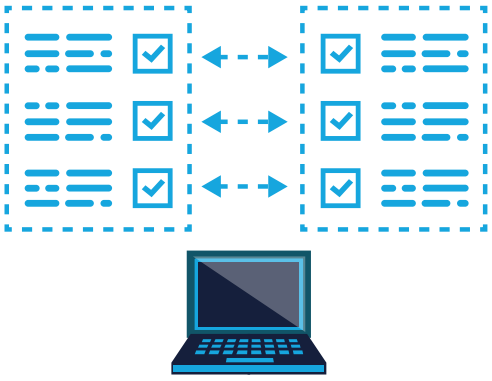
NEXT-LEVEL DETECTION THROUGH MACHINE LEARNING

Synthetic identity fraud impacts organizations worldwide. News headlines highlight cases in which these fraudulent identities steal money. Many financial institutions recognize the risks of synthetic identities and deploy prevention and detection measures to mitigate this type of fraud.

Synthetic identities are created by leveraging pieces of real consumer information (such as names, dates of birth and Social Security numbers) and fictitious information. These identities then are used to apply for financial accounts and products. When approval thresholds are met, these identities open checking accounts and qualify for credit cards. The fraudster may appear to act like a normal consumer while patiently building account and credit history with low dollar amounts and payments. The criminals behind these applications then max out credit or overdraw accounts with no intent to repay the amounts. Detection of synthetics through static rules is challenging, as it is not instantaneous to build the analytics required to identify changing patterns and implement rule changes. However, industry feedback indicates machine learning can successfully detect synthetic identity fraud, as the technology can "learn" fraud patterns at a much faster rate than human beings alone.



BENEFITS OF MACHINE LEARNING

Machine learning is a method of data analysis in which systems and computers identify patterns and generate decisions based on data and statistical models. Machine learning includes automated bots for online customer service, product recommendations and mobile traffic alerts. In fraud detection, machine learning can process large amounts of data without relying on pre-defined detection rules. Manual intervention can assist employee training, validate fraud model performance and improve models based on insights from known fraud cases.

One way to find synthetic identities in the application process or existing accounts is to use alternative (third-party) data to compare identity information to known "bad data." The volume of data is substantial because it can include multiple products, transaction data, applicant data, credit histories, public records and internet profiles. Fraud detection often is based on a set of rules by process and products. Those rules require regular manual monitoring and updates to react to the changing fraud environment. Thresholds and parameters need routine updates to include changes in fraud activity and risks, as well as to adjust alerts and false positive rates.

NEXT-LEVEL DETECTION THROUGH MACHINE LEARNING

The starting point for machine learning is to determine the data inputs needed for analysis and collect historical data useful for training and updating the model. These inputs include data collected during the application process, transactional data and other public/digital data sources for comparison. Applications and accounts can be reviewed for risk factors that may be characteristic of a synthetic identity. The large amount of data that can be used for comparison requires a fast, efficient processing solution to reduce the impact on potential or existing customers.

DATA INPUTS FOR MACHINE LEARNING MODELS SHOULD BE ESTABLISHED BASED ON THE BUSINESS NEED AND EXPECTED OUTPUT, AND MAY INCLUDE:

- Product data across the enterprise, such as checking and savings accounts, credit cards, loans, home equity lines of credit (HELOC) and merchant services.
- Applicant or account holder data, including credit history.
- Account holder details for other accounts, including device ID and IP address, that can be used to find matches that might indicate risk.
- Transaction details, including account contact data changes, purchases, payment activity, adding authorized users, fraud claims, account holder requests for credit increases and approved credit increases.
- Alternative or third-party data, such as public records and digital data, to compare to the account holder information for name, mailing address, birth date, phone number, email address and Social Security number.
    - Legitimate identities typically have a digital footprint, including previous addresses, phone numbers and social media accounts.

While machine learning offers promising efficiencies and insights, it is a complex technology and must include appropriate governance and human oversight. In particular, strong consideration should be given to data management and organization, inclusive of key elements such as quality, explainability and avoiding bias. Additionally, it is important that organizations using machine learning consider the real-world scenarios surrounding the data and rules and identify outputs that may need additional human input or manipulation.

NEXT-LEVEL DETECTION THROUGH MACHINE LEARNING

CONCLUSION

Detection of synthetic identities is complex given the nature of this type of fraud. No instant check can confirm that a name, date of birth and Social Security number matches a true identity. Fraud detection approaches that leverage machine learning offer the ability to relatively quickly process the large volume of data needed to compare an applicant or existing account holder with available records to find indicators of synthetic identity fraud while learning from the data. Automated/machine learning and model improvements can more effectively address evolving fraud threats. Static rules and negative lists are still useful components of a layered fraud detection framework, but often require frequent manual updates to their parameters and detailed data matching. Machine learning for synthetic identity fraud detection may provide faster and more accurate results – as well as greater agility to adapt to new fraud threats. As synthetic identity fraud expands, financial institutions have the option to consider machine learning for fraud detection.