THE FUTURE OF FRAUD FIGHTING

an intro to machine learning

sift science

table of contents

introduction

Machine learning

—a buzzword to be sure. In this digital age, do you really know what machine learning is and how it can be used?

Chances are good that machine learning is already integral to your day-to-day life. Crazy, right?

The goal of this ebook is to help you understand machine learning and its application in reducing fraud on your website. With just a few keystrokes and the right tools, you can stop fraudsters in their tracks.

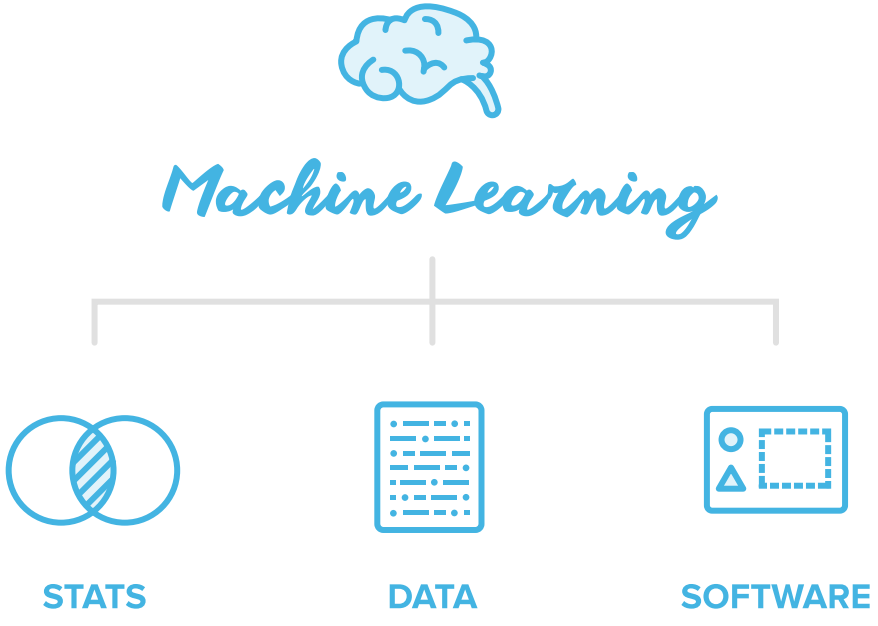
Fraud is a $200 billion dollar, truly global industry. Cyber criminals steal from more than one million people...every day! If you're an individual hit by fraud, hopefully your bank or insurance company can help you out. However, businesses — especially small business owners — find that every dollar lost compounds to a staggering degree.

At Sift Science, we think of ourselves as Fraud Fighters, wielding our trusty sword of machine learning against the crooks that want to rob your site. Follow us as we journey through the intricacies of the interwebs to not only grasp the *what* of machine learning but also how to put it to work for you fast.

01 what is machine learning?

So what exactly is *machine learning?*

Machine Learning = Statistics + Data + Software.
Machine learning (ML) falls under the umbrella of computer science. At its core, machine learning refers to the practice of training computers via software to recognize patterns and infer predictions, emulating a human-like ability to learn from "experience". In ML, the "experience" that machines get is from humans not only inputting but also indicating useful data. But why doesn't data alone make for machine learning?



01 what is machine learning?

Let's think about data itself. Chances are, your computer has spreadsheets full of numbers and names that encompass your past sales and communications. However, you cannot predict your site's future unless you take your data to the next level. Machine learning is that next step.

With ML, computers — via specially created algorithms and mathematical formulas — can learn from historical data and suggest likely future scenarios. The magic of machine learning is that after the initial set-up, your machine learning system needn't further explicit programming outside of occasional training and course-correcting.

02 why do we want machine learning?

What's so great about machine learning?

If I told you to imagine a tiger, your brain can conjure up an image, right? What are some of the key elements that connote *tiger* in your brain? *Big cat*, *stripes*, *tail* — all of these are relevant descriptors. But how does your brain distinguish between a tiger and a lion (they are both big cats), or between a tiger and a zebra (stripes for the win!).

