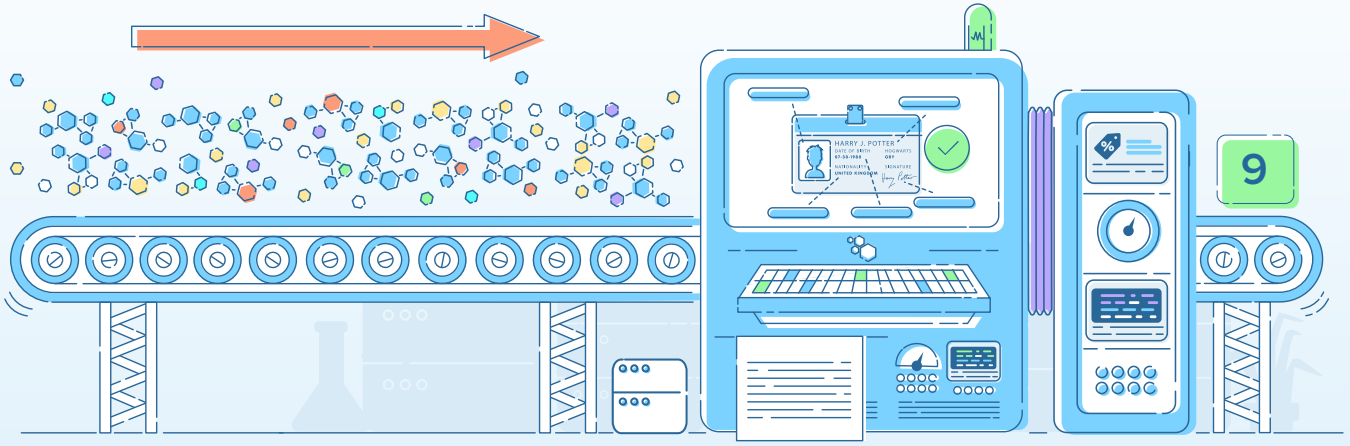


Machine Learning for Fraud Prevention:

What's Next

sift science

Contents

Introduction: Machine learning is powerful and ubiquitous

"A breakthrough in machine learning will be worth 10 Microsofts."

When Bill Gates made this dramatic statement to a group of MIT students in 2004, Facebook had just launched. Google had just filed for its IPO. Amazon Prime didn't exist. Technologically and culturally, it was a different world.

Did we ever think that ML would be used for predicting whether a criminal is a flight risk or diagnosing psychopathy, as it is today? Google now claims to be "AI-first," Facebook has its own AI research lab, and Amazon offers a machine learning platform.

5 Machine learning achievements from the past decade

2011	IBM's Watson wins Jeopardy, against two humans.
2012	A neural network developed by the Google Brain team learns to recognize cats from unlabeled images.
2014	Facebook's DeepFace neural network system learns to identify faces with 97.35% accuracy.
2015	Microsoft's Distributed Machine Learning Toolkit allows machine learning problems to be solved across multiple clusters of computers.
2016	Google's AlphaGo becomes the first computer program to to beat a human professional Go player.

Advancements in technology – including improvements in computing power and wide-scale availability of big data – have radically changed how companies use machine learning.

Processing. GPUs – which possess nearly 200 times more processors per chip than CPUs – have become the engine for machine learning, and deep learning in particular. Machine learning applications can see performance improvements of 5-10X when they run on GPUs, compared with CPUs.

Data availability. Though many machine learning technologies have existed for years, it's only now, with a gigantic and ever-expanding amount of data available from a variety of sources, that advanced capabilities like real-time updating have been unlocked.

Cost. As a bonus, it can be less expensive than ever to employ these advanced technologies. However, it would be a mistake for companies to assume that implementing machine learning on their own is easy or inexpensive. Building and maintaining the right infrastructure can be extremely complex.

The complexities of implementing machine learning

As more data arises, so does the complexity of understanding it all. When data is small, this is easier than when it grows by 10x, 100x, etc.

Companies that want to build their own machine learning infrastructure need to think from the beginning about supporting massive data growth, and an increasingly complex surface area of problems to tackle.

