobile apps, text messages, social media pages, emails, paid ads and customer service routes.

Behavioral data. This category includes transactional details such as purchase histories, product usage information (e.g., repeated actions) and qualitative data (e.g., mouse movement information).

Attitudinal data. This data type encompasses metrics on consumer satisfaction, purchase criteria, product desirability and more.

In credit card domain, fraud transpires when an unauthorized individual obtains access to your information and uses it for purchases. Any business, regardless of its size, is susceptible to credit card theft and fraud, presenting a significant risk. Here are various scenarios in which such fraudulent activities can occur:

Lost or stolen cards: Criminals can acquire credit cards either by discovering lost cards or by stealing them directly from individuals.

Card-not-present fraud: This type of fraud doesn't require physical possession of the credit card. Fraudsters obtain key details such as the cardholder's name, card number, and expiration date, enabling them to commit fraudulent transactions through mail, phone, or online channels.

Counterfeit, doctored, or faked cards: Criminals use devices known as skimmers to illicitly capture credit card details from the magnetic strip. This information is then encoded onto counterfeit, doctored, or faked cards. Detecting a skimmer can be challenging,

Application fraud: Instead of stealing existing credit card information, criminals may apply for new credit in someone else's name, using the victim's personal information. This may include their full name, date of birth, address, and Social Security Number, along with stolen supporting.

II. Literature Review

Be it consumers like you and I or business organizations ranging from small to big enterprises, all face day to day challenge in detecting fraud and taking corrective action. With the ever-evolving technology land scape the fraudsters become more smart / intelligent and try new way of committing fraud to gain instant fame or monetary gain. It becomes complex and herculean task keep fraud at bay by just looking at data and when organization involved with millions or billions of data points, it becomes necessary to look at some automated process which can take various data points into account and provide a recommendation as to if the transaction that is about to happen is legitimate or is there a fraudulent intention is behind that.

Study [3] suggests, organizations often using more than one application to detect fraud. The application used are either specialized in one area or sometimes organizations use multiple platforms specialized in same core area and based on reliability take the response from an application vs other. Manual review of prediction is an integral part of detection.

A business rule is an expression of a business policy. You package up business rules into rulesets. The rule engine takes a ruleset and executes it against a set of objects. Study [4] suggests Rule engines (A methodology used form ages) is still an integral part of fraud detection. Although more modern technology has gathered advancement like Machine Learning blended with Artificial intelligence.

Along with all these detection process, study [3] suggests manual intervention in the process is quite common till date in the industry (Figure 1 above)., though organizations want to move out of manual intervention as the automated fraud detection model becomes more tuned and works as per organization standard.

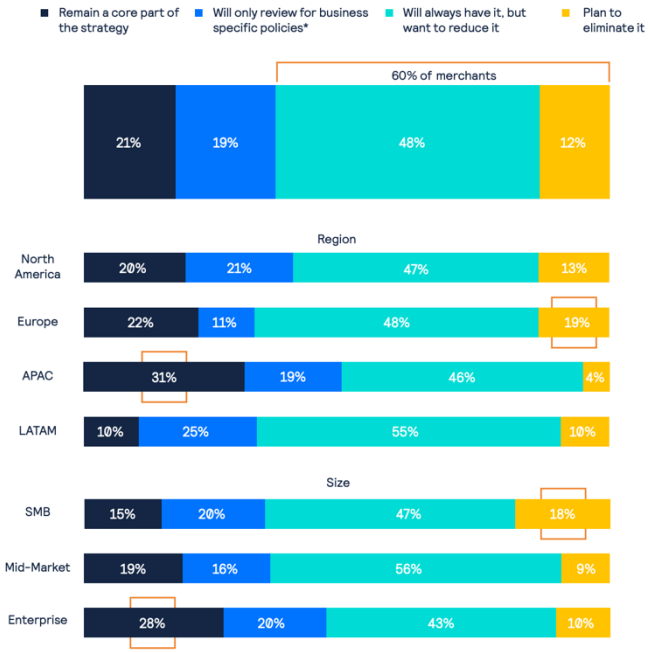

Fig. 1 Role of Manual Review in Future Fraud Management strategy

III. Methods Used in Fraud Detection

3.1. Rule Engines

Rules-based fraud detection detects fraud by analyzing various attributes such as unusual time stamps, account numbers, transaction types, and amounts, among other criteria. Configured Rules can be thought of as a reactive way of detecting fraud. Typically, Subject Master Expert (SME) who have extensive domain knowledge are tasked with managing these rules. An enterprise grade application is used which hosts these rules. Any transaction initiated by POS machine is filtered through the set of predefined rules. In credit card industry, generally POS machines are used for initiating a transaction in any business establishment, Various data around the merchant like merchant category (restaurant, gas station, wholesale food etc.) are used for fraud

detection. The Merchants are provided a Category based on their classification called MCC code. A high-risk merchant category code refers to a specific four-digit number assigned to businesses within certain industries that are considered to carry a higher level of risk. The classification of an MCC code as high risk is typically based on factors such as the nature of the products or services offered by the merchants, historical data of fraudulent activities within that category, Merchant geo location and regulatory guidelines. Identifying and monitoring high-risk MCC codes is crucial for effective risk management and fraud prevention.

