How Sift Works

Contents

Introduction: the Sift technology stack

In 2011, Sift disrupted the fraud prevention industry with a first-of-its-kind machine learning approach that accurately predicted fraud and defended against online abuse in real time.

Today, our global data network, custom machine learning models, automation technologies, and comprehensive reporting fuel the business growth of tens of thousands of websites, allowing them to prevent fraud, streamline operations, and drive revenue growth.

From a technology perspective, detecting fraud is extremely difficult – like finding a needle in a haystack. Additionally, fraudsters are constantly evolving and adapting their techniques. So how do we do what we do effectively? We've designed our technology stack to meet these challenges head-on.

Sift receives billions of events per month and analyzes vast streams of data in real time. Our platform hosts a powerful machine learning engine that allows us to detect and prevent fraud by analyzing nuanced combinations of signals buried in historic and real-time data.

Our approach combines speed, scale, and sophistication to deliver a unique, adaptive solution that allows our customers to accurately distinguish between the users they can trust and those they can't. Savvy businesses use this knowledge to focus on improving the experiences for trusted users, while keeping fraudsters at bay.

Let's take a look into the powerful technology at the heart of the Sift engine.

Enterprise-grade data infrastructure

Sift offers a secure, reliable, and scalable infrastructure that opens up multiple integration points to ensure the capture of critical data from any source that you use.

Our solutions specialists will ensure that you make the most of Sift. As trusted partners, they will guide you as you integrate your desktop and mobile experience using our Javascript snippet and SDKs and connect your backend systems using our REST APIs. Our easy to use integration guides and support center articles ensure that you can also find your own answers along the way. Many organizations are able to get started with Sift with just a few hours of engineering time.

Data infrastructure drivers

Sift has three data infrastructure drivers: reliability, scalability, and security. Let's take a close look at each.

Reliability

Sift supports a fast-growing portfolio of enterprise-level and international clients. We strive to provide a resilient and highly available service to our customers across the globe. Attributes of our service that contribute to our high availability include 24x7 monitoring, backups, and incident response procedures. You can always check our system status by visiting our public status page. Hosted in multiple data centers and PoPs around the world, the Sift data infrastructure ensures topline performance while eliminating a single point of failure.

Scalability

Your growth is our primary success measure. We recognize that the ability to scale with our users' demand is critical. At Sift, our cloud infrastructure is designed to dynamically autoscale as traffic to your business peaks, ensuring unlimited processing power to meet your needs. Highly tailored individual components of our infrastructure – such as a customized HBase and Kafka implementation – ensure our ability to process over 1 billion database requests and handle many thousands of API traffic requests per second. We perform 100,000+ queries per second, all while storing petabytes of data in a read-heavy workload.

Security

We're serious about protecting data. Sift maintains compliance with the SOC 2 framework. We employ strict access control, two-factor authentication, and encryption of data in-flight and at-rest to ensure that the highest standards of customer privacy are maintained at all times. Our annual SOC 2 Type 2 assessment tests our security processes and controls against the SOC 2 security framework, ensuring independent, third-party assurance that we are taking the appropriate steps to protect our systems and our customers' data.

Massive quantities of high-quality data

Machine learning requires access to massive quantities of high-quality data in order to be truly effective at detecting and preventing fraud. The quality of the results you see are directly in line with the quality of data sent to Sift.

Our models learn from real-time data received from 34,000+ domains across the world fighting fraud with Sift. Having pioneered the machine learning approach in the fraud prevention market, we've been collecting this invaluable data for more than seven years and now host the richest set of labeled data in the world, resulting in the highest possible accuracy of our machine learning models.

We are constantly pushing the envelope to discover hidden fraud patterns before they can damage your business. To accurately predict fraudulent behavior in real time, we sift through a variety of data types and formats that together create a holistic picture of emerging fraud threats.

Raw data

In the world of machine learning, the most accurate analyses come from high-quality data. We centralize an array of datasets gathered via SDK, Javascript snippet, and API – augmented with a variety of third-party data – so that our customers get the most comprehensive view of the world. By integrating with Sift, the burden of data collection and management is taken off of your plate. The data that we leverage can be broken down into the following categories:

USER IDENTITY

Attributes that are associated with the identity of a user
Examples: Name, email address, phone number

PROCESSES

Preferences and patterns associated with the user
Examples: Browsing patterns, keyboard preferences, screen tilt

LOCATIONAL DATA

Location attributes associated with the user
Examples: fine and coarse location, GPS coordinates, shipping address, billing address

DEVICE & NETWORK DATA

Properties of the device and network connection associated with the device
Examples: IP Information, Network ID, carrier network, device manufacturer and model

TRANSACTIONAL DATA

Order details and order history associated with the user
Examples: Order value, order velocity, payment instruments

DECISIONS

Business actions that your team takes every day

CUSTOM DATA

Attributes that are unique to your business
Example: For a hotel reservation, the number of nights associated with a booking