Your ability to not only learn patterns in data but distinguish between and weigh the importance of each pattern is essential to your intelligence. **This ability is what engineers strive to emulate when building machine learning systems.**

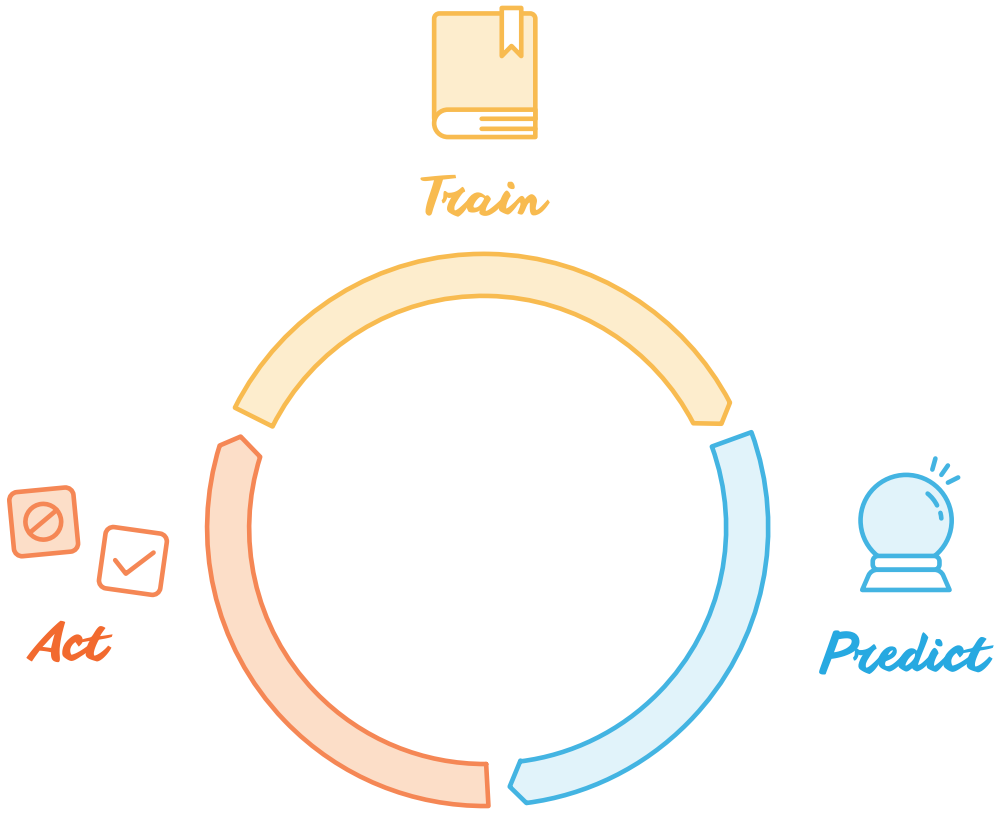
TIGER ZEBRA

Machine learning allows for a reduction in human effort. The time it takes humans to read, synthesize, categorize, and evaluate data is significant — and that's all before any action takes place. ML teaches machines to identify and gauge the importance of patterns in place of humans. Particularly for use-cases where data must be analyzed and acted upon in a short amount of time, having the support of machines allows humans to be more efficient and act with more confidence.

03 what does machine learning do?

Machine learning offers automated, accurate predictions.

Depending on the purpose of a specific machine learning system or algorithm (called models), ML can turn dense and confusing information into a narrative that suggests the likelihood of a future action. By continually adding data and experience, a user further trains the ML system, making its predictions more accurate and more relevant to that specific use-case. At its core, machine learning is a 3-part cycle.



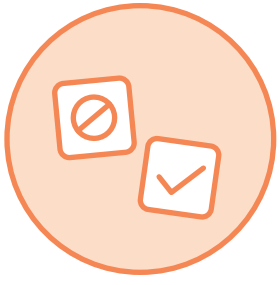
03 what does machine learning do?

Train

The first step of of any ML system is to **train** the model. Herein, the algorithm receives its basic purpose: to extract patterns from data. Subsequent training can come from the user and doesn't require further technical guidance.

Predict

Once the algorithm has sufficient data, it can make predictions. These predictions essentially answer the question, "what matches the data pattern?".



