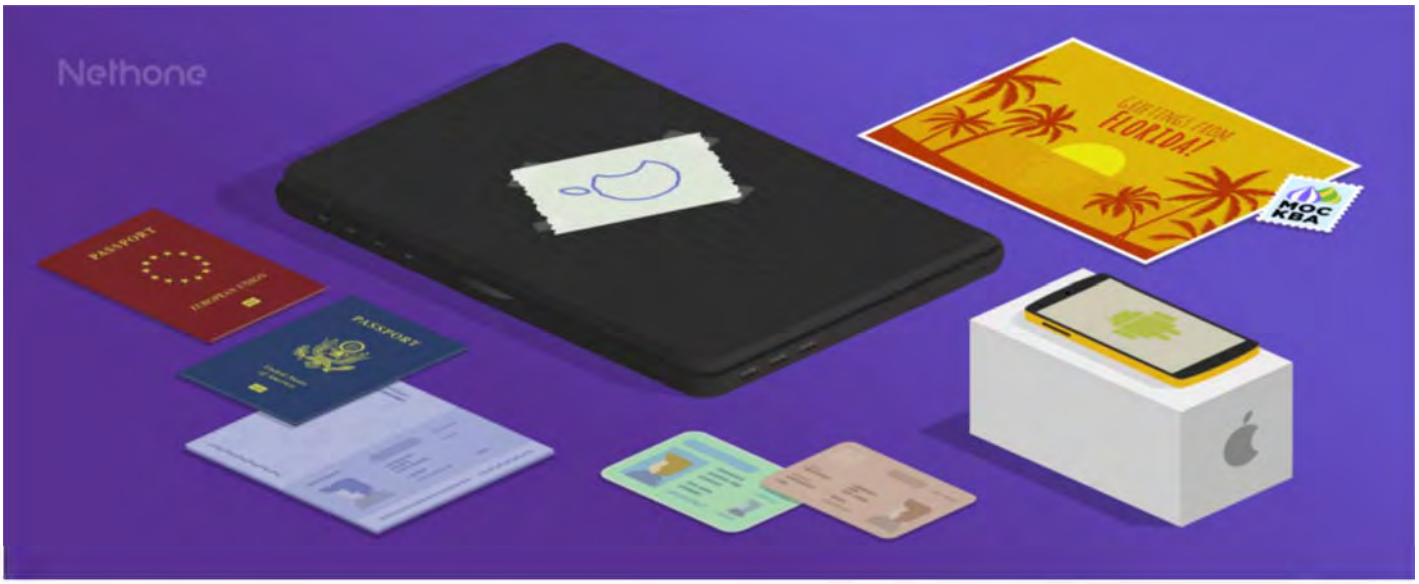
Nethone

Sign up

A Beginner's Guide to Machine Learning in Payment Fraud Detection & Prevention

Aleksander Kijek, Chief Product Officer . Aug 18 2017

While attending numerous conferences dedicated to payments and fraud detection, like ATPS or MRC, one can easily notice that Machine Learning (ML) is on everyone's lips these days. However, the ever increasing popularity of the topic is followed by more and more myths and rumors emerging and proliferating. In order to make the matter as clear as possible we have prepared this short guide to help you get started.



Nowadays, Machine Learning is being applied in nearly all areas of business: customer churn prediction, credit scoring, offer recommendation (e.g. Amazon or Netflix) and more. Machines car pilot an aircraft, drive a car, read texts and recognize their sentiment, and even write short novels or compose music. They have already beaten humans in one of the popular multiplayer cooperative games – DOTA2.

This technology has also proven to be extremely effective when it comes to fighting fraud.

….but what exactly Machine Learning is in the context of detecting fraudulent activities?