THIRD-PARTY DATA

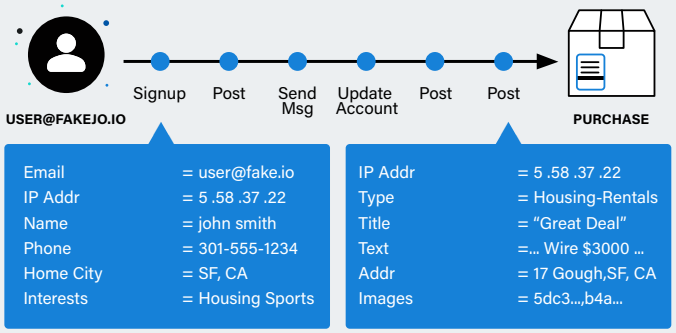
A variety of relevant third-party datasets
Example: geo data, bank data, currency rates and conversions, social data

Knowledge derived from data

Organizing data for the prediction tasks that we deal with helps to establish connections between users, devices, locations, and other attributes. When we look at data, we don't just review the last event, but rather analyze the entirety of behavior and actions over a span of time. This broad-scale perspective allows us to learn new user behavior relative to other users for each customer, and scale those learnings across the entire network. This time series nature of modeling affords us a great deal of flexibility and utility when fed into the machine learning system. Some of the knowledge examples that provide valuable insights into the behaviors of fraudsters include:

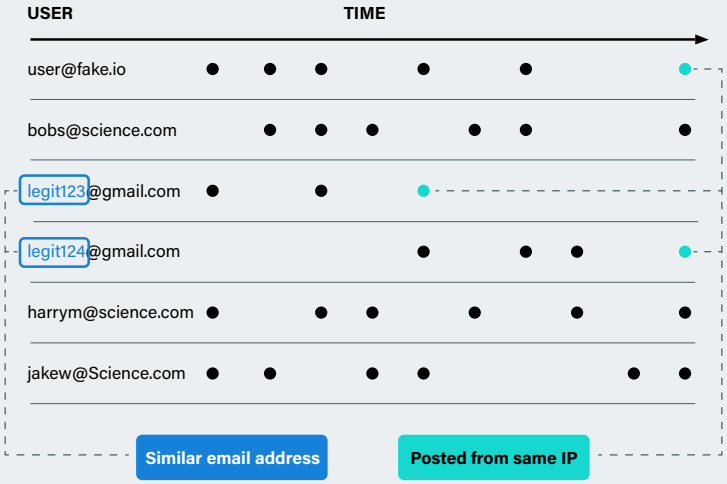
TIME SERIES DATA

As users interact with your website, every single step of that journey is collected and analyzed to reveal insights into your users' traits. This knowledge immediately updates models across the network.



CROSS-USER DATA

Data points across multiple users can be utilized to see data patterns that reveal connections between users and logins.



CROSS-ORGANIZATIONAL DATA

Manually reviewed transactions across organizations get looped back into the system as soon as they are marked by fraud analysts. This becomes a valuable data point that influences risk analysis across the network.

Sophisticated data preparation

Suspicious behavior is often buried within streams of data. Sift has built a complex system that combs through the world's largest database of fraud- and abuse-related data collected from our global network in real time, and maps it against meaning, relationship structure, and relevancy – often in company-specific and market-specific ways. This evidence informs our machine learning models that recognize new patterns and update in real time. Sift pioneered this approach and has proven its success time and again with businesses of all sizes and industries across the world.

Data normalization

Fraudsters are constantly hunting for new methods to get around existing system controls and rules. That means that as fraudsters adapt their tactics, businesses can be vulnerable to new types of fraud attacks. Starting with data normalization, Sift spots the little details that other approaches miss. Here are two examples of data normalization techniques that we frequently use:

Reliability

As an example, a customer might use a rule or a blacklist to block a fraudulent email address, e.g *johndoe123@gmail.com*. In response, a fraudster will often create a similar-looking email address, e.g *johndoe124@gmail.com*, to circumvent the controls enforced by your system. A similar technique is common with physical addresses:

Ralph Wiggum 123, Smith Ln San Francisco, CA	Ralph W Smith Lane, #123 San Francisco, California	Ralph C Wiggum #123, Smith Ln San Francisco, CA

Sift has expertise in identifying this repeat behavior. Our data normalization coupled with our n-gram analysis extracts the key substrings in the data field to identify repeatable data patterns and then apply probabilistic models to weigh the likelihood that two data inputs are correlated with each other.

RAW DATA	NORMALIZED DATA AFTER STRIPPING OUT SPECIAL CHARS, NUMBERS ETC.
johndoe123@gmail.com johndoe124@gmail.com johndoe_123@gmail.com	John doe @ gmail.com

Example data transformation

Therefore, when we spot minor variations to a known fraudulent email address, we are able to accurately match and flag similar email addresses.

 |

Currency conversion

Currency conversion is another critical data point. If a buyer on your site spends money in a currency different from the majority of your users, we will properly adjust the order amount on our end before comparing it with other orders. Since currency conversion rates fluctuate continually, we base this comparison on conversion rates around the time of the sale, not on a single fixed currency conversion chart.