Think about:

- A scalable means of handling orders or magnitude of more data
- Pipelines for extracting, transforming and loading data for training and testing
- Systems in which online and offline analyses can be performed over the same data and the same code can run over both sets
- Infrastructure to enable ad-hoc analysis of an ever-growing data set

Advanced processing, data availability, and promise of lower costs: All of these factors together lead up to a perfect storm of machine learning innovation. And one of the main applications of that innovation has been within fraud prevention.

Why fraud prevention is an ideal application of machine learning

Think about all the areas that machine learning has already transformed: spam filtering (do you ever see spam in your email account these days?), speech recognition (think Siri), marketing personalization (such as Netflix recommendations), and search (Google, of course).

Fraud prevention is another area in which machine learning is revolutionizing both workflows and outcomes, allowing organizations to stay ahead of increasingly technologically advanced criminals.

Fraud: a costly and growing problem

Retailers will lose **$71B** to card-not-present fraud in the next 5 years.

Account takeover losses reached **$5.1 billion** in 2017, up 120% from the year before.

Sources: Juniper Research, Javelin Research

Clearly, there's a lot of money at stake. But businesses have additional reasons to embrace machine learning. As the internet grows, fraudsters are unearthing more opportunities to defraud both businesses and end users. The huge data breaches of the recent past have flooded the dark web with financial and personal information that can be used to commit account takeover and identity theft. And bad actors are turning to fraudulent user-generated content to commit scams, post spam, and commit malicious attacks.

Meanwhile, fraudsters' techniques are growing increasingly technologically advanced. All of the hardware and software needed to commit fraud at scale are on sale, often for disturbingly low prices. Today's online businesses are facing an increasingly sophisticated enemy that attacks, responds, and changes tactics extremely quickly.

With machine learning, companies can stay ahead.

Here are a few more reasons why machine learning is an ideal technology for preventing fraud:

Fraud hides within an immense amount of data

The most effective way to detect fraud is to look at the broad behaviors of end users. Leave no stone unturned. Don't just look at a transaction or order, but instead all the events leading up to and after the transaction. This amounts to a lot of data, and the best way to detect fraud in such a huge volume of data is with machine learning.

Fraud happens fast

When a machine learning system updates in real time, that knowledge can be used within milliseconds to update machine learning models and stop an attack before it does damage.

Fraud is unique

Every business experiences fraud slightly differently. Machine learning models are dynamic and evolve to changing patterns.

Fraud constantly changes

Fraudsters constantly adapt their tactics, making them very difficult to detect by humans – and impossible to detect by static rules-based systems, which don't learn. Machine learning, however, can adapt to changing behavior.

Fraud looks "clean"

To the human eye, fraudsters and good users don't appear any differently from each other. Machine learning has a more nuanced way of viewing data, which helps avoid false positives.

The benefits of machine learning for fraud prevention

- Highly accurate results
- Reduce the need for manual review
- Fewer false positives, through sophisticated behavioral analysis
- Ability to prevent fraud without impeding the user experience.
- Lower operational costs than other approaches
- Frees up teams' time to focus on more strategic tasks
- Can be automated
- Adapts quickly

Machine learning vs. rules in fraud detection

Across many industries, machine learning is displacing legacy solutions that just can't keep pace or deliver the same quality of results. In fraud detection, the outdated approach to fighting fraud is manually updated rules systems, which rely on if-then statements to apply decisions. The system goes through the rules, one by one, and if it determines any rule is tripped it will take the appropriate action and skip all the other rules.

Machine learning, on the other hand, is probabilistic rather than deterministic. It uses statistical models to look at past outcomes and anomalies to predict future outcomes. A machine learning system can learn, predict, and make decisions without being explicitly programmed.

Similar to how email spam filters learn to recognize which messages to deliver to your inbox, a machine learning system can distinguish the characteristics of fraudulent purchases from legitimate ones.

Machine learning is often deployed as part of automated fraud screening systems, identifying high-risk transactions, accounts, and risky logins to prevent payment fraud, account abuse, content abuse, and account takeover. Machine learning can replace even the most complex rules set and produce higher accuracy, fewer false positives, and savings through automation.

Here's a quick table listing the pros and cons of rules vs. machine learning:

Machine Learning	Rules
Proactive - tells you what's happening	Reactive - tells you what already happened
Learning across many data elements	Limited data elements
Scalable	Higher maintenance
Needs statistical significance	Can operate on a small data set
More accuracy at scale	Less accuracy at scale

The Nuts and Bolts of Machine Learning

There are several stages to developing a machine learning system. You collect data, clean the data, and perform feature extraction. Then you build a model and deploy the model. Next, you evaluate and monitor the model. Finally, you iterate until good performance is achieved. It's a complex process.