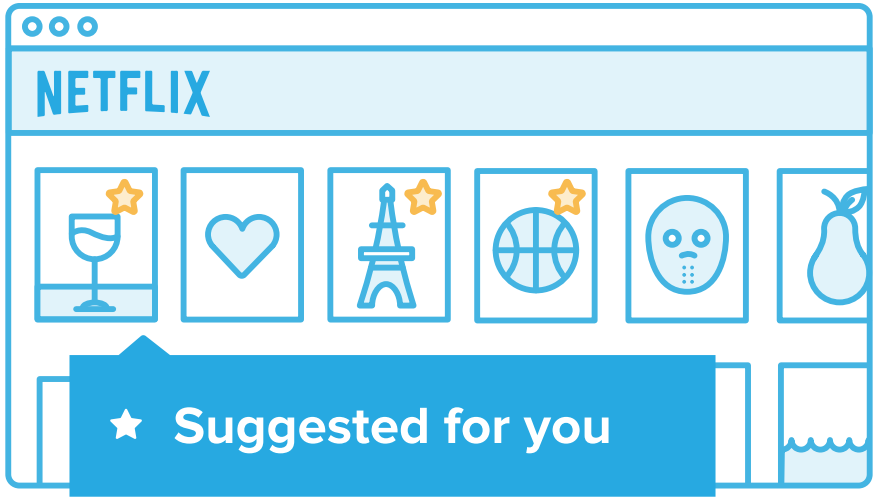
Act

ML is only as accurate as its data. In order to grow even more accurate, the model needs feedback, which is the **Act** phase. In this step, the user indicates whether the prediction is correct or incorrect, which in turn **trains** the model, restarting the cycle.

04 what are examples of machine learning?

How does machine learning affect my day-to-day life?

Excellent question! Do you use Gmail? How about Netflix? Guess what -- both of those sites use machine learning to optimize your experience.



Netflix is a video-streaming site that offers viewers access to TV shows, documentaries, movies, etc. in exchange for a subscription fee. Netflix constantly asks users to review the shows and movies that they've already watched.

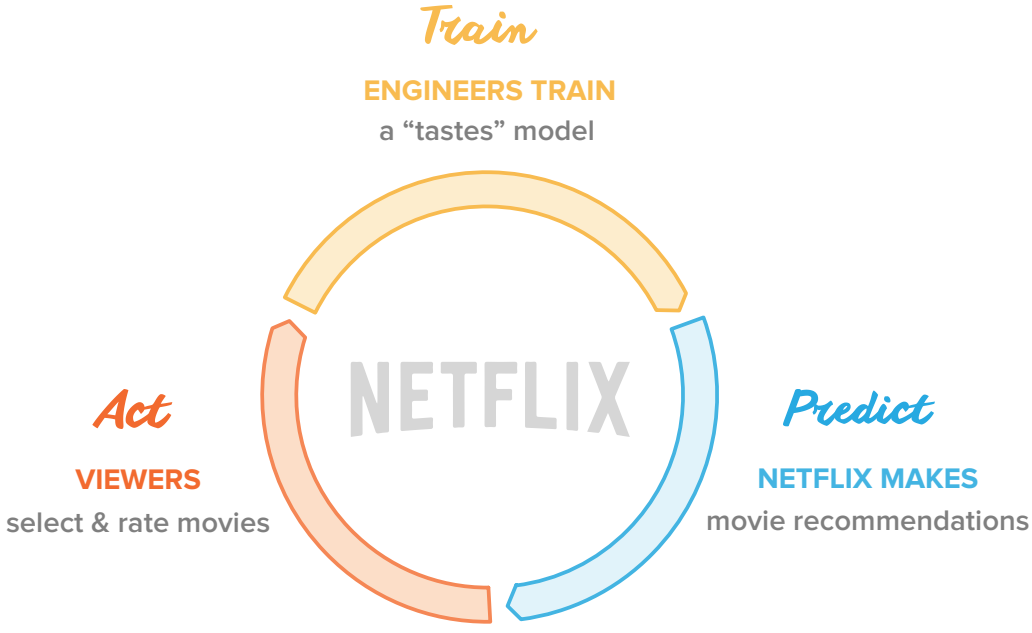
The purpose of those reviews (via a star-rating system) is to train the Netflix recommendation model, based on what you choose to watch and how to react to those choices. Netflix wants to understand your taste and preferences. The Netflix's ***Suggested for You*** is machine learning, automatically gathering data to tailor the experience to you.