Feature engineering

A key challenge in building an effective machine learning system that accurately detects a variety of fraud and abuse vectors is feature extraction – deriving the most useful signals from all kinds of raw data sent to Sift. Feature engineering transforms raw data into structured, machine-processable formats that can be understood by a machine learning algorithm. Why is this important? It allows us to set up building blocks that are powerful indicators of fraud. For example, a count of the number of vowels per email address when applied to a machine learning model could be used as a strong fraud signal.

Fraud isn't static, and new patterns emerge daily. Building an effective set of features that will uncover fraud indicators capable of detecting and blocking tricky behavior requires deep knowledge of the industry, our customers, their end users, and fraudsters. With feature engineering, experience is everything. Sift has a library of over 10,000 features that we use to uncover fraud patterns across many industries and time zones. Analysis of false positives and false negatives identified by our customers further contributes to and greatly improves our detecting capabilities, and those findings are used by machine learning models across the entire network.

Large-scale learning

As customers send us data, we automatically extract and build features over that data without having to know what it is. We are able to leverage any discrete, granular piece of information within the context of everything else we've learned. For instance, if a customer sends us { "pickup_lat": -39.234234, "pickup_lng": 120.234 }, we'll automatically learn that this is a geo location and compute a battery of feature extractors that look at (among other things) the distance, the IP and location of the user, and more. These findings will impact the user's risk score, even if you originally suggested that distance may not be an important factor.

Types of features

While the feature set constantly evolves based on fraud patterns and available customer data, features can be broadly classified into the following categories:

EVENT FEATURES

Properties of a user's most recent event
Examples: content posted, credit card type, billing ZIP Code, shipping type, login device, etc.

STATE FEATURES

Properties of user's current state
Examples: country, time of day, browser type, IP address, login device type, etc.

TEMPORAL FEATURES

Properties of the user's time series up to that point
Examples: the number of transactions in the past hour, number of other user logins associated with the same IP address, etc.

GRAPH FEATURES

How the user relates to others on the site and on other sites
Example: number of times an IP address has been labeled as bad across the Sift network.

IDENTITY FEATURES

A feature such as a fingerprint from your browser or mobile device, usually in a form of an opaque identifier, that tells us who you are

Continued on the next page

BEHAVIORAL FEATURES

A different kind of fingerprint illustrating a type of behavior; one of our strongest classes of features, essential to identifying fraudulent patterns.
Example: user behavior such as type of purchase, time of day, and location.

VELOCITY FEATURES

A category of features that measures the rate of any action happening; Sift computes the velocity of everything that comes into our system
Example: velocity of failed transactions, or a rate at which you are logging in from a particular country per day, per week, per month, etc.

Feature densification

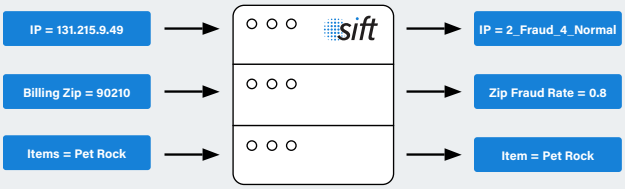
Once the features are derived from a time series of activities, we end up with different classes of features, such as:

- **device ID features**
- **features that track email address characteristics**
- **features that capture physical address characteristics**
- **customer field derived features**

For some customers, we have tens of thousands of features. Many of these features will likely have many distinct values and very little repetition for each value. In other words, the feature is very sparse, which makes it challenging to derive a signal. We define sparse features as any that can take on several hundred unique values (e.g. an email address, IP address, or ZIP Code). To solve this problem, we break sparse features down into smaller features called indicator features. This results in feature vectors with several thousand features. We then transform this feature data using a process called densification, where sparse features are represented as numeric fraud rates.

Example

Here is an example that shows densification in action. Let's say a feature vector comes in with a value of "94105" for the feature "shipping ZIP Code". The first thing we do is look up the number of fraudsters and normal users that we have previously seen with this shipping ZIP Code. If we have previously seen this shipping ZIP Code in either capacity, we can then produce a fraud rate feature that equals the ratio of fraudsters to all users with the same shipping ZIP Code.



Value of a feature

Each feature greatly impacts the model. How do we measure the impact of a feature on a customer?

Model evaluation

We've built many different tools to help us experiment, evaluate, and iterate across various types of models, customers, and data sets. An offline training pipeline bakes in all of the lessons that we've learned from conducting experiments. Today, this pipeline is how we produce valid experiment results at Sift. This investment is at the core of our ability to innovate and constantly deliver meaningful improvements to our customers.

Continued on the next page

Model evaluation

See these posts on our Engineering Blog for more information on ML experiments at Sift:

Part 1: Minimizing Bias
Part 2: Analyzing Thousands of Models
Part 3: Building the Right Tools

An ensemble of machine learning models

Before we dive into our models, let's agree on the high-level concepts behind supervised machine learning, specifically how it applies to Sift. Typically, supervised machine learning centers on a cycle of training, predicting, and acting stages. During training, we use historical data — including customer feedback — to find correlations between inputs and outputs. Inputs encompass both user-generated events and the metadata associated with them; outputs refer to scores between 0 and 100 that reflect the probability that a given user is a fraudster.