This is why so many organizations are leveraging API-based machine learning systems for fraud prevention. It's possible to leverage enterprise-grade machine learning without investing in building your own. For example, you can integrate your desktop and mobile experience using Sift's Javascript snippet and SDKs, and connect your site's backend systems using our REST APIs.

Let's run through the basics.

Data processing

Successful fraud detection requires access to massive quantities of high-quality data. The phrase "garbage-in, garbage out" is applicable here, as the quality of the results you see are directly in line with the quality of data you send.

As data sets grow ever larger and more complex (we're talking petabytes, not megabytes), it becomes more necessary for machine learning platforms to be able to scale.

As an example, here's a small fraction of the data Sift collects via SDK, Javascript snippet, and API:

Raw data collected

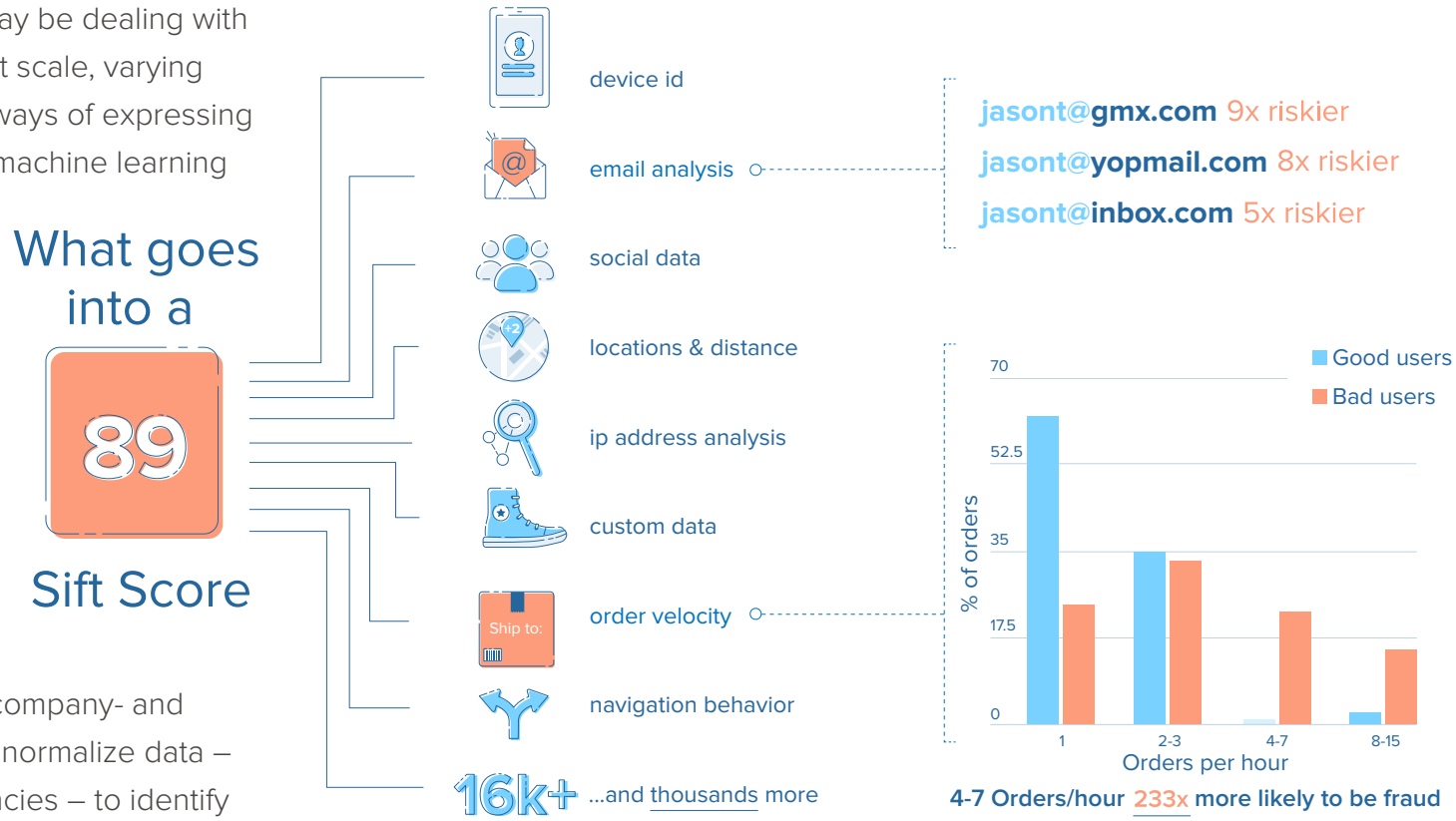
User identity	Attributes that are associated with the identity of a user *Examples: Name, email address, phone number*
Behavioral patterns	Preferences and patterns associated with the user *Examples: Browsing patterns, keyboard preferences, screen tilt*
Locational data	Location attributes associated 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 & 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*

