sift

Not All Machine Learning Solutions Are Created Equal

The unmatched value of real-time fraud prevention and global data

Contents

Facing the Scale of Fraud

From now until 2023, retailers around the world are careening towards more than $130B in losses as a result of card-not-present fraud. The swift and exponential expansion of e-commerce businesses and alternative payment methods has undoubtedly brought major convenience to consumers, not to mention a massive influx of revenue to the merchants they flock to. But with such speed, growth, and scale comes an equally large rise in vulnerability—a fact that cyber criminals are all too aware of, and all too ready to exploit.

Business leaders are being forced to face a harsh and unfortunate reality: whatever technology merchants have access to is just as easily available to fraudsters. From cloud services and web servers to APIs and other advanced solutions, the innovations that move our world forward can be flipped on a dime to create new, highly-scalable ways to commit fraud with unprecedented speed. And in markets that evolve constantly to meet the changing needs of consumers, online businesses need a fraud prevention solution that can effectively detect multiple types of fraud across a variety of channels—and enable trust and safety teams to scale operations beyond what they can do on their own.

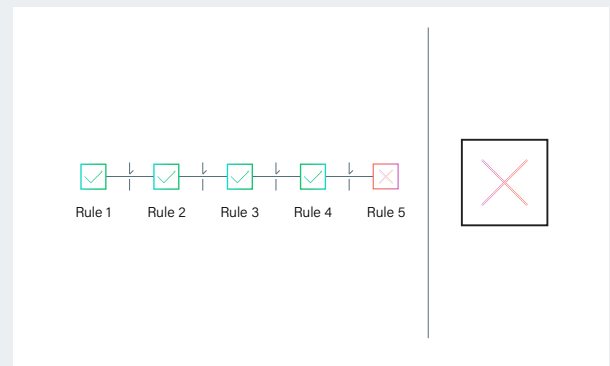# "Checkbox" Machine Learning: When Fraud Solutions Fall Short

Online businesses are beginning to feel the stark limitations of traditional fraud prevention systems built on static rules, or platforms using what's known as "checkbox" machine learning. A single algorithm, a limited data set, and models that only update every couple of weeks in order to adjust rules—all of that still technically counts as machine learning (ML). But are these rules-in-ML-clothing actually complete, real-time, end-to-end solutions that significantly reduce manual review while driving exponential growth? No.

These legacy solutions simply can't keep up with the evolving digital marketplace, because they were developed with the assumption that manual review would always be possible, that user data would always be manageable instead of massive, and under the pretense that the average customer base wouldn't expand as dramatically or as quickly as it does today.

Unlike systems built with complete machine learning models, these rules-driven "checkbox" platforms are not capable of automatically learning from large data sets or manual analyst feedback. Instead, they act like a gate, subjecting every user to the same level of scrutiny based on a narrow range of criteria that's collected during a finite period of time. They're always behind on data, which means their users are never going to catch up with fraud. Their inflexibility makes them frequently inaccurate, and fraudsters have very little trouble adapting their attack strategies to outmaneuver them. Finally, the unconditional friction that these systems rely on causes trusted users to regularly get caught in the net—leading them to abandon your site for competitors that offer a better experience, and driving customer insult rates through the roof.

Fighting Modern Fraud: Rules or Machine Learning?

While rules-based systems allow for some discrimination between legitimate and fraudulent interactions, solutions that use machine learning offer unmatched speed and accuracy.

Rules 1-5

Rules-based systems and "checkbox" machine learning solutions use limited criteria and time frames to treat every user with the same level of scrutiny. In turn, fraudsters can figure out how to outsmart them, and customer insult rates go up. As shown above, even if a trusted user passes most rules (green), it only takes one failed rule to block their journey on your site.
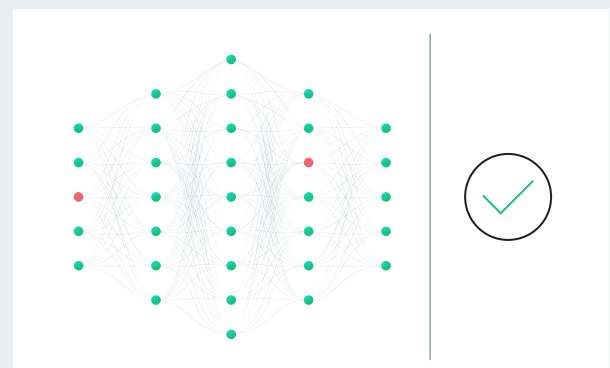
End-to-end machine learning (ML) assesses interactions using numerous diverse signals, surfacing patterns that enable holistic risk assessment and giving you the speed and accuracy that only machine learning can provide. Above, green dots represent low-risk signals, while red dots represent higher risk signals.