At Sift, instead of relying on a single machine learning model, we use an ensemble of several predictive models. Some models are trained with a general understanding of fraud patterns across our network of customers, some are built for the industry you operate in, and others are tuned to your organization's data specifically. This ensemble of models allows Sift to accurately score a transaction, a user, or a session while taking a holistic approach when analyzing risk.

Custom models

Each business is unique. We make every effort to customize our approach to catch the fraud that is specific to your needs. Say you're a shoe company – we might recognize that buying size 15 shoes at a particular time of day is associated with fraudulent activity. Sift's flexible machine learning models can detect such patterns of normal and abnormal behaviour that are very unique to your business model, industry, and audience.

Network models

Network learning models are one of our flagship features. They incorporate data from across the spectrum of Sift's vast customer base. The network learning models are important because fraudulent users could have accounts on multiple websites, and spotting bad behavior on one site helps to identify it on other sites as well. For example, Sift has learned from millions of examples which mailing addresses look like "reshippers" – intermediaries who ship goods on behalf of credit card thieves around the world. Other examples of data shared across the network are email addresses, IP addresses, and device fingerprints. Subtle patterns reflected in these data points and known fraudulent addresses are captured in our network-specific learning models and shared across our customer base. The result: when we learn about a fraudster, we share the learnings network-wide, before that fraudster can move on and affect another business.

Sift regularly measures the efficacy of network learning. We use a notion of Lift (formally: reduction in variance of error) that currently sits at 30%. In other words, a given customer would be 30% less effective at predicting fraud if they didn't benefit from the entire Sift network. This shared knowledge equates to more power for everyone!

Vertical models

Vertical-specific models bundle together businesses that are in the same vertical and tune scores based on learnings specific to organizations of that type.

Regional models

Over 40% of Sift's customers reside internationally. Region-specific models oversee businesses in various geographical segments, since the nature of fraud varies across the world. We closely monitor global fraudulent activity patterns to enhance regional models and increase the accuracy of our predictions.

A diverse model stack

Sift operates across multiple vectors of fraud and abuse: payment fraud, account takeover, fake accounts, fraudulent content, and promotion abuse. Each of these vectors demand a unique approach to identifying fraudulent behavior. Therefore, we have built a separate modeling stack for each, applying a battery of different machine learning algorithms that together deliver a so-called meta-model. Practically, for each customer, the meta-model represents a dynamic combination of the weighted outputs of the following machine learning models: logistic regression, random decision forests, deep learning (RNNs), N-gram, and Naive Bayes.

There's no world where one single algorithm can fix everything. Combining all of these models together far outperforms any individual model. What you have to do – and what you have to be really good at – is constantly deploying, scaling, operating, evaluating, and making improvements to algorithms across a broad cross-section of different verticals, geographies, and businesses with very different risk profiles and users. This comprehensive approach allows us to be adaptable to the needs of a very heterogeneous portfolio of customers. And that is a part of Sift's secret sauce.

Logistic regression

Logistic regression is particularly useful in cases where only a limited set of information is available for risk analysis – such as a case with sparse features (e.g. a guest check-out experience while shopping). Similar to Naive Bayes, logistic regression models provide easily interpretable results so that customers can accept Sift Scores with confidence and review the analysis that contributed to that score.