04 where does machine learning work?

Why does Netflix want to create such a **you**-specific experience?

Netflix "estimates that 75% of viewer activity is driven by recommendations." Those recommendations take into account not just the viewer's historical taste but also the viewer's ongoing viewing trends, changing as viewer selections change. Sometimes, we can confuse the algorithms, like when a friend picks the movie for movie night. A different user may account for some unexpected future recommendations, but that's how you know that the ML system is working.

Retrain your Netflix model by selecting the *Not Interested* option for suggestions that are out-of-line with your tastes. There you go — you're using machine learning!



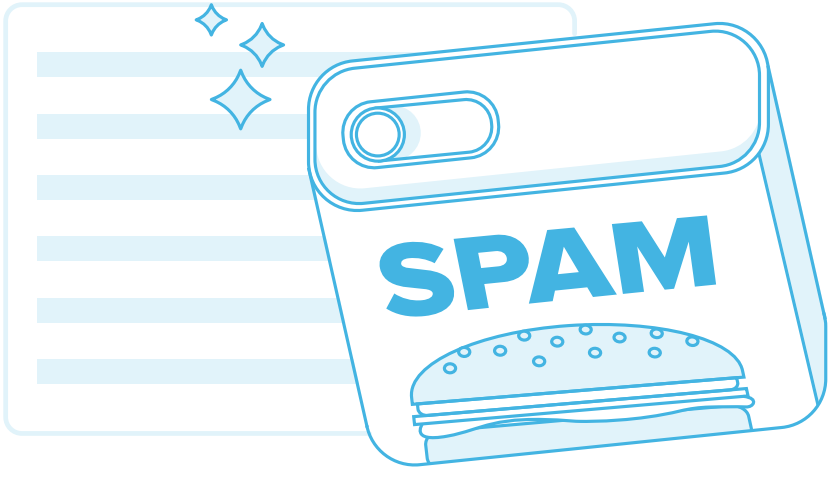
05 how accurate is machine learning?

Is machine learning actually accurate?

Yes, very! For many machine learning models, the data is always incoming and users are constantly acting on the predictions. Take for example your email. In your email inbox, how often do you currently see spam messages? Maybe once every few weeks? And how rarely do good messages get stuck in your spam folder?

On the whole, your email account is pretty accurate in weeding out spam or fake emails, and you haven't done a thing to teach it wrong from right. That's because it's predicting spam with machine learning.

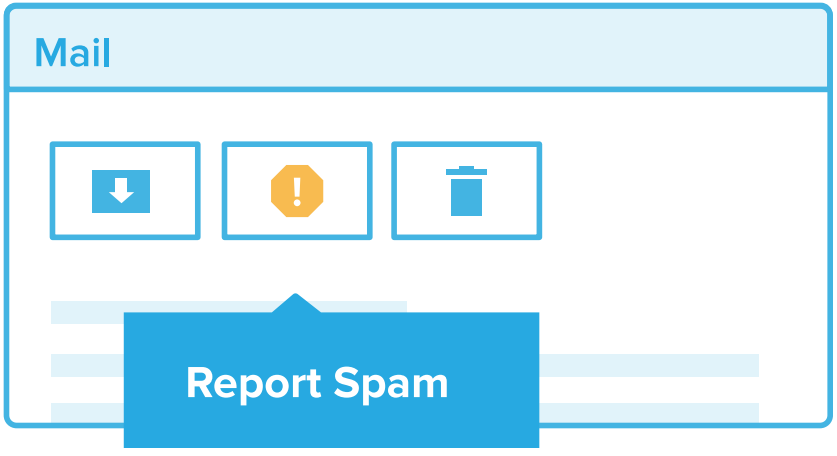
If you didn't have machine learning working to keep your inbox clear, you'd see lots more Viagra offers, desperate notes from deposed princes, and "You've won!" emails daily. The 2013 Symantec Intelligence Report estimated that spam made up 68% of all email traffic in 2012. That the ML algorithms can catch the vast majority of these spam messages is pretty impressive!



05 how accurate is machine learning?

How do these algorithms work?

Your email ML system is always looking for patterns in data. Certain keywords appear in spam messages more frequently than good messages. Terms like “Viagra” or “belly fat loss,” indicators such as the message length or email domain, and data from the sender’s email address itself may suggest spam.