Simply put:

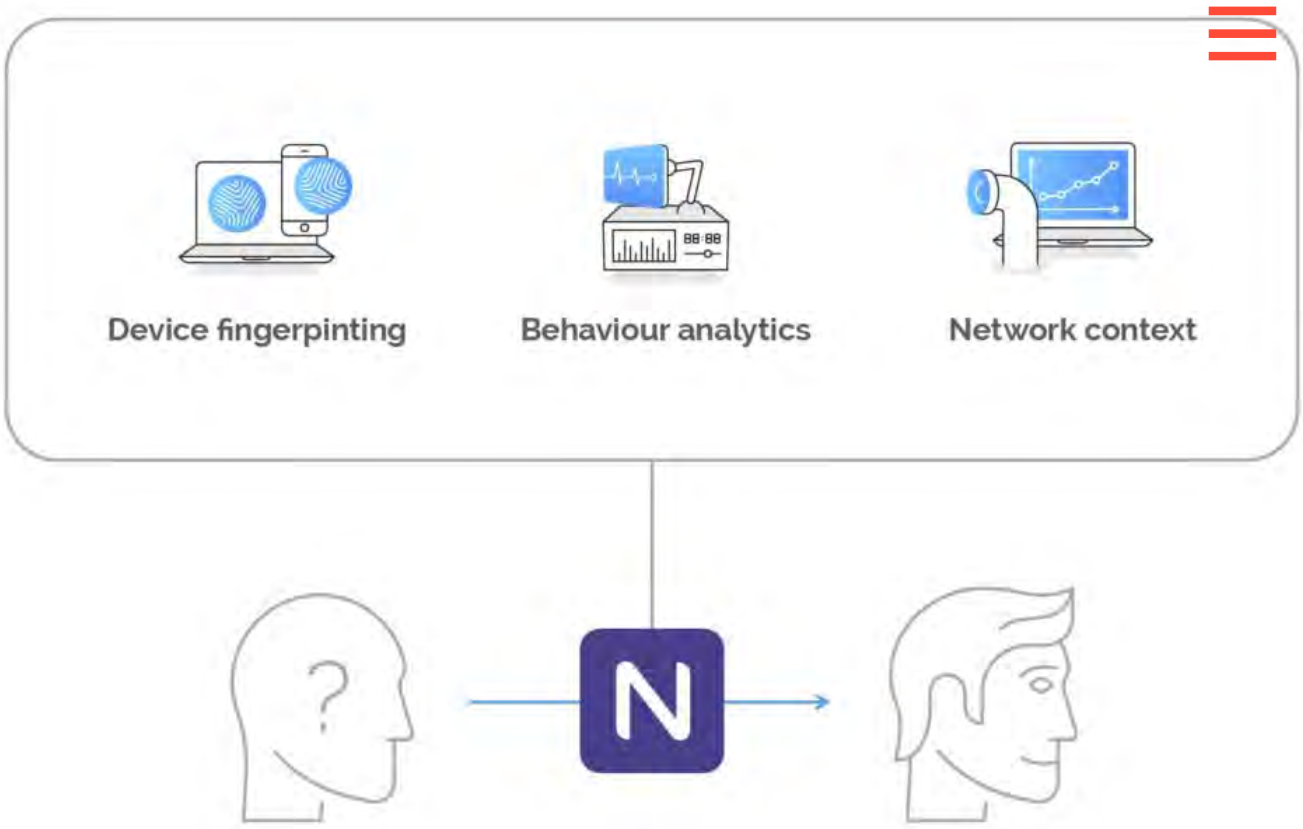
Machine Learning is a subfield of computer science that allows the machine to learn to tell fraudsters from legitimate users without explicitly telling it what designates to look for.

Let's dive deeper…

The idea is that there are certain characteristics of fraudulent transactions that differentiate them from legitimate ones. Machine Learning algorithms recognize patterns in the data that allow ther to discern fraudsters from legitimate clients, based on thousands of pieces of information, that sometimes may seem completely unrelated to a human being. The algorithm is searching for patterns in fraudsters' behavior, their hardware characteristics etc.

Applying Machine Learning to business

Whenever a customer carries out a transaction – the Machine Learning model thoroughly x-rays their profile searching for suspicious patterns.



Depending on the severity of the discovered "fraud-like" patterns, such a transaction can be accepted, blocked or handed over for a manual review. Everything is done in milliseconds.

What makes Machine Learning so special, is that it allows spotting fraudulent transactions wit a very high accuracy. Take Almundo.com case. This popular Online Travel Agent from Latin America has reduced fraud, chargebacks and manual reviews by 70%, thanks to Machine Learning.

Such a reduction leads to better customer experience (less false positives), optimisation of operational costs and significant increase in revenue.

Machine Learning is not aimed at replacing risk managers – it provides them with a more powerful tool to do their job!

Why does Machine Learning matter?

fraud detection strategy. I've selected some that I find most important.