Random decision forests

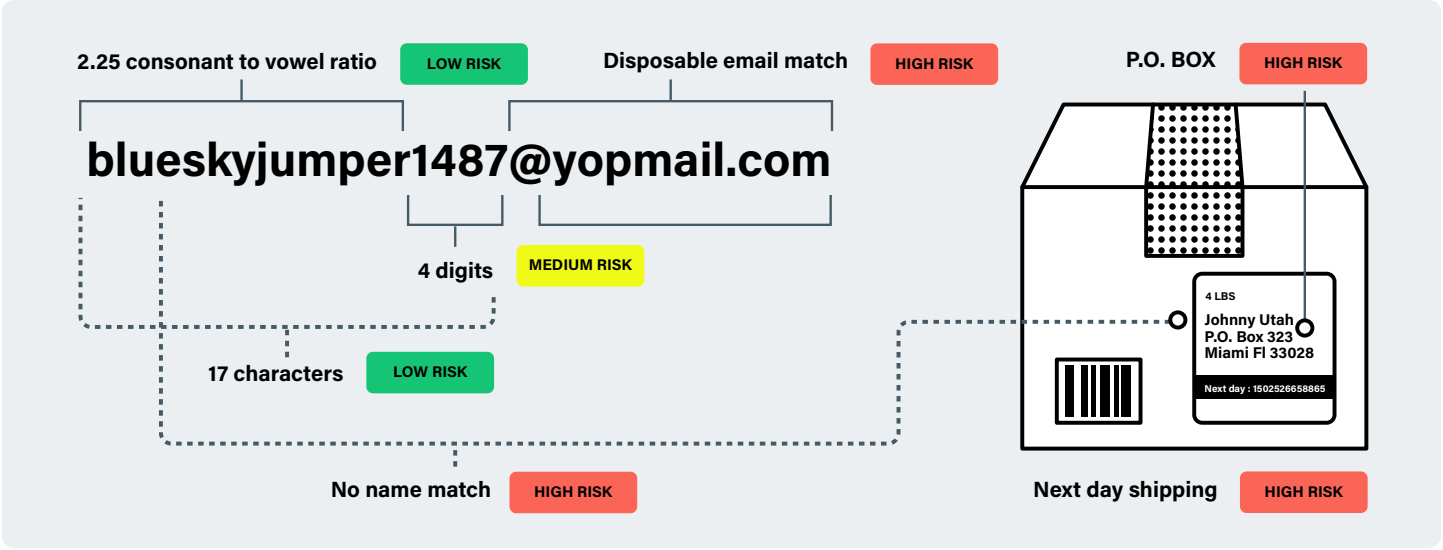
As a big data company, we leverage large-scale computation on deep datasets to draw the most accurate results in the industry. Random decision forests are a powerful, scalable, and intuitive model for us. They can model interactions among features, are relatively inexpensive to train, and are one of the most interpretable machine learning models around. Beyond that, random decision forests are highly accurate on our datasets, remove bias, and are widely used in large-scale applications. Another benefit is the ability to easily incorporate domain knowledge and the fact that a decision forest can be thought of as a large system of automatically learned rules. This last point is helpful when explaining machine learning to new customers who have only used antiquated rule-based fraud detection systems in the past.

RNNs

RNNs (Recurrent Neural Networks) are deep learning algorithms designed to learn on sequences of events in the Sift system. An event is usually a discrete action that someone has done – for instance, a transaction performed, a piece of content published, or a login into a user account. RNNs look at the time series of the events for all your users and then use that information to create a model that predicts what sequences of events are correlated with good or bad behavior. We do not write any features of our own here, but rather allow RNN models to discover new ways of identifying the signal through the noise, thus saving hours of manual construction from relevant factors. Instead, the network discovers them. RNNs proved to be extremely effective for us at predicting certain types of fraud.

N-gram analysis

N-gram is a type of natural language processing that looks at all of the combinations of adjacent words or letters of length n. This approach 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 such repeat behavior, and can typically preemptively flag fraudulent users who come back to a website or app – even if they change their device or identifying information.



Naive Bayes

Model accuracy and performance are important qualities that determine which models are included in a customer's modeling stack. Although Naive Bayes is a trivial model, we find that it contributes to producing the most accurate results in a large stack of algorithms, particularly while onboarding new customers with limited training data or providing reasons why and how Sift arrived at a particular risk score for a transaction or user.