Every time you click “Report Spam,” you’re acting to train your email’s ML system to recognize the ever-changing signs of junk mail.

When you find a good message stuck in your Spam folder and mark as “Not Spam,” you’re doing the same. So even when it’s right out of the box, machine learning can be exceptionally accurate. Likely, you receive all of the emails that you want and don’t often see those spam messages -- that’s machine learning at work!

06 how does machine learning apply to fraud?

What can machine learning do against fraud?

As evidenced by the email spam example, bad users leave behind signs. These calling cards of fraudulent activity can be used to train a machine learning model to spot credit card fraud, fake accounts, referral fraud, and other types of illicit activity that plague websites. Having a human review every new order or newly created account — known as *manual review* — is too time consuming as traffic and transactional volume increase.

The solution? Put a system in place that identifies the most glaring examples of fraud, while flagging those that require further, closer examination.

Although creating a simple set of business rules based on common fraudulent characteristics may help you at first, such a solution isn't scalable. Why, you ask? Well, as you learn and identify fraud signals, you have to manually create an **If X then Y** system. Adding a new rule for every IP address or email domain where fraud has occurred is unscalable and unreasonable.

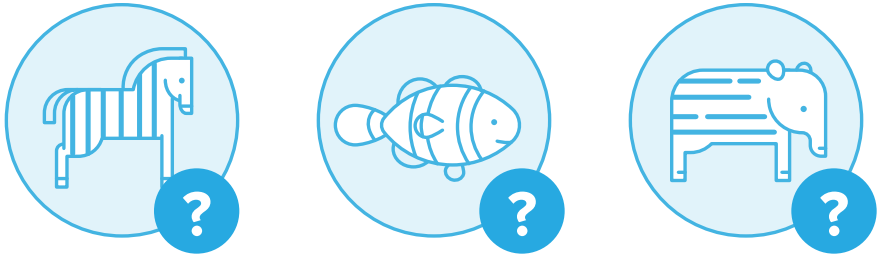
Imagine that your online store is hit by a fraudster from Venezuela. If you were using an **If X then Y**, rules-based system, you might set up a business rule to deny any and all orders coming from Venezuela. But what if only 15% of all Venezuelan orders are fraudulent—you'd be missing out on the profits of that other 85% of orders, and creating a negative customer experience for those good shoppers.

06 how does machine learning apply to fraud?

Additionally, what if those many indicators — shipping address, IP address, social network data, shopper credit card issuer — carry different weights? Let's return to our animal example to better visualize the impact of static **If X then Y** systems on detecting online fraud.

If I tell you, "*Animal X* is approaching; it has stripes," do you automatically assume that *Animal X* is a tiger and therefore terrifying? Of course not! It could be any number of striped animals — a zebra or a clown fish or a tapir.

There are countless other and more important data points that would help you decide whether you should run away or stick around to finally meet a tapir. Even if a non-ML system had data points that included, "*Animal X* has stripes, is orange and black, and can be found in trees," this mystery animal could still be a monarch butterfly or a tree frog. But with sufficiently weighted data (*Animal X* has a long tail, a **loud roar**, is a **big cat**), your brain would be able to deduce that, "Hey, I think that's a tiger. Time to get outta here!"



ZEBRA CLOWNFISH TAPIR

06 how does machine learning apply to fraud?

The ability to take in data and understand the relative values of each data point in relation to the whole is where intellect comes into play. ***That*** is what machine learning can do for fraud detection. With a trained ML system, analysts can accurately assess which orders are tigers and which are tapirs.

Machine learning is a voice of reason in fraud detection. When combined with human intuition, statistical insights can bring about better decisions and long term financial savings.

Rather than canceling orders from good customers simply because they may share an insignificant attribute with a fraudulent or charged back transaction, machine learning can weigh the many other indicators that contribute to a customer's profile.

“The ability to take in data and understand the relative values of each data point in relation to the whole is where intellect comes into play.”

06 how does machine learning apply to fraud?

Human Judgement

explains

creates

can tell what's meant

Machine Learning

describes

replicates

can tell what's said

Fraud is rarely cut and dry. ML gives users probabilities, not black and white judgements. Probabilities allow you to understand the nuances in your data, enabling you to make better decisions and teach the ML model to make even more accurate predictions in the future.

