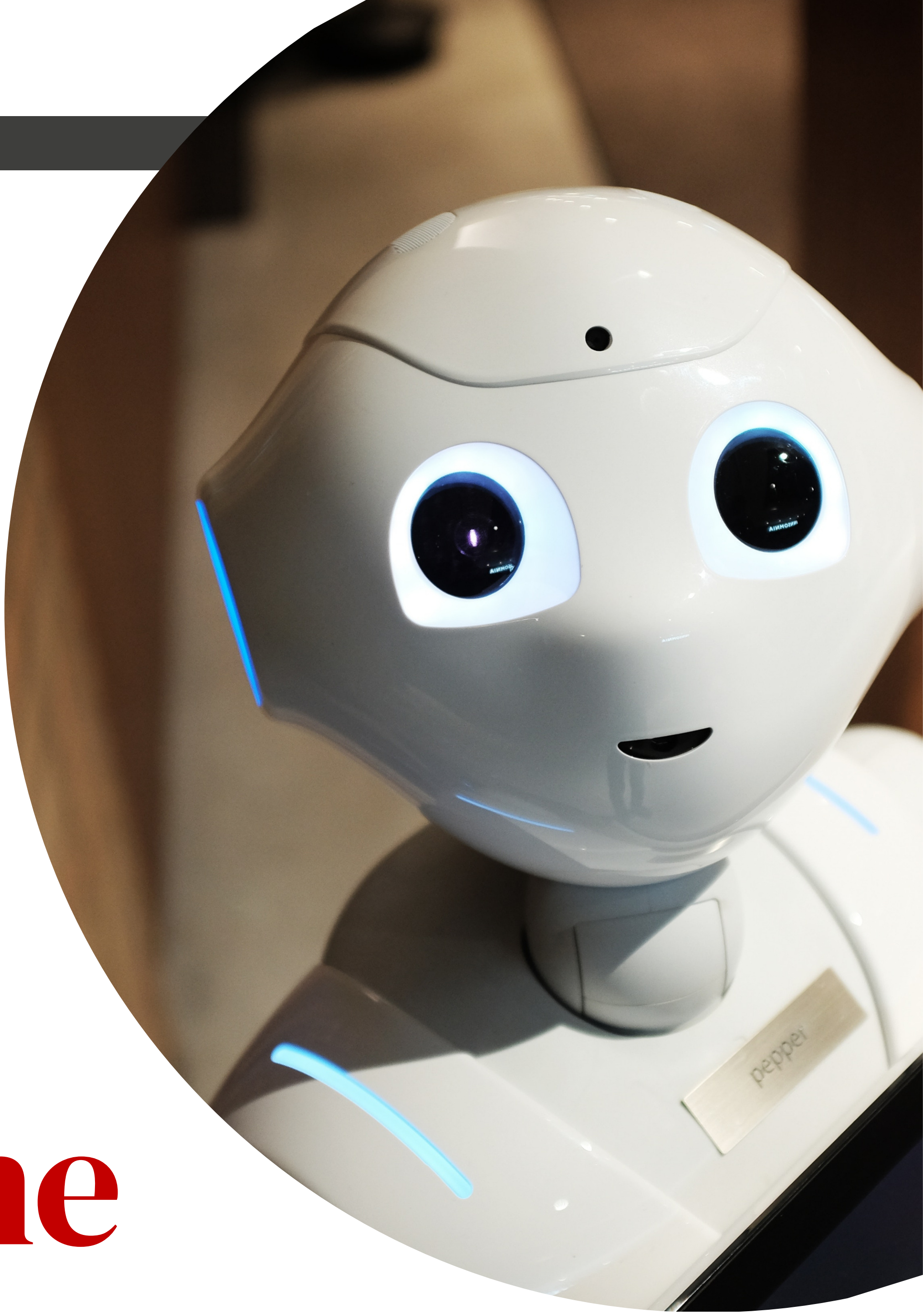


Machine Learning:

Answering Your Most Burning Questions

enuda

April 2022

Table of contents

enuda

WHY ML?

