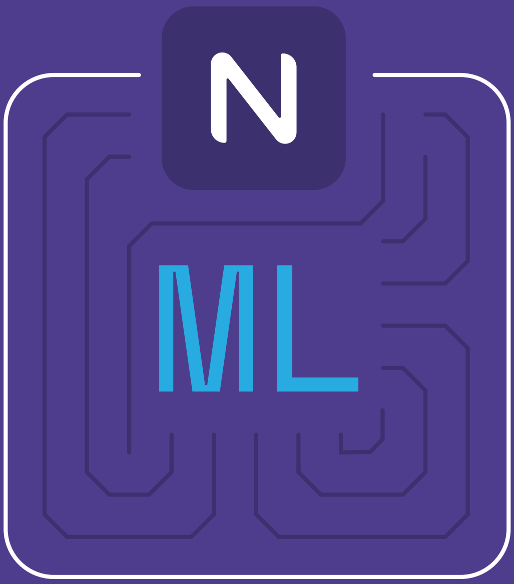
A Beginner's Guide to Machine Learning

in Payment Fraud Detection and Prevention



Nethone

Numerous conferences dedicated to payments and fraud detection, like MRC London 2019, show 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 rumours emerging and proliferating.

In order to make the matter as clear as possible, we have prepared this short guide to help you get started.

Table of Contents

01 Machine Learning for fraud detection – the definition

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 can pilot an aircraft, drive a car, read texts, and even write short novels or compose music. They have already beaten humans in one of the popular multiplayer cooperative games – DOTA2 (for now in a 1-on-1 scenario and we cannot wait until it will beat human teams in the 5-on-5 scenario).

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

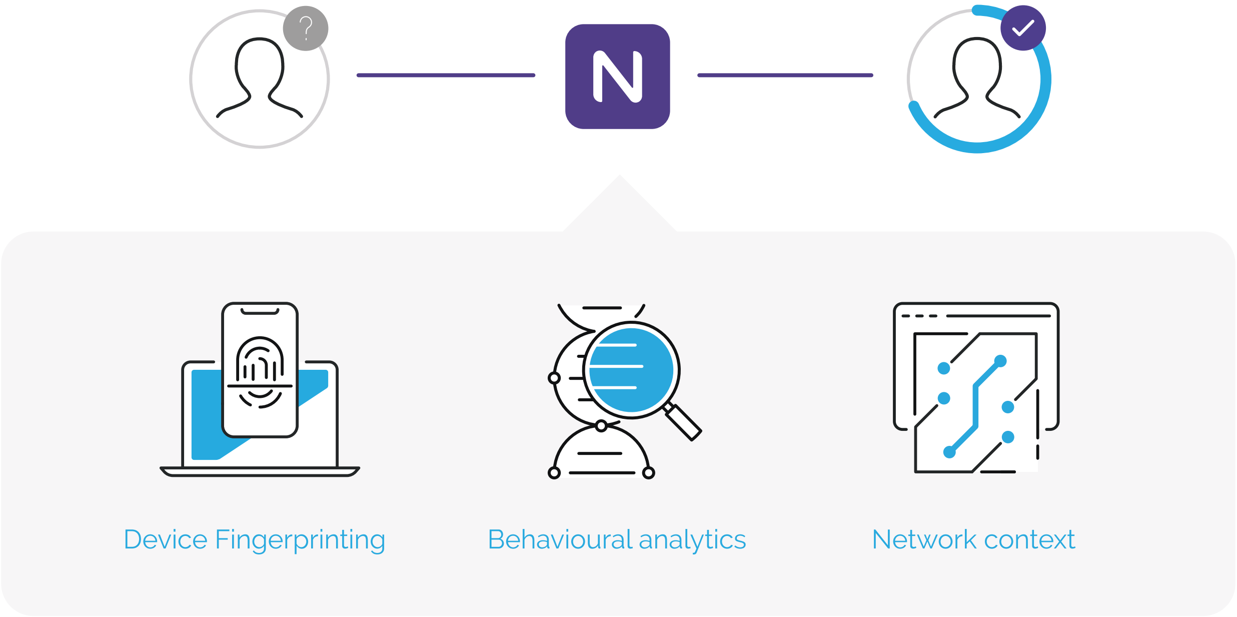
But what exactly Machine Learning is in the context of detecting fraudulent activities?

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.

The idea is that there are certain characteristics of fraudulent transactions that differentiate them from legitimate ones. Machine Learning algorithms recognise patterns in the data that allow them 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' behaviour, their hardware characteristics etc.

