Not All Machine Learning Systems Are Created Equal

Contents

Introduction

The rise of e-commerce, alternative payment methods, and card not present transactions has caused the number of fraud losses by retailers to increase greatly. According to Juniper Research, retailers will lose $130 billion globally in card-not-present fraud between 2018 and 2023.

Another factor attributing to the increase in fraud losses by retailers is the easy access to modern technologies. Fraudsters now have access to cloud services, web servers, APIs, and advanced technologies. This allows them to not only find new and innovative ways of committing fraud, but also to commit fraud on a massive scale and with super speed.

In this challenging and competitive landscape, online businesses need a fraud prevention solution that can detect many different types of fraud and is effective across multiple channels. The right fraud prevention system can help businesses reduce fraud losses, lost sales, fraud review costs, and chargebacks. The right system can also help prevent businesses from losing customers due to troll accounts, fake listings, and content spam.

The shortcomings of legacy fraud detection systems

Online businesses are discovering that traditional fraud prevention systems simply can't keep pace or deliver the results they need. Rules engines, the most common legacy approach to fighting fraud, are not as effective as machine learning-based systems at fighting the many different types of online fraud or keeping up with massive amounts of user-generated data.

Unlike machine learning-based systems, rules-based systems are not capable of automatically learning from large data sets or analyst feedback. Instead, rules are hard-coded and inflexible. That means that as fraudsters adapt their tactics, businesses are left vulnerable to new types of fraud attacks that can easily slip through the cracks. Rules-based systems are also not capable of detecting the subtle nuances of fraud, treating it in "black and white" terms that often cause a greater number of false positives.

Qualities of a powerful machine learning solution

There are quite a few machine learning-based fraud prevention solutions available today. But not all machine learning solutions are created equal. So, what makes a machine learning system effective at catching fraud?

Speed: works in real time.
Fraudsters constantly find new and innovative ways to commit fraudulent transactions, so an effective tool must be able to respond to changing fraud patterns as they occur.

Scale: draws upon vast and varied networks.
The key to doing machine learning well is to leverage large volumes of high-quality data. The more data the tool has access to, the more accurate it will become.

Sophistication: can accurately analyze and process enormous data sets.
Since suspicious behavior is often buried within streams of data, an effective tool must be able to extract individual elements, understand their significance, and deliver precise results.

While the types of machine learning algorithms a fraud prevention system uses are important, access to massive quantities of high-quality data is crucial when it comes to fraud prevention. Algorithms are only as good as the data and training provided to them.

The Sift solution

Sift uses large-scale machine learning and a global network of fraud data to provide online businesses real-time, adaptive fraud protection. Sift features a vast library of fraud detection and prevention models covering many different types of businesses and industries. Sift can also be customized to suit each company's specific fraud prevention needs. Sift can help any online business fight fraud, while still providing a great customer experience.

Sift learns from more than 5,000 fraud signals, which can be clustered into two basic categories: behavioral signals and identity signals. Behavioral signals include things like what a user clicks or taps on, the rate at which they buy things over a certain period of time, when they signed up for an account, and other actions they take on a website or mobile app. Identity signals include things like email addresses, device information, and billing and shipping addresses.

What makes Sift different

Here are just a few of the features that set Sift apart when it comes to detecting and preventing online fraud:

Online Learning

Sift uses online learning to update deployed models with new information obtained from customers, third parties, and other real-world sources of data. When the system is notified that a transaction has been deemed to be fraudulent, the system learns what characteristics and attributes the fraudulent transaction contains. That knowledge is relayed across the entire network in milliseconds, allowing customers to receive updates in real time.

Large-Scale Machine Learning

Large-scale machine learning allows Sift to leverage thousands of signals in order to quickly discover new fraud patterns and detect fraudulent behavior. Thanks to large-scale machine learning, Sift features a fraud library that contains millions of fraud patterns. Large-scale machine learning also allows the platform to uncover fraud signals specific to each client, automatically and with no additional integration work.

Rescoring

Sift features real-time learning and scoring so that new fraud patterns and signals can be learned, detected, and predicted, and knowledge is shared through the network in milliseconds. Sift constantly reevaluates fraud factors, signals, and other data — then rescores the probability that certain users are conducting fraudulent transactions.

Labels

A label is the representation of a human judgment about a specific user. Customers can apply a label indicating whether specific users are "bad" or "not bad." If a user is marked as "bad," the system considers all of the behavior and signals of the user to be associated with bad behavior in the future. Labels help Sift achieve a high level of accuracy when it comes to learning and predicting future fraudulent behavior.

> "Labels are extremely valuable. Customers benefit from a network of thousands of fraud analysts across the globe, spread throughout different geographies, and various verticals and businesses. The magic of Sift is we're able to represent and weight this aggregation of human judgment so that we're not overly biased one way or the other."