Preparing data: Data normalization and feature extraction

Before a machine learning model can do its work, data needs to be cleaned and formatted to be fed into the model. Suspicious behavior is often buried within streams of data.

Data normalization

Raw data is messy. You may be dealing with measurements of different scale, varying currencies, and different ways of expressing addresses. In order for a machine learning algorithm to understand that Oak St. and Oak Street are the same (for example), the data needs to be normalized.



At Sift we map our database of fraud and abuse-related data against meaning, relationship structure, and relevancy – often in company- and market-specific ways. We normalize data – like addresses and currencies – to identify repeatable data patterns and weigh the likelihood that two data inputs are correlated with each other.

Get a deep dive into data normalization at Sift.

Feature extraction

Feature engineering transforms raw data into structured, machine-processable formats that can be understood by a machine learning algorithm.

Feature extraction for fraud prevention is both challenging and resource intensive. It requires deep domain knowledge and experience, since fraud patterns are constantly changing. After years of dealing with fraud across different verticals, geographies, and applications across mobile and desktop, Sift has built a library of over 10,000 features that we use to uncover fraud patterns. 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.

Learn more about feature engineering at Sift.

> In a survey of 16,700 members of the data science community, Kaggle found that "dirty data" was the top challenge faced at work.

Model building

When talking about building models, the first thing to remember is that the algorithms aren't as important as the data a system has access to. It's a combination of the use of algorithms and vast sets of data that differentiate machine learning systems.

Typically, supervised machine learning centers on a cycle of training, testing, evaluating, and iterating to drive performance. At Sift, instead of relying on a single machine learning model, we use ensembles of several predictive models across different types of fraud.

Some models are trained with a general understanding of fraud patterns across our network of customers, some are built for the industry a particular business operates in, and others are tuned to a specific organization's data. This ensemble of models allows us to accurately score a transaction, a user, or a session while taking a holistic approach when analyzing risk.

Performance varies from one algorithm to another and there is no "silver bullet" solution for machine learning. We take a comprehensive approach, where we use weighted mixes of many models (deep and "shallow") models to find the best combination of predictors for each customer in each vertical for each type of fraud we prevent.