With this in mind, it's important to know that the fraud prevention solutions available today don't necessarily fall squarely into either checkbox or end-to-end machine learning. Many platforms claim to use various levels of machine learning or artificial intelligence coupled with rules, so it's critical that you understand what technology is truly at the core of the product—as well as how crucial rules really are to the overall functionality of the solution.

To determine what you're looking at, it's helpful to understand what you should be looking *for*. You'll want to explore the way the technology works, the platform's data and security compliance, how easily and fully it will integrate with your existing technology stack, and what support you can expect during onboarding and beyond.

Questions to Ask

- How robust is the platform's data network compared to what we currently have access to?

- Can we customize the machine learning model for our specific business needs without sacrificing growth or negatively impacting the customer experience?

- Can the platform expand a single data point into multiple fraud signals for better accuracy, or will we have to manually build it out on our own?

- How quickly can the model adapt to, and catch, new fraud patterns?

- How difficult will it be to integrate into our current technology stack, and what resources will be required?

- Does the platform comply with necessary federal and state security requirements, as well as our organization's internal security needs?

- How comprehensive are the reporting features compared to what we can currently do?

- What kind of support and resources can I expect from the solution provider?

Key Capabilities of a World-Class Machine Learning Platform

There are two primary things to look for when researching machine learning-based fraud prevention platforms: accuracy and transparency. Above all, the solution must be able to differentiate fraud from trusted actions. It should also be simple to understand the predictions and red flags that emerge. The patterns and recommendations the solution surfaces should be accurate, accessible, and actionable. It should be clear how and why the model has come to various conclusions, as well as how those findings influence refinements to its structure. This level of visibility will enable you to make smarter decisions, and confidently make the right moves to drive your business forward.

That said, if all machine learning solutions aren't created equal, what makes one truly effective at catching fraud? Scale, sophistication, and finally, speed.

Scale: It's built upon vast and varied networks of data.

The ability to leverage large volumes of high-quality data is required for a machine learning solution to be as accurate as online businesses need it to be. Simply put, the more data a model has access to, the more precise it can become—ultimately leading to exponentially greater growth and significantly less fraud.

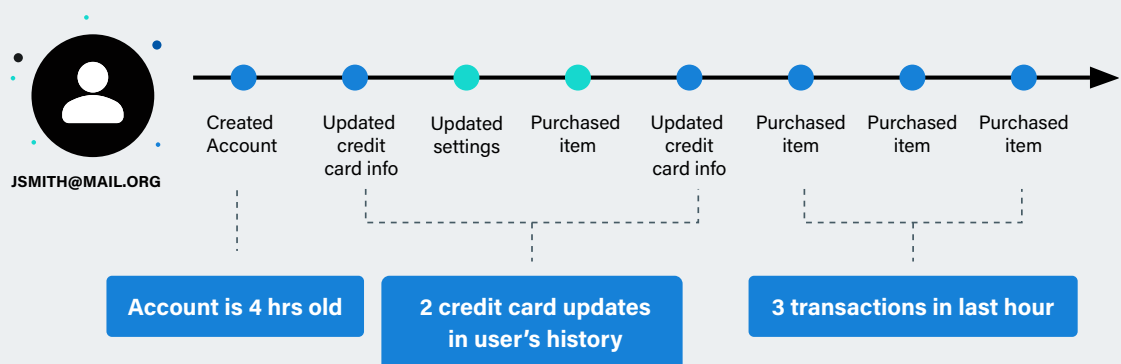## Sophistication: It can deftly analyze and process enormous data sets.