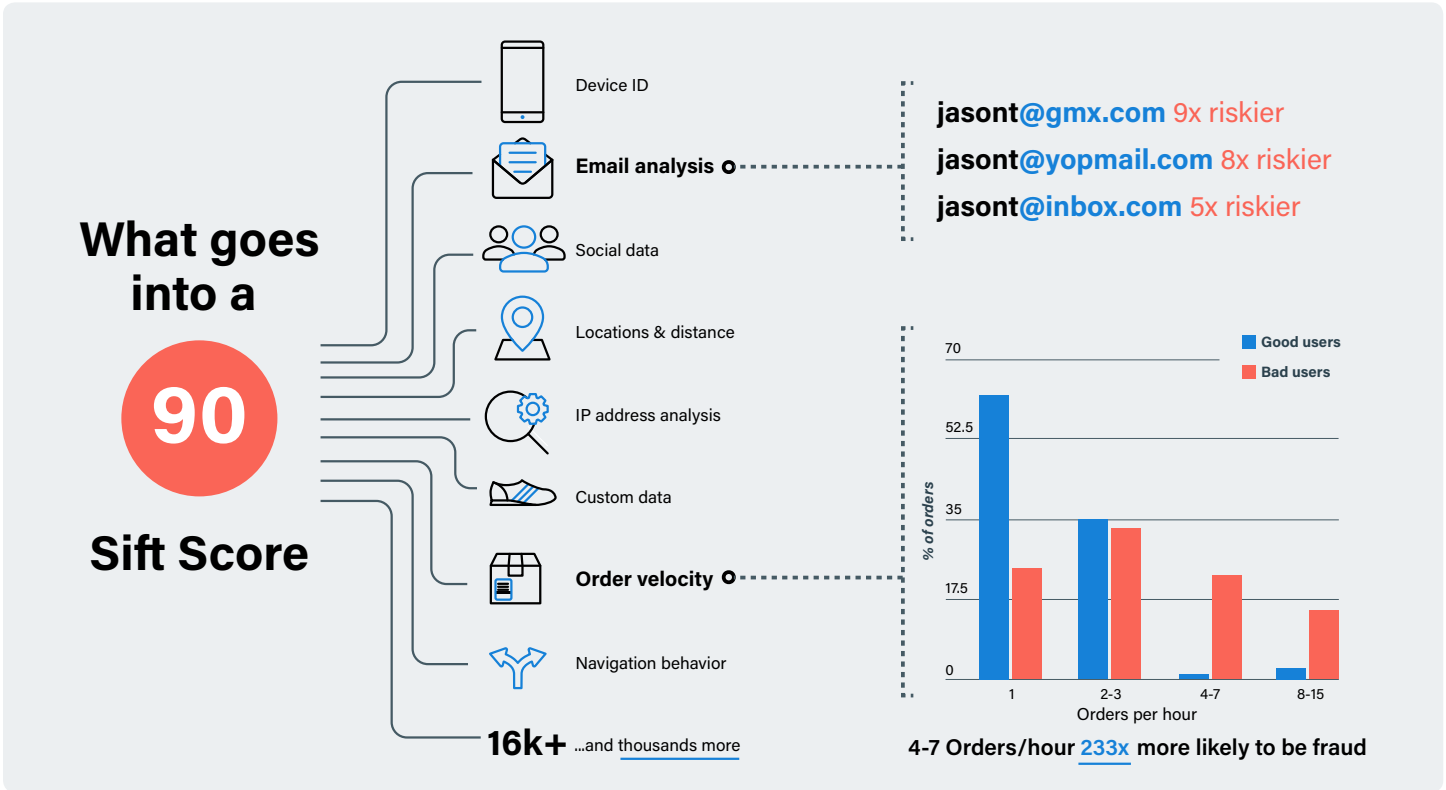
Striving for accuracy

Our fraud models improve as we get more data from our customers, and as we add more features to the models. We are constantly reevaluating the quality of our fraud models – not just how well they're doing overall, but also how well they're doing for each customer individually. When we run these evaluations, we are careful to look at the metrics that matter most to our customers – not just precision and recall (precision is a measure of result relevancy, while recall is a measure of how many truly relevant results are returned), but also holistic measures of classifier accuracy that are aligned with the efficiency of your review team. We monitor our evaluation metrics both "offline" (how well models trained on old data predict fraud from more recent data) and "online" (the bottom-line performance of the models in production). We're careful to avoid common pitfalls characteristic of simpler methods to evaluate time series prediction tasks.

Applying our ML to assess risk: Sift Score

A Sift Score is a number between 0 and 100 that reflects the probability of that user or transaction to be potentially fraudulent. At the bottom end, 0 indicates a low likelihood, while 100 indicates a high likelihood.

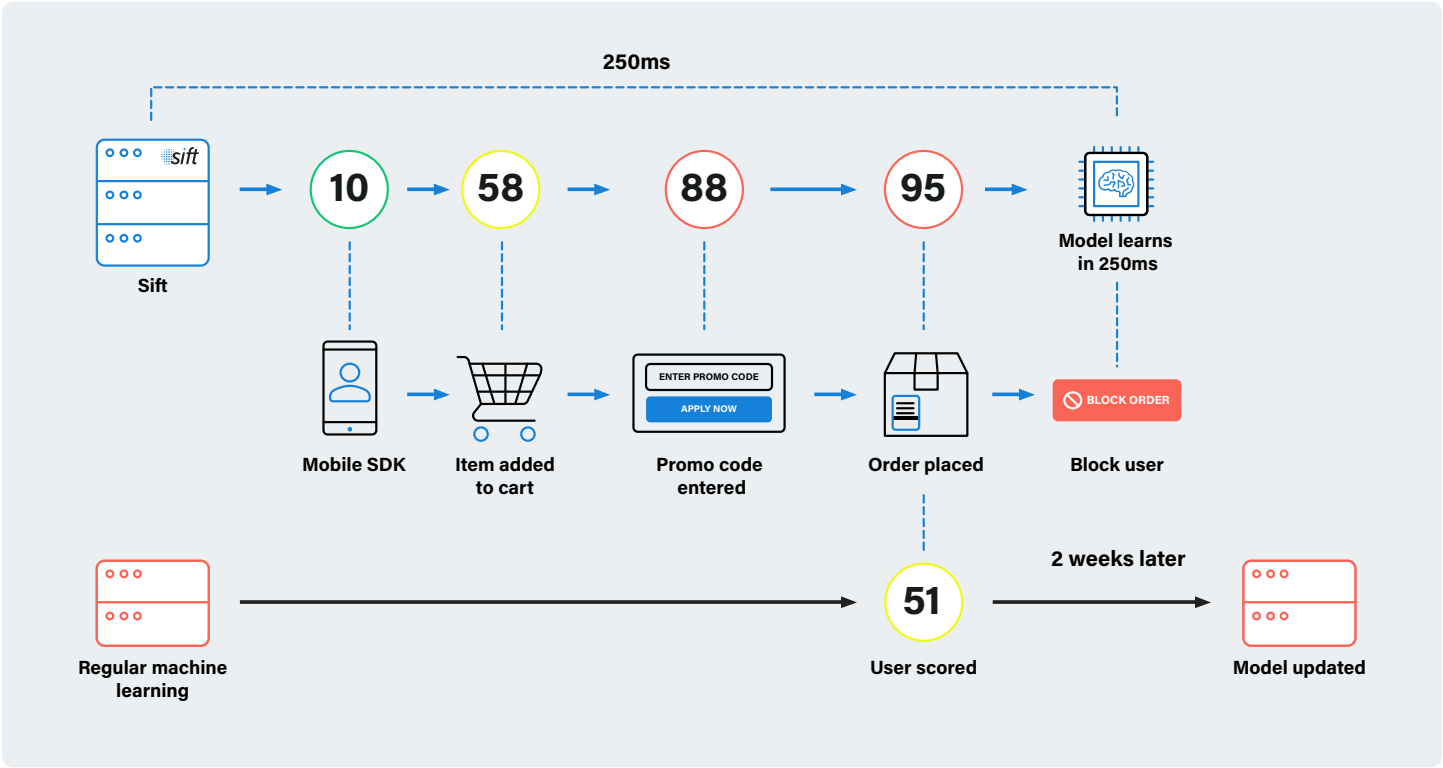
Each time we get an event – be it a page view or an API event – we extract features related to those events and apply a dynamic combination of the machine learning models described above. These features are then weighed against historic fraud we've seen both on your site and within our global network to determine a user or transaction's Sift Score. We also recalculate scores when related users are over a certain score threshold. There are two ways to access a risk score: via our API and through the Console.