02 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 with very high accuracy. Take eSky case. This popular Online Travel Agency in Latin America has decreased manual review rate by 47%, increased the approval rate by 23%, and at the same time significantly decreased chargeback ratio. Such a reduction leads to better customer experience (less false positives), optimisation of operational costs, and a 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!

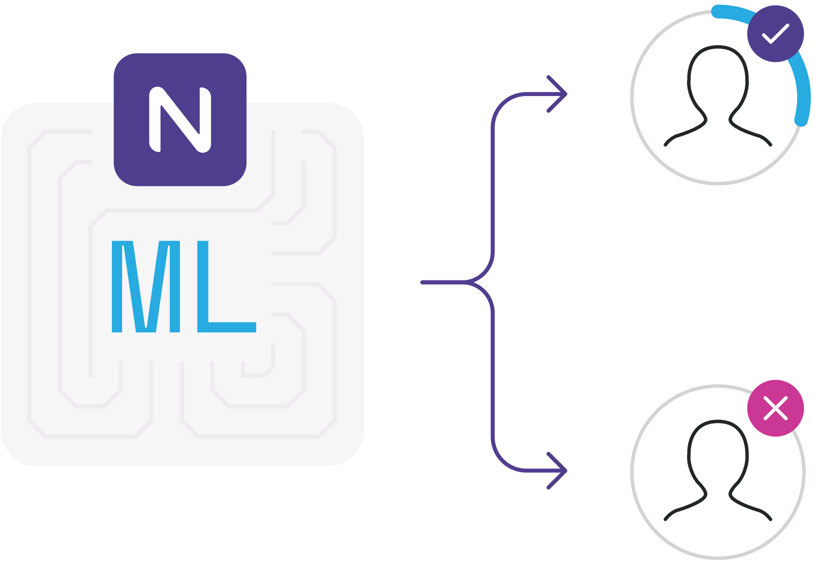
03 Why does Machine Learning matter?

There are few reasons why companies should consider including Machine Learning in their fraud detection strategy.

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.**

A skilful analyst can embrace up to 40 attributes of every user.

Machine Learning allows analyzing thousands of features.



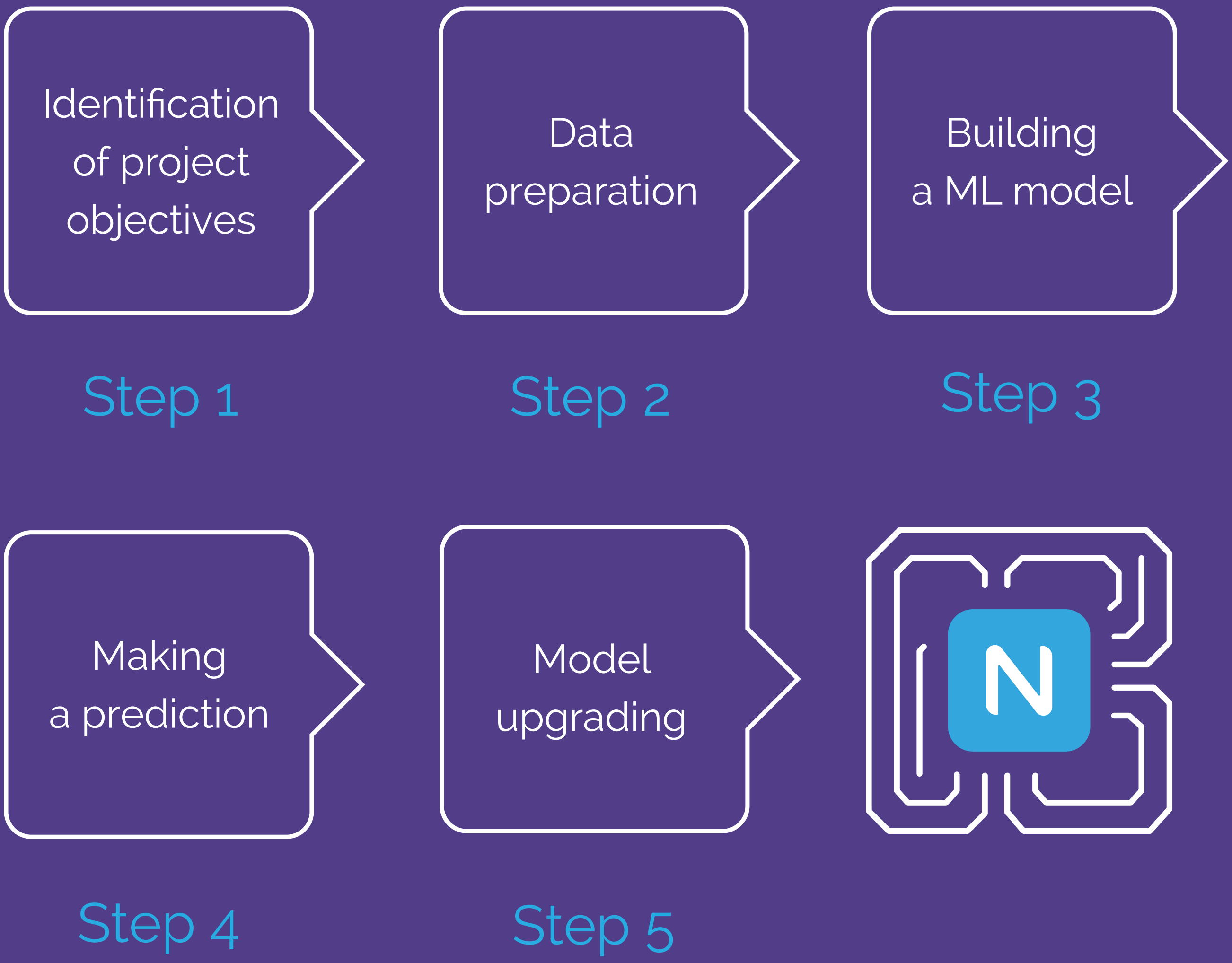
The traditional approach to fraud detection, using static rules-based systems (also known as production or expert systems), has its disadvantages which make it less effective:

1. **There is a lag** between identifying a need for a new rule and its implementation – contrary to machines that do it almost immediately.
2. **Static rule-based systems are heavily dependent** on human labour, 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 these market-specific patterns that must be analysed.
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 money loss and an increase in false positives.
4. **Rule 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 analyse over time. With Machine Learning you are able to verify its performance faster and adjust to the changing reality more quickly.

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 maximise the revenue. It means, for instance, that you are able to understand how many fraudulent transactions will be caught, at a particular trafic rejection level.

04 How to predict fraud with Machine Learning

For the purpose of this guide, we present a simplified version of Machine Learning process, to give you a general concept of what it is all about.



Identification of project objectives

Step 1

Data preparation

Step 2

Building a ML model

Step 3

Making a prediction

Step 4

Model upgrading

Step 5

Step 1: Identification of project objectives

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

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

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

- What is your company's need?
- What are the main KPIs?
- What are the revenue sources and the biggest revenue blockers?
- 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, etc.

The same refers to machines – **in order to create fraudsters' profiles, they need historical data about previous fraudulent events.** The more features and data collected by a company to analyse, the better. It could be time, frequency or value of the transaction, the history of the previous purchases, geolocalization information, chargebacks report and many other data resources.

This raw data should be then cleaned and prepared into the form that is understandable for machines. It takes some time (usually it is 60% – 80% of the whole Machine Learning process) and requires certain technical skills. 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	Currency	Products quantity	Shipping address - city	Date of transaction	IP address	Target
7892151	702.9	USD	VISA	5	GLASGOW	9/22/2018	22.234.996.087	0
7398210	54	USD	DINERS	1	EDINBURGH	9/22/2018	09.091.662.125	0
5973254	122.09	EUR	MASTERCARD	2	PARIS	2/18/2019	12.067.267.145	1
4402178	110	USD	NULL	3	HASTINGS	3/5/2019	88.092.528.026	0
7398234	4.9	PLN	DINERS	1	WARSAW	5/23/2018	09.190.672.146	1
936826	10.2	USD	VISA	2	LONDON	5/30/2018	98.382.037.227	0
4729113	100.21	USD	VISA	5	LONDON	9/1/2018	12.826.010.371	0
6810093	89.99	EUR	MASTERCARD	2	LONDON	7/24/2019	93.361.020.157	1
2718325	16	USD	NULL	1	CAMBRIDGE	6/15/2018	55.936.016.361	0

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 the possibility of identifying more categories e.g. – a good customer, a regular customer, a 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 Step 2 to find out what are the characteristics of marked fraudulent transactions and what are the best predictors of fraud. As there might be hundreds of features describing transactions, customers and their behaviour, analysing 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 modelling technique will be most suitable for the particular business case and data, what will be the best set of model parameters and more.

Step 4: Making a prediction

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

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 will 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.

As mentioned before, fraud attacks are getting more sophisticated, therefore one needs more data to successfully detect fraud. For instance, detailed device features (e.g. GPU capabilities, processing 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 5,000 attributes and analyse them in order to create more precise and detailed fraudsters' profiles.

Nethone has developed proprietary AI-driven Know Your Users analytical system that helps convert threats and challenges into well-informed, profitable decisions and protect travel, digital lending, and e-commerce industries from fraud.

Reach out to

contact@nethone.com

to learn how Nethone can provide for you a comprehensive Know Your Users analytics.