The merchants are provided the category code when they apply for and establish a network connection with major card networks as Master Card, Via, American Express. Study [6] suggest they are given various categories and are governed by that category specific rules. Furthermore, transactions may be categorized as Card present transaction vs card not present transaction (like using of card details online to purchase product or services). When a customer swipes his card in the POS machines (Card present transaction) this is a crucial data for fraud rule. likewise, the geolocation where card is swiped from the location of the card holders residential address, current amount of transaction vs users' historical transaction amount is a data point to be considered.

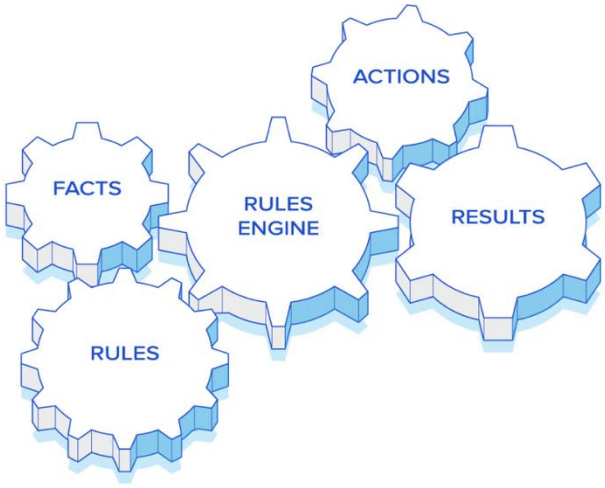

Fig. 2 Rule engine setup

All these pertinent details (meta information) are captured along with the transaction details and are validated against a set of rules defined in the rule engine as shown in Figure 2 above.
There are numerous tools that are paid tools to help set up the rules and integrated with transaction fraud management applications to detect fraud. Drools [7] (an opensource tool from Apache) or decisioning rules is a BRMS solution that helps in providing a BRE.

3.1.1 Rules in a glance.

```
rule "Name of rule"
when
                "The if" part of the business logic goes here.
then
                The "then" part of the business logic goes here.
end
```

Figure. 3 Basic rule authoring

SME's write the rules using the Drools User Interface. Application code will be responsible for loading appropriate facts into Working Memory so that SMEs can write rules that query these facts. Only facts relevant to application business logic should be loaded into Working Memory, to keep the rules engine running at top speed.

3.1.2 Enhancing rules with Eventing.

Rules work as a template against which the facts are run. For the decision to happen in real time basis and in a quicker manner, we need further enhancement to rule engine. like eventing from various data sources. For example, when a transaction happened various systems listen to the data and execute further computation on the data. the various response that comes back from the systems that are listening is sent.

For instance, on a Stockbroker application, when a sale operation is executed, it causes a change of state in the domain. This change of state can be observed on several entities in the domain, like the price of the securities that changed to match the value of the operation, the ownership of the traded assets that changed from the seller to the buyer, the balance of the accounts from both seller and buyer that are credited and debited, etc.

Depending on how the domain is modelled, this change of state may be represented by a single event, multiple atomic events or even hierarchies of correlated events. In any case, in the context of this guide, Event is the record of the change of a particular piece of data in the domain.

Events are processed by computer systems since they were invented, and throughout the history, systems responsible for that were given different names and different methodologies were employed. It wasn't until the 90's though, that a more focused work started on EDA (Event Driven Architecture) with a more formal definition on the requirements and goals for event processing. Old messaging systems started to change to address such requirements and new systems started to be developed with the single purpose of event processing. Two trends were born under the names of Event Stream Processing and Complex Event Processing.

This engine offers the capability to identify, link, simplify, combine, and respond to events. Essentially, it provides methods to deduce complex events from basic ones, respond to relevant events, and initiate actions. The key distinction between Complex Event Processing (CEP) and typical rule execution lies in their treatment of time. While standard rules execution in Decision Manager revolves around facts and the logic applied to them, the CEP engine centers on events. An event signifies a notable change in status at a specific moment or within a specific timeframe.

As rules are pre-defined, its necessities to be on top of all trending fraud pattern to configure the rules so that patterns can be detected, and necessary corrective action can be taken. This is a time consuming and resource intensive work. We need SME to always looking for patterns and continually working on them. With help of Machine learning we can train system to look at data points and take necessary action.