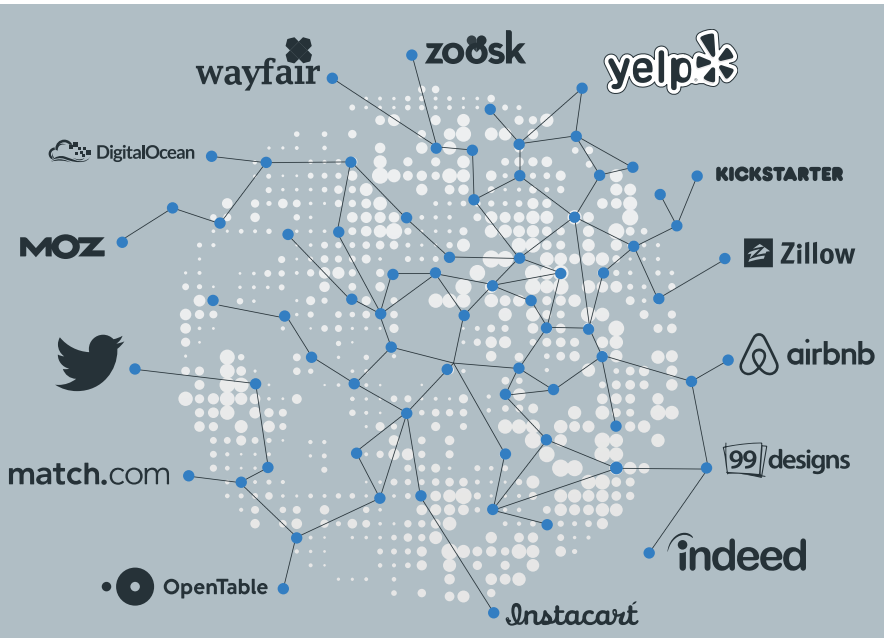
An Ensemble of Models and Global Models

Sift features an ensemble of predictive models that detect different types of fraud based on specific signals and behavior. The platform currently learns from over 5,000 fraud signals and that number is growing rapidly. Each customer receives their own machine learning model and a global model that shares the data collected from each transaction and model across the network. The platform also uses in-depth custom learning to create models that are defined by a small number of signals and can target business specific data for each customer.

Global models make it possible for customers who do not apply labels to their own data to leverage all of the other models shared across the network, providing them with effective fraud prevention.

> "One of the things that really separates Sift is our global model. As customers label data, and that data is shared across the global network, new customers can begin to derive value from the product straight out of the gate. That is very unique to our system."

Fred Sadaghiani
CTO at Sift



Console / Data Visualization

One of the most important features of Sift is a console that tells the complete story about the results through data visualizations, relevant signals, and access to raw data. Many fraud prevention platforms do not do a good enough job of explaining the results, nor do they provide a means for users to explore the data.

Users need to understand why transactions are scored the way they were and why certain transactions are deemed to be fraud. Users should be able to explore the data gaining insights from fraud signals and data visualizations.

N-Gram Analysis

When it comes to detecting fraud, Sift does more than just simple correlation. The platform uses n-gram analysis, a type of natural language processing that looks at all of the combinations of adjacent words or letters of length n. This allows for a detailed, nuanced representation of the data.

N-gram analysis is especially useful when it comes to spam detection and identifying multiple fake accounts. For example, when a fraudster is blocked, they will often create another account on the same site, and may change a few details (for example, by tweaking johndoe123@gmx.com to johndoe124@gmx.com). Sift is one of the few vendors employing n-gram analysis to identify repeat behavior like this, and can typically flag fraudulent users who come back to a website or app — even if they change their device or identifying information.

> "One challenge in the field of machine learning is explaining the results. How do you describe why the algorithm surfaced a particular example — a particular user, order, or transaction — as risky? We really take pride in how we're able to tell that story and make it easy for our customers to interpret our results."

Fred Sadaghiani
CTO at Sift

Conclusion

The quantity, velocity, and variety of transactions and fraud data continue to increase at an extremely rapid pace, and legacy fraud prevention systems simply can't keep up. Machine learning-based systems like Sift are designed for analyzing vast streams of data generated from billions of transactions in real time. They are also capable of detecting fraud from nuanced combinations of signals buried in data.

While machine learning is far more effective at fraud protection than traditional methods, it is not a silver bullet. Machine learning alone is not going to solve all of your fraud problems. Machine learning requires access to massive quantities of high-quality data in order for it to be truly effective at detecting and preventing fraud.

Sift features a growing network of thousands of companies that provide a wealth of quality data and models helping the platform accurately predict and prevent fraudulent transactions in real time. Sift can help protect your online business from the rapidly evolving strategies and methods of fraudsters. Learn more about how Sift can help your online business fight payment fraud, account abuse, and many other types of fraud at www.sift.com.

 |