 |

The unique qualities of Sift's machine learning

Traditional supervised machine learning models are trained on a batch of historical data and then shipped to production servers, which can classify incoming data based on those trained models. This process generally works, except if a new type of fraud emerges after the model was trained because the model will not fully adapt until the next batch training.



The integral part to our approach – what distinguishes us from any other solution in the marketplace – is our real-time machine learning, comprised of three critical components: real-time learning, proactive scoring, and rescoring. Together, they are responsible for our ability to detect and prevent fraud before it happens on your website or application – even if it's a brand-new, never-before-seen approach.

Real-time learning

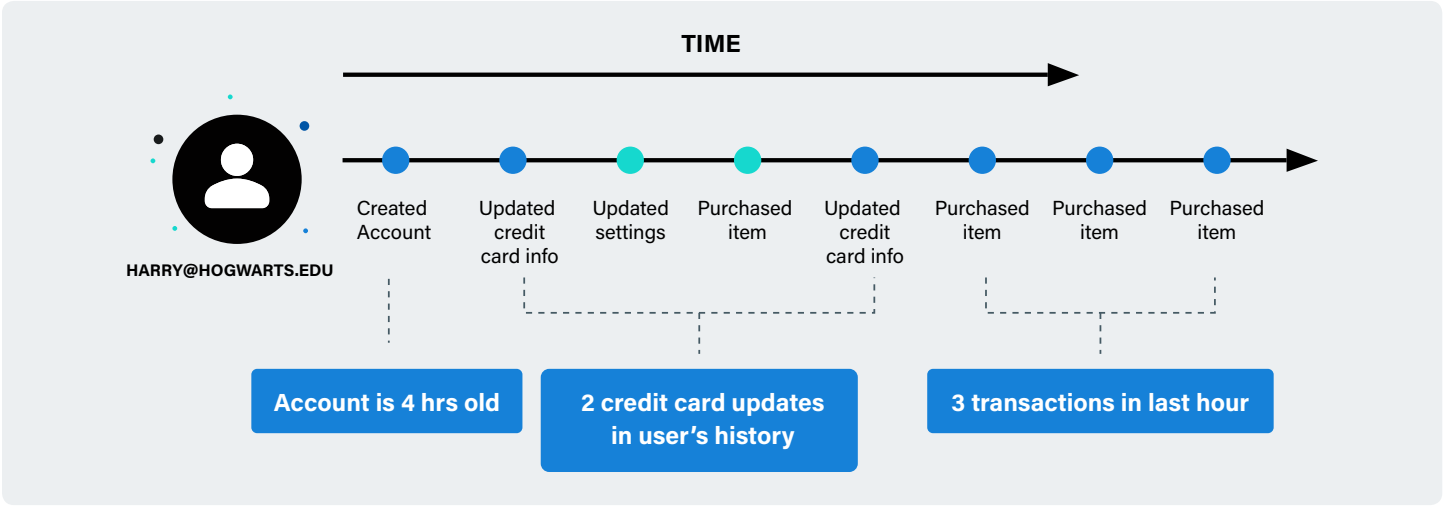
This is enabledthrough learnings from across the global network of websites and applications that use Sift. We share knowledge about bad behavior so that together we become smarter and better equipped to keep fraudsters at bay. The information is pushed in real time to update our network models and is propagated to all of our customers in just a few milliseconds. For example, a credit card tester is likely to have tested their cards on several different sites before hitting yours. With Sift, such card-testing behavior is identified on other sites and the fraudster's user information is propagated to your model in under a second; if and when the fraudster reaches your site, you're ready.