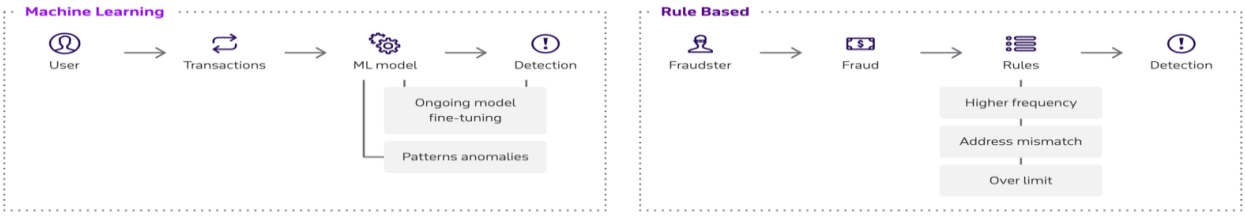

Fig 4 Rule Vs Machine learning detection

3.1.3 Rules with Machine Learning For fraud detection

Machine learning for fraud detection is the practice of using computers to analyze and learn from data to make informed decisions about fraudulent activity. It's the most efficient, accurate way to achieve the results (Figure 4 above).

3.1.3.1. An interaction happens.

A user interacts with your business — anything from creating an account, requesting a quote, or making a purchase.

3.1.3.2. Data is collected.

The technology gathers information about the user — such as device type, IP address, location, shopping behavior, and more.

3.1.3.3 Data is reviewed.

The technology analyzes the data using two types of machine learning models — supervised and unsupervised.

3.1.3.4. Fraud risk is determined.

The technology looks for red flags that could signal fraud and determines the interaction's level of risk.

3.1.3.5. Policies are consulted.

The technology consults your business policies to check acceptable risk thresholds.

3***.1.3.6. Fraud is stopped.***

The technology blocks, accepts, or flags an interaction for further review.

3.1.3.7 Interaction data is recorded.

The outcome of the interaction is recorded so that machine learning algorithms can learn and improve decision accuracy. And unlike the human brain, the more data you feed a machine learning algorithm, the better and more accurate it becomes.

That said, ML is not always perfect. Drawbacks of ML for fraud detection include the potential for false positives, describing when a system mistakenly marks legitimate actions as fraudulent. False positives open the possibility for a negative feedback loop—if one detection mistake isn't spotted, the algorithm thinks it responded correctly and the behavior was legitimate, teaching itself to repeat the same response in the future.

When new fraud pattern suddenly emerges, it become difficult to quickly train an algorithm. The data must be collected and to make sure data is diverse is nature. These necessities a way to detect fraud some other way and analyze the data pattern till sufficient data across different angle is collected and then feed to a model

to accurately determine the response. In this scenario, Rule engine plays a big role. As data scientist and SME can work in tandem to stop the fraud and do manual review of samples while in the process harvesting data for model to train upon.

Thankfully, human insight (Manual review along with predefined rules) alongside machine learning can help overcome this problem.

IV. Conclusion

Traditional fraud detection systems have serious limitations:

First, they're based on static rules. While the rules may work great initially, they become less useful over time as technology evolves and attack methods change. Bad actors want to achieve their goals with as little effort as possible, so they won't waste resources trying the same approach that doesn't work again and again—they will find a way around the static barrier.

Traditional systems rely heavily on human labor, so they are limited by the expertise, time, and energy of the people who create and manage their rules. Manually operated systems can eventually become so complex that it is nearly impossible for new users to understand how to manage them.

Machine learning for fraud detection solves these issues. ML is faster, more accurate, and more cost-effective, eliminating the need for a human to supervise every decision, processing new data automatically, and updating detection models in real-time.

References

[1] Indrajeet Deshpande, what is customer Data? Definition, types, collection, validation, and analysis - Spiceworks. (2021, March 16). Spiceworks. Blog. [Online]. Available: https://www.spiceworks.com/marketing/customer-data/articles/what-is-customer-data

[2] cybersource.com [PDF document]. Available: https://www.cybersource.com/content/dam/documents/campaign/fraud-report/global-fraud-report-2022.pdf

[3] Fraud.com International. (2023, September 25). Are fraud rule engines a thing of the past? [Online]. Available: https://www.fraud.com/post/fraud-rule-engines-a-thing-of-the-past

[4] Honick, L. (2024, February 26). Common High-Risk MCC Codes [year] Update. Host Merchant Services. Available: https://www.hostmerchantservices.com/articles/high-risk-mcc-codes/

[5] Drools - Business Rules Management System (JavaTM, Open Source). (n.d.). Available: https://www.drools.org/