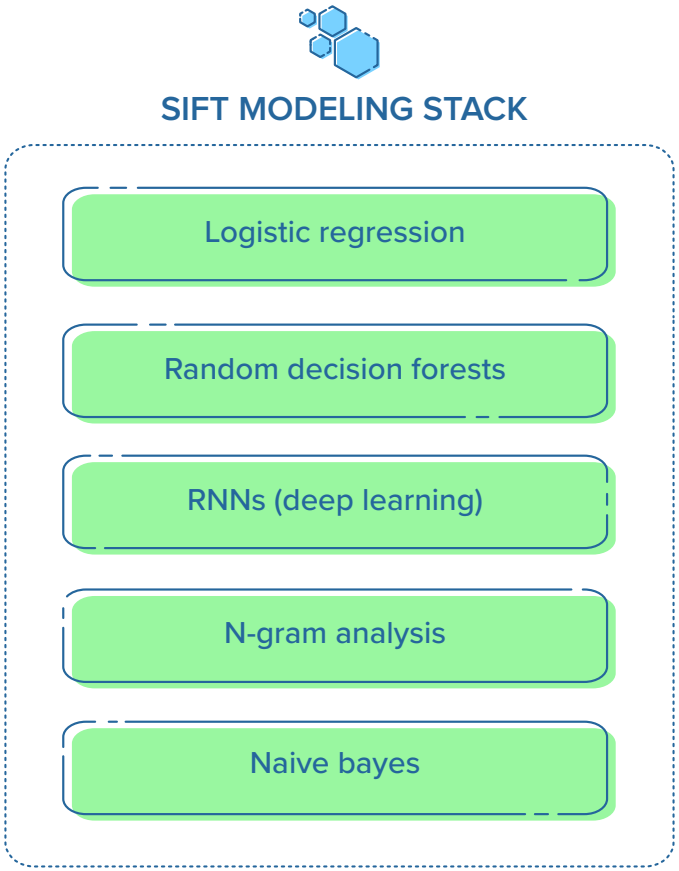
Machine learning algorithms

No model is perfectly accurate. There are strengths and weaknesses to different types of models. At Sift, we've discovered that combining multiple models together far outperforms any individual model.

That said, here are some of the algorithms we employ:

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 (for example, a guest check-out experience while shopping). Logistic regression models provide easily interpretable results.

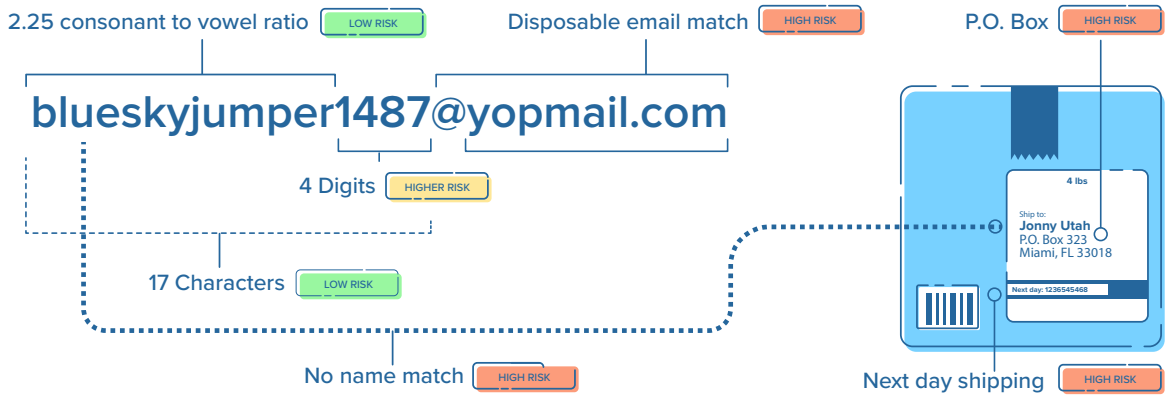


Random decision forests

Random decision forests are a powerful, scalable, and intuitive model. 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.

N-gram

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 detecting fake accounts and fraudulent content. 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).



Naive Bayes

Naive Bayes is often considered a trivial model, but we find that it contributes to producing accurate results in a large stack of algorithms, particularly while onboarding new customers with limited training data or providing reasons why and how we arrived at a particular risk score.

Latest developments: Deep learning

Machine learning has clearly come a long way in a short time, and is currently being used to solve challenging problems in fraud detection. But what's next?

Many of the algorithms listed above have been in use for years. But recently, some industry-leading companies have been introducing a different technique to their fraud prevention approaches.

Neural networks, more popularly known as deep learning, are unique in that they require no human programming – and still deliver impressive results. At Sift, we are already using RNNs, with impressive results.

Though neural networks are complex to implement, the open-sourcing of Google's Tensor Flow project, including a distributed trainer, makes it easier for companies to leverage the power of deep learning.

> At Sift, we've discovered that deep learning improves the accuracy of our multi-model approach to fighting fraud for our customers by as much as **10%**

Focus on deep learning

AI-focused startups received **$1B** in funding in Q2 of 2016

Google was working on more than **1,000** deep learning projects in 2017

Sources: CB Insights, Fortune

What's deep learning?

While most machine learning models need to be trained with data in order to make accurate predictions, a deep learning model can uncover patterns and learn independently – much like the human brain. These algorithms are built on multi-layered artificial recurrent neural networks (RNNs) to effectively gather data insights and learn complex relationships between inputs and outputs. Deep learning

requires significant computational power, such as GPUs.

