INNOVATIVE DATA DRIVEN SOLUTIONS FOR FRAUD ANALYTICS

NATIONAL TECHNICAL INFORMATION SERVICE (NTIS)

NTIS offers premier skillsets from the industry, academia, and not-for-profit organizations to support meeting data missions in the Federal government through its unique Joint Venture Program Authority[1].

FEDERAL GOVERNMENT ADDRESSES FRAUD

The Federal government may get involved when such frauds occur in transactions related to, amongst others, the Financial Sector, Unemployment, or Healthcare related claims.

In March of 2023, the Biden administration asked Congress to give Federal watchdogs and prosecutors funding of $1.6 billion to go after billions of dollars in COVID-19 spending lost to fraudsters, stating that "The data shows that for every dollar we put into fighting fraud, the taxpayers get back at least 10 times as much," [President] Biden said last month."

Some examples of fraud in different industries include:

- **Healthcare Sector**: Misrepresenting diagnoses or procedures to maximize payments.
- **Financial Sector**: Creating synthetic identities by using a combination of real and fake personal information.
- **Education Sector**: Misrepresenting income status of one's family while filling the application form to file for aid.
- **Manufacturing sector**: Noncash frauds, including the stealing of intellectual property, such as trade secrets or technology.

TYPES OF FRAUD

First-party fraud is when someone misidentifies themselves or gives false information to appear eligible for certain services or money. Some examples of First-party fraud include Chargeback Fraud, Sleeper Fraud, Application Fraud, De-shopping, Goods Lost in Transit Fraud or Fronting.

As mobile transaction volume grew in payment transactions, the study[5] noted a 33% and 24% fraudulent activity increase in the retail and ecommerce segments, respectively.

Second-party fraud refers to fraud schemes in which a person knowingly allows a second party to use their identity or personal information to impersonate a valid user and commit fraud. Thus, by its nature, all Second-party fraud scams involve two perpetrators: the fraudster and an accomplice. Some of the common types of Second-party fraud include Money Muling, Fake Merchant Scams, Gift Card Laundering, and Leading the Witness.

Third-party fraud, generally known as identity theft, occurs when a person uses another person's identifying information to carry out financial transactions without the knowledge of the individual whose information is being used. Some examples of Third-party fraud include Account takeover fraud, Synthetic Identity Creation, False Identity Fraud, Credit Card Fraud, and New Application Fraud.

THE COST OF FRAUD

A recent report by LexisNexis[3] describes the cost of fraud as being more than the actual dollar value of a fraudulent transaction. In addition to the amount of the fraudulent transaction, it also includes additional costs related to labor incurred in the investigation, fees incurred during the applications, underwriting and/or processing stages, legal fees, and external recovery expenses. The total cost of fraud is expressed by saying that for every $1 of lost value due to fraud, the actual cost is higher based on a multiplier representing these additional costs.

FTC reported fraud losses increased more than 70 percent in 2021 to more than $5.8 billion[2].

- In the financial services and lending sector, the cost of fraud is highest amongst U.S. banks, where every $1 of fraud loss actually costs $4.36.
- In the mortgage firm segment, every $1 of fraud loss costs $4.20.

In the e-commerce segment[4], merchants have a Fraud Multiplier of $3.85 for every $1 of fraudulent activity.

FRAUD DETECTION AND PREVENTION THROUGH ANALYTICS AND DATA INSIGHTS

Fraud analytics refers to the use of data analytics in the context of fraud prevention that may range from screening available data to find anomalies – as signals outside of expected/normal ranges are considered suspicious or risky to big data techniques and machine learning to automatically detect and prevent fraud at scale.

INNOVATIVE DATA DRIVEN SOLUTIONS FOR FRAUD ANALYTICS

With today's massive volume of transactional data, much of which may be in real-time, it is not humanly possible to screen the data, which makes it necessary to implement modern computational techniques noted above. A second complexity is the fact that other types of data are needed for successful fraud analytics and prediction.

These include semi and unstructured data that is not ideal for analysis using traditional relational databases but may be better suited to machine learning and AI solutions. A Seagate keynote address[5] refers to an IDC prediction that the global datasphere will grow to 163 zettabytes by 2025 and majority of that will be unstructured and will be machine generated.

A third aspect of data in fraud analytics is the use of behavioral data, that provides information about customers' interaction with businesses - obtained through marketing systems, social media, websites, mobile apps, CRM systems, call centers, emails, and behaviors observed in a physical setting.

During the COVID-19 pandemic, unemployment insurance fraud may have been in excess of $60 billion[6].

Even with the hyperscale storage architecture platforms available today, only a portion of the generated data will be stored meaning that analytics may have to be performed while the bulk of the necessary data is still available.

Fraud Analytics, while highly complex, deals with some of the following regarding user access:

Identity validation: This is the process of checking a person to validate that they exist in the real world. This can be done by checking databases such as postal address files, phone records, or even basic credit data.

Identity verification: This is a much more in-depth assessment of linking a person to the information they provide.

Authentication: The process of checking a customer's identity against information that only the user should have or know. This is to check they are who they say they are.

From a transactional/data point of view, identifying fraudulent records or transactions involve:

Identification: Where analytics is used to look for anomalies that indicate potential fraud risk schemes and identify high-risk areas for inclusion in the fraud risk assessment.

Validation: Analytics is used to validate the identification of high-risk schemes, evaluate the accuracy of risk assessment process findings, and indicate the need for additional procedures.

Monitoring: Tests and tools are typically developed to continuously monitor high-risk schemes and behaviors, aid in assessing the effectiveness of the fraud action plan and provide proactive alerts for possible exceptions and violations on an ongoing basis.

In terms of computational approaches, a number of techniques may be used, including:

Statistical techniques: These include clustering and classification methods to find patterns and associations among groups of data.

Artificial intelligence techniques: These include data mining, machine learning (supervised and unsupervised), neural networks, expert systems, pattern recognition, and anomaly detection for fraud detection and prevention.

HOW NTIS CAN HELP

NTIS has supported HHS OIG and VA OIG with multi-year projects focused on Fraud Analytics.

It is evident that Fraud Analytics is a highly complex and challenging problem to address and yet we are obligated to address this growing concern and save taxpayer's money. NTIS, with its carefully selected pool of Partners (JVPs), can support projects requiring Fraud Analytics. For further information, please email BusinessOpportunities@ntis.gov

1. 15 USC 3704b: National Technical Information Service (house.gov)
2. New Data Shows FTC Received 2.8 Million Fraud Reports from Consumers in 2021 | Federal Trade Commission
3. 6th Annual True Cost of Fraud Study: Financial Services and Lending Report (U.S. and Canada Edition 2022)
4. The LexisNexis Thirteenth Annual True Cost of Fraud™ Study for Ecommerce and Retail
5. Datasphere 2020, We're Evolving: Seagate Corporate Strategy Update and Announcements
6. GAO Pegs Unemployment Insurance Fraud Tally at More Than $60 billion