Risk signals are often buried within massive streams of data. An effective fraud prevention solution must be able to gather a multitude of signals from a single data point, understand their significance, deliver insights about them, and relay this information to an ensemble of global and custom models for increased accuracy. For example, an email address can be a single data point (e.g., "Have we seen this email address before?"), or it can be a dozen, including analysis of the domain and the username itself. A large data set and sophisticated algorithm are worth virtually nothing if an ML system isn't able to extract the necessary data points from them to help your business stop fraud.

Speed: It surfaces patterns and adjusts in real time.

Fraud is adversarial and constantly changing. The criminals who commit it are always looking for new and innovative ways to circumvent prevention strategies, so an effective platform must be able to identify and respond to changing fraud patterns as they happen. To accurately fight fraud without sacrificing customer experience or business growth, scalability and sophistication need to happen at great speed—something that's incredibly difficult to do, but that can be accomplished with a world-class machine learning fraud prevention solution in place.

Online learning in action

JSMITH@MAIL.ORG

Created Account — Updated credit card info — Updated settings — Purchased item — Updated credit card info — Purchased item — Purchased item — Purchased item

Account is 4 hrs old

2 credit card updates in user's history

3 transactions in last hour

The Sift Solution: Unstoppable Growth, Powerful Protection, and Global Expertise

Sift uses large-scale, real-time machine learning and an ever-expanding global network of data to adapt our models to emerging trends and share key insights across markets. Sift develops custom learning models to meet our customers' specific needs and provide online businesses with unparalleled accuracy for proactive, adaptive protection against fraud.

Here are a handful of the capabilities that set Sift apart when it comes to detecting and preventing online fraud with machine learning:

Online Learning

"Checkbox" fraud prevention solutions utilize bare-minimum machine learning capabilities. Their models only update every couple of weeks, leaving the teams that use them trailing far behind fraud trends, and stuck in the dark about how fraud is truly impacting their business. Sift learns online, powering real-time risk assessment at every interaction (e.g., account creation, transaction, login, content creation, etc.), allowing our global data network to update constantly and refine deployed models based on new information obtained from customers, third parties, and other real-world sources. When the system is notified that a transaction has been flagged as fraudulent or risky, it learns what characteristics and attributes the fraudulent transaction contains, helping the entire network to get smarter and more accurate over 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. Our customers benefit from a constantly-evolving library of information that includes millions of fraud patterns surfaced from merchants in nearly every vertical around the globe. Finally, large-scale ML allows the Sift platform to uncover fraud signals specific to each of our clients—automatically and with no additional integration work.

> "
> One of the things that puts Sift unequivocally ahead of the pack is our global model. As customers make decisions about data, and the model informs the global network of those findings, new customers can begin to derive value from Sift immediately. That is very unique to our platform.
>
> **Geoff Huang**
> VP of Product at Sift

Sift's real-time learning versus checkbox machine learning
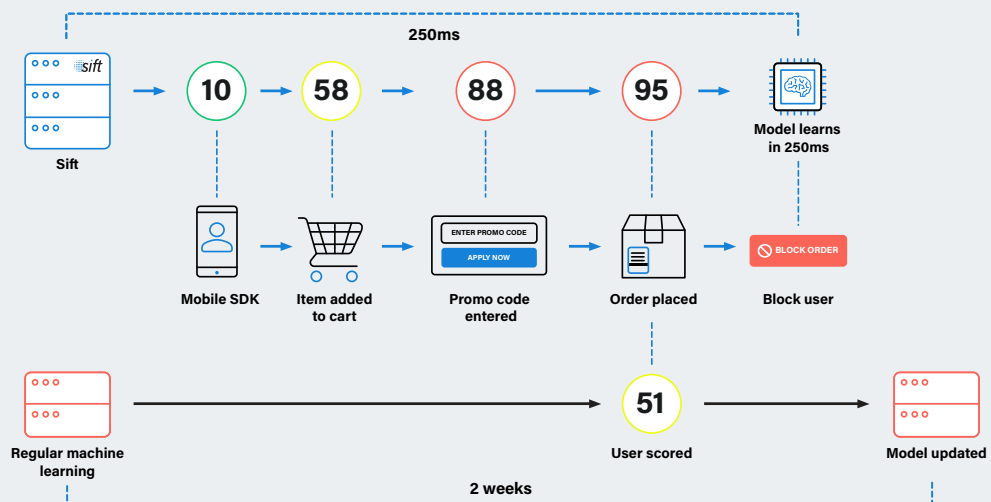
Rescoring

Sift learns online, rescoring in real time and leveraging robust feature engineering capabilities to transform raw data into actionable insights that better represent the underlying issues being surfaced. This allows Sift customers to fully understand how fraud is affecting their businesses and stay well ahead of emerging fraud signals, all while detecting and predicting attacks with pinpoint accuracy.

