

6 Machine Learning Myths

Machine learning is everywhere – but many people aren't even aware of it. It's on smartphones, automatically tagging photos and powering the voice-activated intelligent assistant. It's on numerous websites, powering chatbots and providing personalized recommendations. It's in connected cars, allowing drivers to control the infotainment system using voice-activated commands.

Let's start with a basic definition: Machine learning is a discipline that involves getting computers to learn without the need to be explicitly programmed. Andrew Ng, VP and chief scientist of Baidu, [describes](#) machine learning as “the technology that lets a computer get smarter and smarter all by itself just by looking at data.”



5-myths-machine-learning

But even if you're armed with a general understanding of what machine learning is, there are still a lot of misconceptions floating around out there about what it actually takes to successfully apply it. Here are a few of those myths:

Myth #1: It's all about the right algorithm.

Nope – it would be more accurate to say that it's all about the data. As Sift Science CTO Fred Sadaghiani puts it, “data is orders of magnitude more important than the algorithm you use or any technique that you're applying.”

In terms of data, think both quantity and quality. The more data you provide the system, the better results you'll get. And providing the *right* data is equally (or even more) important. To properly train a model, it needs to have access to the best representative data from real-world scenarios.

Myth #2: Machine learning always performs tasks in real time.

The perception that machine learning always performs tasks in real time is fairly common. However, only systems that are built for online learning can actually perform in real time.

For example, some fraud prevention systems use machine learning to create models for automatically recognizing and identifying fraudulent transactions. When a transaction comes back as fraudulent, that information needs to be added to the models so they can recognize and identify the characteristics of that type of fraud in the future. Systems that use batch learning would group that transaction together with other examples, and the learnings would be applied all together, at a point in the future.

Using online learning, however, a fraud prevention solution can *automatically* apply new fraud information to models across the system in near real time – without the need to be updated manually.

Myth #3: Machine learning technologies are a black box.

They can be, but they don't have to be. Numerous companies have developed proprietary machine learning technologies, but do not provide any insights into how the machine learning algorithms and models work. This can make users feel like they're essentially placing their trust in a "black box."

But some companies using machine learning technology are transparent about how the technologies work, by providing explanations of the results. For example, [Sift Science features a dashboard](#) with visualizations providing insights into why transactions are scored the way they are. Providing the "why" helps make machine learning less of a mystery.

Myth #4: Machine learning is absent of human bias.

It is extremely difficult, if not impossible, to completely remove human bias when using machine learning. Quality data is crucial to machine learning; data filled with human bias can greatly impact machine learning applications. One of the most well-known examples of machine learning going awry due to human bias is Microsoft's now defunct AI-powered chatbot "Tay."

Microsoft released Tay earlier this year with the goal of conducting research on conversational understanding. The bot was able to learn, to a certain extent, from interactions with users on Twitter, GroupMe, and Kik. It didn't take long for Twitter users to teach Tay to be racist and offensive. Microsoft pulled the plug on Tay within 24 hours after its launch.

Tay wasn't programmed to be racist and offensive. However, the bot was influenced by the conversations it had with humans, conversations that in

many cases contained human bias.

Myth #5: Only skilled data scientists can use machine learning.

Not long ago, machine learning was only for professional data scientists and other technology experts – which meant that only major technology companies and enterprises could take advantage of it. The science of machine learning is complex, constantly evolving, and can be difficult to understand.

However, there are numerous organizations aiming to democratize machine learning so that anyone can use it. Major technology companies like Amazon, Baidu, Facebook, Google, and Microsoft are building amazing machine learning-powered platforms and applications. Machine learning is now available to everyone via open source projects as well as affordable machine learning-based solutions and APIs.

Examples of open source machine learning projects include Apache Spark's [MLlib](#), [H2O](#), [scikit-learn](#), Google's [TensorFlow](#), and Baidu's [Warp-CTC](#), just to name a few.

Myth #6: A single machine learning model will solve all your problems.

There's no single machine learning model that will act as a silver bullet for fraud. This is why Sift Science uses an ensemble of machine learning models (naive bayes, logistic regression, and decision forests) to detect different types of fraud based on specific signals and behavior.

Each customer also receives a custom model tailored to their business, but they also benefit from a global model that shares data collected from across the many types of businesses across the network – plus models targeting different types of fraud (payment fraud, account abuse, content abuse, or promo abuse).

Conclusion

Many companies are using machine learning to build solutions that can help protect businesses from fraud. Some companies however, are not using machine learning as effectively as they should; the term “machine learning” could be considered more of a marketing buzzword.

Machine learning systems are not all equal, so when it comes to choosing a machine learning-based fraud prevention solution, do your homework: research the company, try the demo, and make sure the company is using this sophisticated technology effectively.

To find out more about how machine learning can be applied to fraud prevention, check out the [Machine Learning – The Future of Fraud Fighting](#) ebook. Curious about Sift Science’s machine learning approach? Read more in [Not All Machine Learning Systems Are Created Equal](#).



Janet Wagner

Janet Wagner is a technical writer who specializes in creating well-researched, in-depth content about machine learning, deep learning, GIS/maps, analytics, APIs and other advanced technologies.