Although the concept of neural nets has existed since the 1950s, and most of the big advancements were made in the 1990s, a combination of increased processing power and the ability to successfully access and store big data has resulted in a deep learning Renaissance of sorts.

How deep learning is used at Sift

At Sift, we realized that relying exclusively on hand-engineered features can leave gaps in modeling user behavior. For example, a fraudster may browse in a very unusual way, with atypical timing between page views, as in a scripted attack. We explored signals derived more directly from this sequence of interactions with our customers' services.

Many fraudulent activity sequence patterns are usually immediately obvious to a fraud analyst, but modeling them through hand-engineered features can quickly devolve into a wild-goose chase. Take credit card testing: Fraudsters want to do as little work as possible to achieve their goal, so they may use the same account to quickly go through stacks of stolen credit cards in an easily recognizable pattern, alternating between making transactions and adding new cards to the account.

Another example of questionable behavior is when an attacker testing stolen credentials produces many suspicious logins and then posts an atypical quantity of content, or quickly navigates to value-extraction pages (to checkout, to transfer money to

5 familiar applications of deep learning

1. Self-driving cars
2. Voice search
3. Automatic text translation
4. Image recognition
5. Recommendation engines

other users, etc). These patterns may be intuitive at a high level, but manually implementing a mechanism to detect them easily becomes impractical to do in a general fashion.

We use deep learning to learn on sequences of events in our 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.

Our RNNs look at the time series of the events for all of a customer's users, and then we use that information to automatically abstract features and create a model that predicts what sequences of events are correlated with good or bad behavior.

We're also currently working on Long Short Term Memory (LSTM) models and deep learning to prevent fraudulent content – like spam and scams – from being posted.

Benefits of deep learning

Deep learning enables us to solve complex problems, using a higher level of abstraction.

Automatic feature engineering

As we explained earlier, feature extraction is one of the most important aspects of creating a successful machine learning system for fraud detection. But this step involves extensive engineering investment, taking the time to identify the most relevant and impactful features to use. With deep learning, there's no need for humans to perform feature extraction, which saves considerable time and resources.

Effective in ensemble models

At Sift, RNNs have proved to be extremely effective at predicting certain types of fraud. We combine deep learning with other models to create what we call an ensemble model. Those ensemble models that included RNNs were able to significantly outperform those that didn't.

Great potential for learning from biometric data

Deep learning sets the stage for using behavioral biometric data – like typing, mouse activity, and how someone holds a mobile device – to prevent fraud. Biometric data is incredibly complex, making it extremely difficult to extract the right features to feed into a traditional machine learning model.

RNN models, however, are perfectly suited to automatically engineer high-level features. This ensures that the most valuable patterns will be identified from amongst a seemingly unlimited amount of potentially useful data. No individual element will be overlooked.

Deep learning in action: Time series analysis

Unlike our previous models, our deep learning models work exclusively on a sequence of user events. We can use deep learning to look at how a user moves from page to page, including both the sequence of pages they view and the tabs they are on.

Example of a good user's behavior:

- Create an account
- Sign in a few seconds later
- Browse for a few minutes
- Log out a few minutes after that
- Sign in a few days later
- Search for an item and add it to their cart
- Check out a few minutes later

Example of a fraudulent user's behavior:

- Create an account
- Sign in a half-second later
- Navigate directly to a product page
- Add item to the card and check out immediately
- Update their credit card on file a few minutes later
- Order another item with this credit card within a few seconds

When we combine the sequence model's prediction with predictions from our other models, we achieve extremely accurate results. Learn more about how we test and deploy RNNs at Sift.

Conclusion

> "I think AI is akin to building a rocketship. If you have a large engine and a tiny amount of fuel, you won't make it to orbit. If you have a tiny engine and a ton of fuel, you can't even lift off. To build a rocket you need a huge engine and a lot of fuel."
>
> **- machine learning expert** Andrew Ng in Wired magazine

We finally have both the engine and the fuel for an AI renaissance. While machine learning has come very far and is impacting our lives in countless, in reality it's still getting started. Soon, the question of whether to leverage machine learning and deep learning at your organization will not even be a question – it will be a given.

In the meantime, forward-thinking online businesses are enjoying the many benefits of applying machine learning to fraud prevention. Would you like to learn more about how Sift's Live Machine Learning works? Check out this deep dive into our technology stack.