> " 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 that we're able to represent and weigh this aggregation of human judgment so that the platform assesses risk while reducing unnecessary bias.
>
> **Kevin Lee**
> Trust & Safety Architect at Sift

An Ensemble of Custom and Global Models

Sift features an ensemble of predictive models that detect different types of fraud based on specific signals and behaviors. The platform currently learns from tens of thousands of signals that span transactional, behavioral, identity, and relational attributes. Each customer receives their own custom machine learning model, as well as a global model that leverages data collected from each transaction and model across Sift's customers, and shares it across the network without ever identifying individual customers or end users. Sift 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 company, such as which travel routes are riskiest or which email domains are most often associated with risk. Our global models make it possible to leverage learnings shared across the network, providing every Sift customer with the most comprehensive, accurate, and effective fraud prevention available on the market.
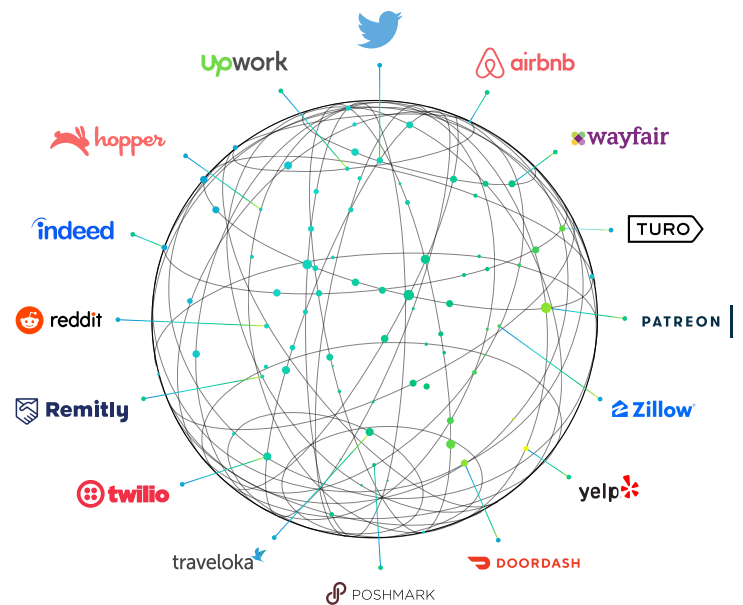
The Sift Console and Data Visualization

Our primary goal is to provide the most accurate risk assessments possible to our customers. But one of the most important features of Sift is a Console that helps you understand the insights our machine learning models surface, rather than hiding findings behind a single number or prediction. For example, a "checkbox" solution might give you a risk score without any insight into how that score was determined, or identify an action as fraudulent without detailing why. The Sift Console tells the story behind the data, empowering our customers to take action by automating more decisions, investigating and exploring cases, and visualizing fraud data so they can make smarter choices about how to refine their fraud mitigation strategies.
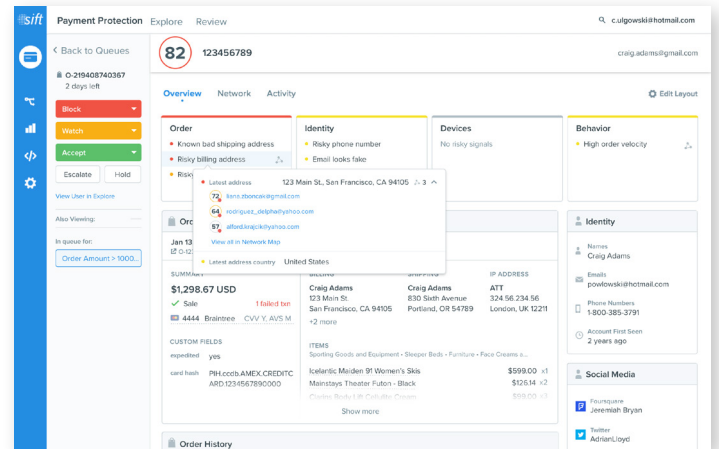
> "
> One challenge in the field of machine learning is explaining the results. How do you describe why the algorithm surfaced a signal as risky? We really take pride in how we're able to tell that story and make it easy for our customers to interpret Sift's results.
>
> **Kevin Lee**
> Trust & Safety Architect at Sift