Online fraud has become more sophisticated due to the rapid advances in the technology available to fraudsters. **Therefore to stay one step ahead of them, companies need to analyze much more data to successfully detect fraudulent attempts**. However, a skillful analyst can embrace, say, up to 10-20 pieces of information. While, **Machine Learning allows to analyze thousands of features**, and it will do it in a blink of an eye.

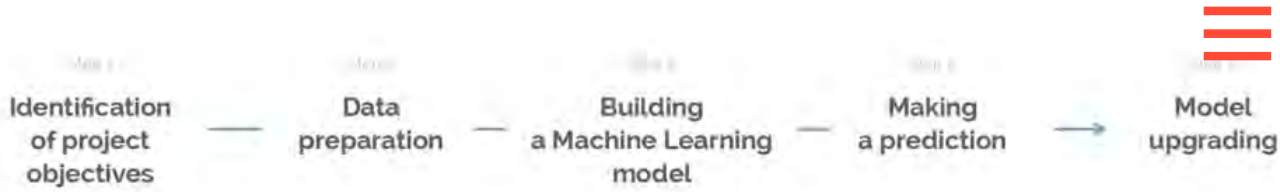
The traditional approach to fraud detection, using static rules-based systems (also known as production or expert systems) has its disadvantages which make it less effective:

1. There is a lag between identifying a need for a new rule and its implementation – machines will do it almost instantly.
2. Static rules-based systems are heavily dependent on human labor, which is expensive. Especially, if a given merchant is expanding to new markets as it implies the need for hiring more risk analysts due to this market specific patterns that must be analyzed.
3. Rules are created by humans who use their experience, knowledge and analytical skills. However, because fraud attacks have become more sophisticated, rules have also become more complex and error prone. That leads to more money loss and increase in false positives.
4. Rules systems grow to uncontrollable sizes, each new detected fraud scheme turns into a rule. After a while, the merchant is left with 150 rules the impact of which is hard to analyze over time. With Machine Learning you are able to faster verify its performance and adjust to the changing reality more quickly.

And last but not least, Machine Learning allows to clearly devise a business strategy based on KPIs and generated predictions of fraud attempts. It is possible to foresee the levels of refusal, acceptance or manual review to maximize the revenue. It means, for instance, that you are able t understand at what level of refusals, how many of fraud transactions will be caught.

How to predict fraud with Machine Learning

For the purpose of this blog post, I'm presenting a simplified version of Machine Learning proces



Step 1: Identification of project objectives

First of all, you need to determine your business objectives. Your goals may include, for instance

1. Minimizing the estimated chargeback ratio.
2. Minimizing the false positive rate („false alerts").
3. Keeping the manual review ratio (operating costs) at a controlled level.
4. Defining clients segments that generate most of the revenue, etc.

Here are some common questions that need to be answered during Step 1:

1. What is your company's need?
2. What are the main KPIs?
3. What are the revenue sources and the biggest revenue blockers?
4. What are the project's success criteria?

...and more.

On a technical level, **our main goal is to predict whether a given transaction is a part of the revenue or a fraud attempt.**

Step 2: Data preparation

Imagine that you want to learn a new skill. What do you do? You look for educational information Read books, guides, various articles, ask questions on forums, talk to professionals in this area et

previous purchases, geolocalization information, chargebacks report etc.

This raw data should be then cleaned and prepared into the form understandable for machines. may take some time (usually it is 60% – 80% of the whole Machine Learning process) and require certain technical skills. So it is advisable to build such competency inside your company or outsource it to an external vendor.

The result of Step 2 is a source dataset that will be used for further analysis (see Step 3). Below you will find a simplified example of what one can receive as a result of data preparation. Please keep in mind, that in practice, such a dataset may include hundreds or thousands of columns and even millions of rows.