> **You need human judgement to know if the person you're calling is a fraudster.**

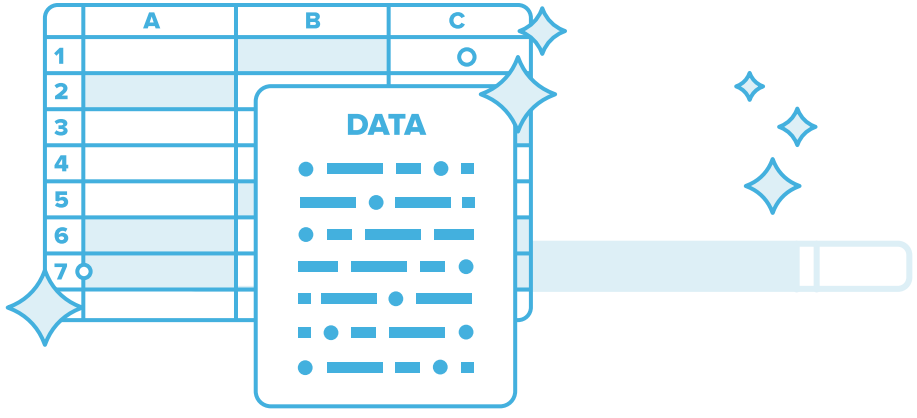
07 does it have any weaknesses?

Machine learning is only as strong as its data. If an online business owner is only predicting based on her own, limited pool of shoppers, she must experience each unique fraud attempt before being able to predict future attacks. This method is exceptionally unsustainable.

With machine learning operating on a global pool of data, however, a shop owner in Portugal can benefit from the fraud experience of a shop owner in Portland or Odessa or Kyoto.

The vastness of data from a wider net of experience allows predictive behavior without actual losses. That way, shop owners that have yet to experience fraud from a specific IP address or email address can benefit from the knowledge of those who have.

Perhaps you've heard of the network effect? In order for machine learning to effectively and accurately detect fraud before it happens, a critical mass of data and users must be achieved. Unless you're an Amazon.com, operating with millions of worldwide data points, using machine learning without any input from outside of your business is likely ineffectual. Machine learning needs a deep set of data in order to be truly magical.



07 does it have any weaknesses?

Once connected to a vast network of order and transaction history, ML for fraud is unstoppable.

Certain fraud fighting tools can even train fraud detection systems in real time. What does this mean? Say a shop owner in Shanghai sees a chargeback on an order placed by Freddie Fraudster in Hong Kong. Once that Shanghainese shop owner marks Freddie as “bad” in the global ML fraud detection system, immediately and automatically Freddie’s details will be updated for every other shop owner on the network worldwide. So when Freddie tries to buy something from a Londoner’s website 2 minutes later, the Londoner will see that Freddie is a bad user and can cancel Freddie’s order before she too suffers a chargeback.

08 fraud fighting in the future

Criminals are constantly cooking up new ways to rob hard-working folks. Whether it's diluting good content with spam messages, stealing credit card information, or scamming a referral or coupon system, fraud is an ever-evolving, ever-growing industry.

Thankfully, machine learning enables the good guys to stay one step ahead of the fraudsters. When properly leveraged, ML brings together sets of information so vast and so deep that they would be otherwise incomprehensible. Allowing manual reviewers to focus on only flagged data points saves time and money. By combining the power of pattern recognition with human intuition, machine learning is driving the evolution of how we process data.

Machine Learning

Machine learning is a necessary tool in the arsenal of today's savvy business leaders. Whether used to stop criminals before they hit your bottom line, project market trends, or flag spam emails, ML is how we optimize our lives.

Interested in learning more about the power of machine learning? Have a fraud problem that needs an accurate and innovative solution?

We're here to answer any questions—just drop us a line at scientists@siftscience.com or visit our website at https://siftscience.com.

references and further reading

Medium offers an excellent overview and exercises in introductory machine learning.

Symantec's regular intelligence reports offer invaluable insight into the impact of fraud.

Wired Magazine's article on the Netflix suggestion algorithms is an interesting read.

Sift Science offers a "Machine Learning Demystified" overview on YouTube.

Business Insider's machine learning overview with embedded Nova video are helpful.