If we learned anything from the previous Industrial Revolutions, it’s that technology inspires innovation.

It empowers us to re-discover the world around us from a whole new perspective.

The technological evolution, together with the fascinating human curiosity, have given us... machine learning.

Machine learning is here today, and it's shaping and simplifying the way we live, work, travel, and communicate.

Simply put, machine learning looks for patterns in data and tries to draw conclusions. ML is not explicitly programmed by people, but the algorithm learns from examples. This is good for us, humans, because it is a lot easier to provide examples of what we want than to write explicit code.

WHY ML?

Machine learning is not just about advanced algorithms.

It’s about breaking down barriers.

Because the truth is... The world will not get less complicated. So, the more complicated your business gets, the more complex the data gets. Humans alone can't keep up with that, but a machine learning model thrives in that scenario.

ML gives us the ability to understand our environment much more intimately than we ever could before.

WHY ML?

Think about Predictive Maintenance, for example. What if your machine could do it on its own by analysing bigger and more complex data? What if you could just watch your machine learn when and how to take action to prevent errors? Today, it's completely possible, thanks to the advanced algorithms we can create.

And the best thing about it? It doesn't have to be overwhelming. Start small, and build along the way. It's a marathon, not a sprint.

Don't know WHERE to start? That's exactly why you're here.

OK... WHAT EVEN IS IT?

Let’s try to explain what machine learning is in simpler terms.

Basically, the simplest definition we can use is that machine learning is learning from data.

What does that mean?

That means that we, as engineers, don't have to explicitly program some specific model for the system or the process we are working on, and we don't even have to know the model or the functioning of that system.

Instead, our machine learning model will take the past data points, and, from those learn, on its own how our system works.

This way, in some future data points, machine learning will enable our system to automatically find some features, connections, patterns in the data. It will allow us to analyse and predict some upcoming states and situations.

WHAT PROBLEMS CAN IT SOLVE?

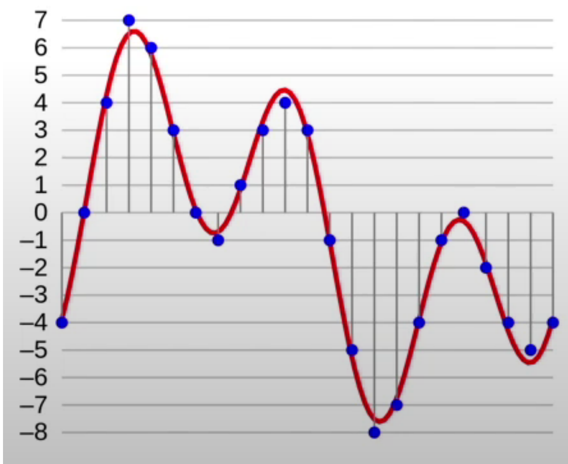
1. Forecasting

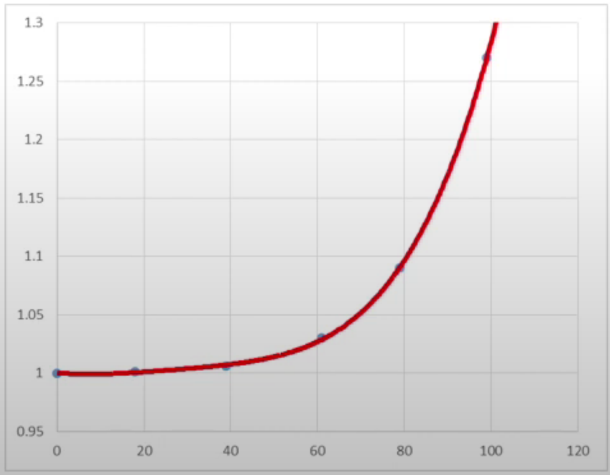
The first type of problems we can solve is forecasting or predicting the value of some specific parameter.