Proactive scoring

Most solutions calculate a risk score only at the moment of purchase on your site. Sift, on the other hand, recalculates the score every time the user acts on your site, based on new knowledge of fraudulent users and patterns. In many cases, this allows Sift to accurately identify repeat fraudsters before they do any other damage, and you can take action more quickly.

Rescoring

Sift also rescores users as we learn new things about them. For example, if Fred and Bill are related on a DeviceID and Fred scored highly, we go back and proactively rescore Bill since he's probably bad, too. This is part of the reason why our system and all computations that we do are constantly evolving.



With such high computing power needs, functioning at scale is a key requirement. Sift built a system that can maintain model parameters in a highly distributed way, but is still able to model users for new features with extremely low latency.

Telling a story with data

Most machine learning systems tend to be black boxes, revealing limited reasoning behind the decision to label a user or transaction as good or bad. Sift tells a clear and concise data story in its console. Built using D3 and React, the console provides context and the top signals that contributed to each Sift Score. You can review a user's order history, payment instruments, location information, and many other signals to get a holistic picture about the user or transaction. Our customers find this data critical to understanding not only their fraudsters but also their good users.

Increasing efficiency through automation

Sift's automation tools free up your analysts by reducing the amount of manual review and unlocking seamless customer experiences. Using the Sift Console, you can build and manage your business logic to automatically take action on transactions (accept, reject, etc.) or queue a subset of transactions for manual review.

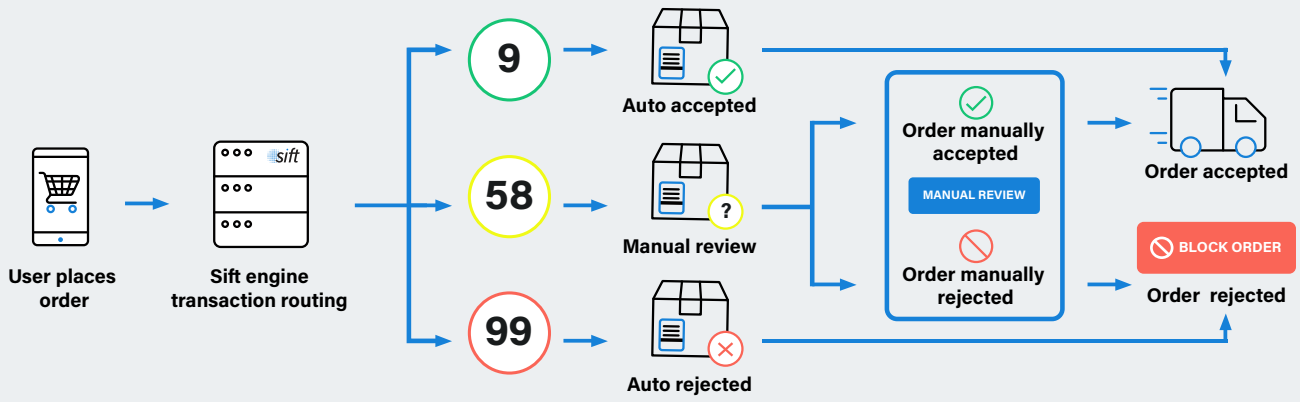
There are three components that comprise the automation layer of the Sift engine:

1. **Workflows**
   Build and manage your business logic within Sift. Take automatic action on users based on a set of customizable criteria (e.g., risk score > 80, first-time user, country = Japan, etc).

2. **Review Queues**
   This feature is the most efficient way for your analysts to manually review users and orders. The best part: they're built right into the platform, so you don't have to build your own solution or integrate with another vendor.

3. **Decision webhooks**
   Connect the actions taken within Sift to your other systems and services — no more switching between multiple interfaces. Whenever an analyst makes a decision or a Workflow applies an automated decision, we'll send you a webhook with the result.

Let's say you have a user with a very high Sift Score. You likely want to auto-block them and not waste time manually reviewing the order. However, if you have a first-time user with a low score but their order value is high, you may want to send them to a Sift Review Queue just to be sure. And if it's a return user with a very low score, you can auto-accept the order right away – and even remove friction that can stand in the way of conversion, such as entering a CVV code at checkout to make it as easy as possible for trusted users to make a purchase. Workflows let fraud managers create and maintain all of this logic themselves without developers having to build and maintain separate routing logic.



 |

Summary

Building a highly accurate system for preventing online fraud and abuse is a complex endeavor that requires constant monitoring, tuning, and engineering resources. You must be able to ingest large volumes of data, use that data in various real-time machine learning models and algorithms, manage automated business logic and decisions, and enable your review teams to investigate and act with speed and accuracy. Here at Sift, we're always thinking about how to empower our customers and have taken into account these many challenges and more. Businesses leverage Sift to prevent multiple types of fraud and abuse, while creating outstanding customer experiences.

 |