DATAVISOR

5 Common Misconceptions about Unsupervised Machine Learning in Fraud Detection

Companies have been using rules engines and machine learning in fraud detection for decades. The simplicity of rules engines and the agility to take quick action have made it an indispensable part of the fraud management toolkit. Machine learning has also been adopted by organizations for decades — approaches with which organizations are most familiar range from neural networks to supervised learning.

The problem, according to DataVisor co-founder and CTO Fang Yu, is that traditional rules-based fraud detection isn't enough to capture all instances of fraud. Because fraud patterns and activities can change rapidly, it's important that fraud detection tools adapt just as rapidly to meet these challenges as they occur in real time.

Machine learning has entered the fraud detection toolbox to help companies keep pace with evolving threats. The two most common types of machine learning forms now are supervised machine learning (SML) or unsupervised machine learning (UML).

Supervised machine learning enables algorithms to learn from existing data and apply that knowledge to new data. For it to work effectively, SML only works when fraud patterns are known and accurate historic loss labels are available. Rules-based engine cannot detect new or evolving threats until being trained to do so. This is where UML holds the greatest advantage. UML doesn't rely on extensive data training and rules, and it can self-adapt to new threats as they emerge.

It's no surprise that the benefits of unsupervised machine learning approaches aren't yet well known, and there are common misconceptions about the usage, effectiveness, and explainability of this approach.

It's time to put these misconceptions to rest and get a better understanding of how UML works and its role in proactive fraud detection.

1 MISCONCEPTION #1
UML causes high false positives

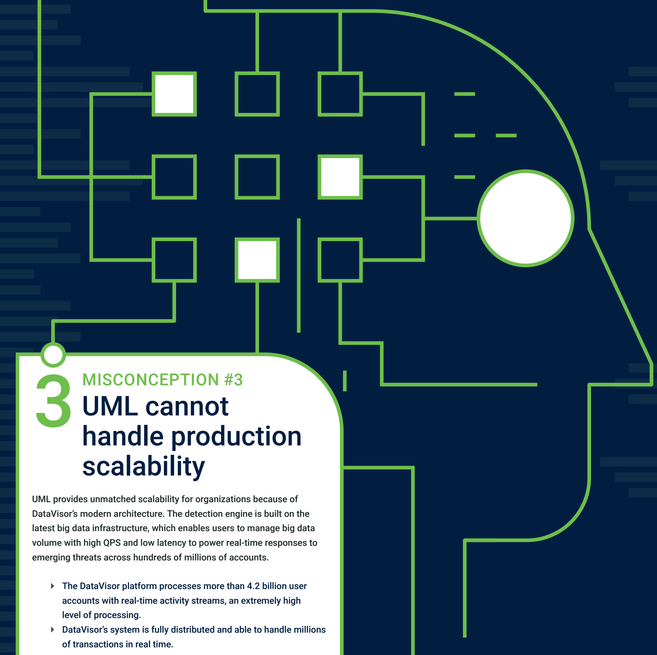
UML does not require labeled data input and is designed to discover patterns within large amounts of unlabeled data. This enables the discovery of new and unknown threats. This often leads to the misconception that using UML results in false positives. That is not the case, as UML combines the power of anomaly detection, clustering analysis, and graph analysis to detect the relationship between anomalies or activities. It's this deeper level relationship-based insight that creates fewer false positives and detects more fraudulent activities than SML.

- The false-positive rate is much lower than with supervised machine learning because UML looks at data on a much larger scale instead of reviewing individual cases or relying solely on anomaly-based detection.
- Because UML is based on large-scale observations, behaviors and patterns that are common with large organized crime rings that would otherwise go undetected under single case review can be identified quickly, before fraud actually occurs.
- Since UML automatically identifies clusters of emerging threats in real-time, there is no human bias associated with model training.

2 MISCONCEPTION #2
UML is a black-box and cannot pass model governance

The goal of model governance is to ensure the quality and reliability of the model. To do this, users must understand how their fraud model work and how it arrives at certain conclusions. It's commonly believed that UML exists in a "black box" that you can't peek inside to see how it works, but that's not the case. Here's why:

- UML provides specific reason codes why a certain transaction or activity was flagged as fraudulent.
- These reasons are easily explained based on activities, behaviors, timing, and other factors.
- Data in a UML model isn't sensitive to outliers or data skews because data is dynamically observed in real time.
- Each feature in the DataVisor platform uses a unique algorithm to locate patterns within specific data sets, with each algorithm featuring multiple layers of engineering that feeds into a clustering algorithm that creates the model.
- DataVisor translates the data into visual representations to help users connect the dots.



3 MISCONCEPTION #3
UML cannot handle production scalability

UML provides unmatched scalability for organizations because of DataVisor's modern architecture. The detection engine is built on the latest big data infrastructure, which enables users to manage big data volume with high QPS and low latency to power real-time responses to emerging threats across hundreds of millions of accounts.

- The DataVisor platform processes more than 4.2 billion user accounts with real-time activity streams, an extremely high level of processing.
- DataVisor's system is fully distributed and able to handle millions of transactions in real time.

4 MISCONCEPTION #4
UML models are difficult to build and implement

DataVisor's platform has built-in flexibility that simplifies fraud modeling, including the option to build your own fraud models. Because UML doesn't require extensive training or data labeling, organizations can usually get set up and start seeing results in as little as two weeks.

- DataVisor's complete fraud modeling platform simplifies UML and allows users to build a UML model themselves.
- A manual mode allows clients to integrate data, engineer features, and build or fine-tune models on their own.
- DataVisor also offers templates that clients can follow for different scenarios.
- Models can be tweaked to fit a company's unique needs.
- Users can get up and running in as little as two weeks.

5 MISCONCEPTION #5
UML is similar to anomaly detection

Anomaly detection is one part of UML, but it's not the end-all detection tool. Anomalies can happen for a number of reasons, and often result in a high false positive rate. However, when the relationships between anomalies are viewed, organizations can lower their false-positive rate because they have more insight as to whether an anomaly is truly suspicious.

- Anomaly detection typically isn't effective in fraud detection because it creates a lot of false positives.
- Because of the high false-positive rate, anomaly detection often requires extensive manual intervention.
- Fraud rings synchronize their behaviors instead of creating one-off fraud transactions, which may not appear as an anomaly when looking at individual user activities. This is why it's important to go a few layers deeper and define the relationships between activities.
- UML isn't anomaly detection because it looks at clusters of activities rather than individual activities that don't fit any specific patterns.

Learn more about how DataVisor can help you fight fraud.

REQUEST A DEMO