Transaction ID	Order value	Currency	Card type	Products Quantity	Shipping address - City	Date of transaction	IP Adress	Target
7892151	702.9	DOL	VISA	5	GLASGOW	9/22/2016	22.234.996.087	0
7398210	54	DOL	DINERS	1	EDINBURGH	9/23/2016	09.091.662.125	0
5973254	122. 09	EUR	MASTERCARD	2	PARIS	2/18/2017	12.067.267.145	1
4402178	110	DOL	NULL	3	HASTINGS	3/5/2017	88.092.528.026	0
7398234	4.9	PLN	DINERS	1	WARSAW	5/23/2016	09.190.672.146	1
936826	10.2	DOL	VISA	2	LONDON	5/30/2016	98.382.037.227	0
4729113	100.21	DOL	VISA	5	LONDON	9/1/2016	12.826.010.371	0
6810093	89.99	EUR	MASTERCARD	2	LONDON	7/24/2017	95.361.020.157	1
2718325	16	DOL	NULL	1	CAMBRIDGE	6/15/2016	55.936.016.361	0
[illegible]	50.45	PLN	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

As you can see, in our example, each transaction (row) is described by a set of features (columns The last column is called **the target**. It indicates whether a particular transaction turned out to be a fraud or not. It is not important how you will mark a fraud in your data, it's up to you. The target can take a value of "1", "F", "Fraud" etc. It is not important which transactions your business considers as fraudulent — machine learning algorithms will look for patterns that discern the "1" class from "0". However, it's worth noting that the accuracy of the algorithm depends on the quality of the "Target" column. Of course, the strength of ML comes also in a possibility of identifying more categories e.g. – good customer, a regular customer, fraudster.

Step 3: Building a Machine Learning model

This is what the whole ML process is about, its final product. Once provided with information about a new transaction, the model will generate a recommendation stating whether you are dealing with a fraud attempt or not.

During the process of building such a model, one takes the dataset from the Step 2 to find out what characterizes marked fraudulent transactions and what the best predictors of fraud are. As there might be hundreds of features describing transactions, customers and their behavior, analyzing and drawing a meaningful conclusion is not a trivial task.

This process requires proper technology and Data Scientists with domain knowledge to know how to combine different kinds of data, which modeling technique will be most suitable for the particular business case and data, what will be the best set of the model parameters and more.

Step 4. Making a prediction

Ok, so we have a Machine Learning model....now what?

Make it work for your business! The model should be now deployed and integrated with your IT infrastructure.

Every time a customer buys a product/service in your e-store, the data about this transaction wil be sent to the model. **The model will generate a recommendation based on which your transaction system will make a decision about approving, blocking it or marking for manual review.**

This process is called **data scoring**.

But that's not the end. During a manual review, if a fraud detection team member marks the suspicious transaction as a legitimate one (false positive)**, Machine Learning model will take this information into account to make a better, more accurate decision next time.**

Step 5. Model upgrading

Models working in the production environment are under instantaneous feedback loop with new chargebacks and are constantly retrained to be able to detect new emerging fraudulent patterns. Just like in real life, humans without learning stimulus degrade their intellectual capabilities, same goes for models.

successfully detect fraud. For instance, detailed device features (e.g. GPU capabilities, processin

power, connection type, use of a virtual machine or a VPN connection) can bring a lot of new insights about the consumer and increase the accuracy of prediction.

It is recommended to look for new sources of information or use one of the available anti-fraud systems, which gather even 3000 data points and analyse them in order to create more precise and detailed fraudsters' profiles.

From now go to...

I hope you have enjoyed this guide. If you would like to learn more about Machine Learning, I recommend the Visual intro to machine learning from R2D3, which will give you more insights or the topic. Also, check out the article from Harvard Business Review to learn what ML can and cannot do for your organization.

And don't forget to follow us on Twitter or Linkedin, where we share more knowledge of how to effectively fight fraud using collected data and Machine Learning.

...or just drop us a line at contact@nethone.com. We will gladly answer all your questions concerning the application of ML in your business.

Aleksander Kijek

Chief Product Officer

Aleksander is a highly-skilled programmer and a Linux enthusiast fascinated by FinTech and Neuroscience. Prior to joining Nethone, he developed his technical and soft skills as leader of PISAK project (an initiative stimulating the social inclusion of heavily disabled people through technology) and coordinator of multiple projects at American Jewish Joint Distribution Committee.
At Nethone, Aleksander is responsible for business and product development, workflow management and ensuring comprehensive operational excellence at the company.

Nethone

Follow us

LinkedIn Twitter Facebook

Contact

contact@nethone.com

+48 22 112 15 15

Nethone

Solutions

Features

Dashboard

Developers

Technology

About us

Subsidised Projects

Case studies

Large American Airline

Resources

Newsroom

Blog

Website Terms & Conditio

Privacy and cookie policy