Dynamic Friction

Leading businesses from a variety of diverse markets are adopting Sift's Digital Trust & Safety mindset, which also enables Dynamic Friction. This proactive, real-time approach gives merchants the ability to apply friction (e.g., multi-factor authentication) when it's needed, and remove it when it's not. Using patented, powerful machine learning, merchants can provide tailored account creation, login, and purchasing experiences for each user based on their risk level.

Looking Ahead: Fighting Fraud with Speed, Science, and Sift

The volume, velocity, and variety of transactions and fraud data impacting e-commerce businesses will continue to increase at an extremely rapid pace. Legacy fraud prevention systems built on checkbox machine learning are already incapable of keeping up, leaving the companies that use them vulnerable, and putting them at a major disadvantage against competitors. It takes sophisticated machine learning to accurately analyze the vast streams of data generated from billions of transactions in real time, and give companies the tools they need to stand strong against digital fraud—without compromising what's good for business.
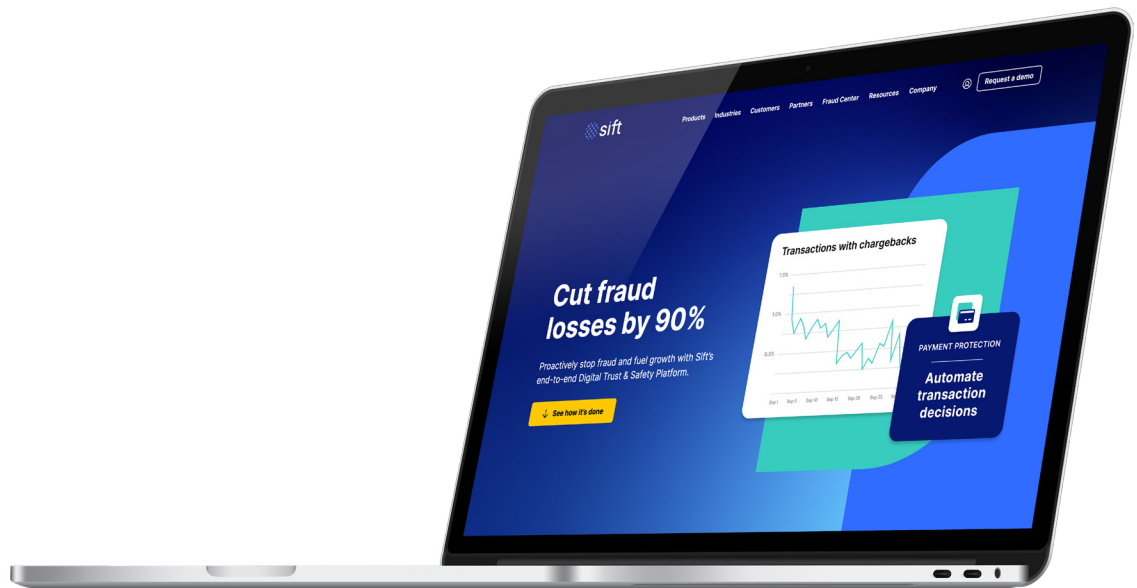
End-to-end intelligent automation with Sift

Sift is the leader in Digital Trust & Safety, empowering digital disruptors to Fortune 500 companies to unlock new revenue without risk. Sift dynamically prevents fraud and abuse through industry-leading technology and expertise, an unrivaled global data network of one trillion (1T) events per year, and a commitment to long-term customer partnerships. Global brands such as DoorDash, Poshmark, and Twitter rely on Sift to gain a competitive advantage in their markets.

Visit us at sift.com and follow us on LinkedIn.