For example, we want to find the temperature at a specific time in the future.

Or vice versa, we want to know:

- what is the trend of our temperature change?
- when will the temperature reach a specific value, at what data point in the future?

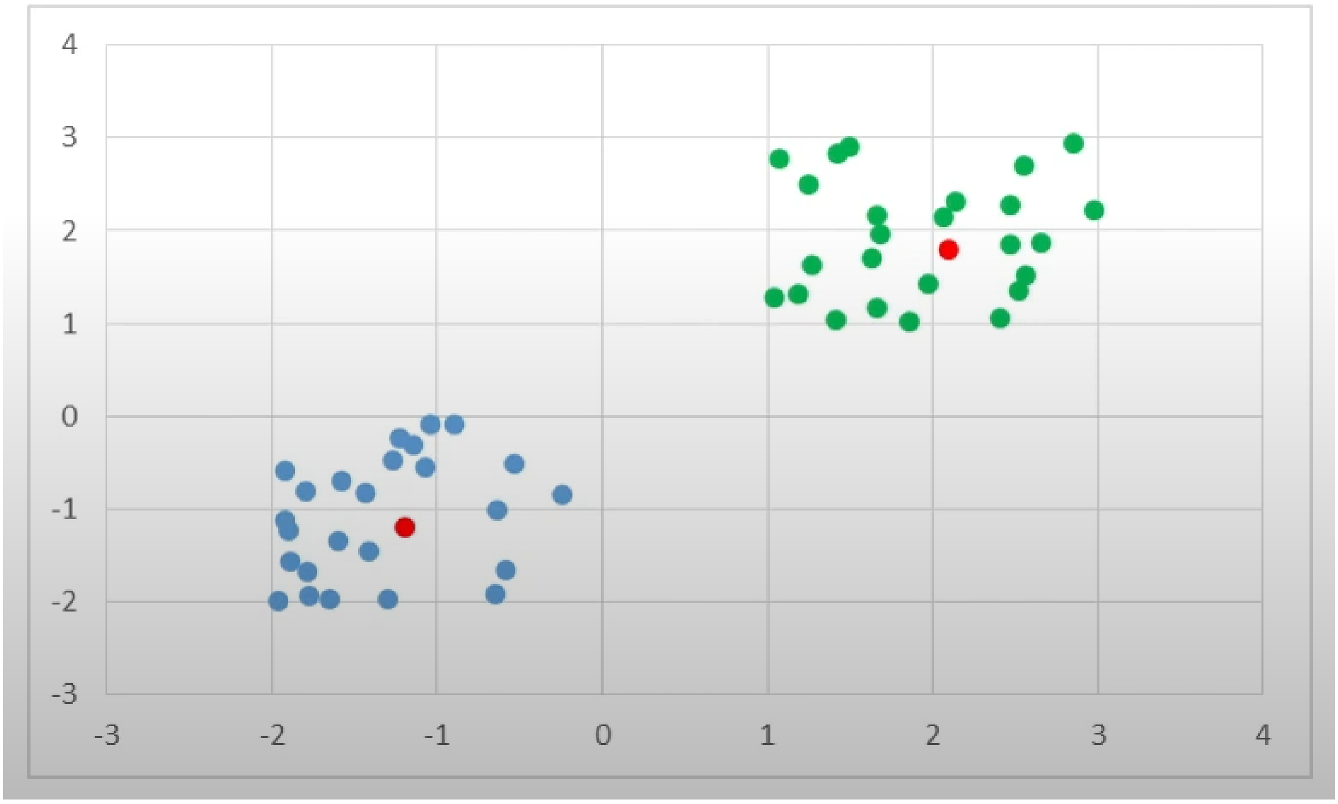




2. Classifying events or situations.

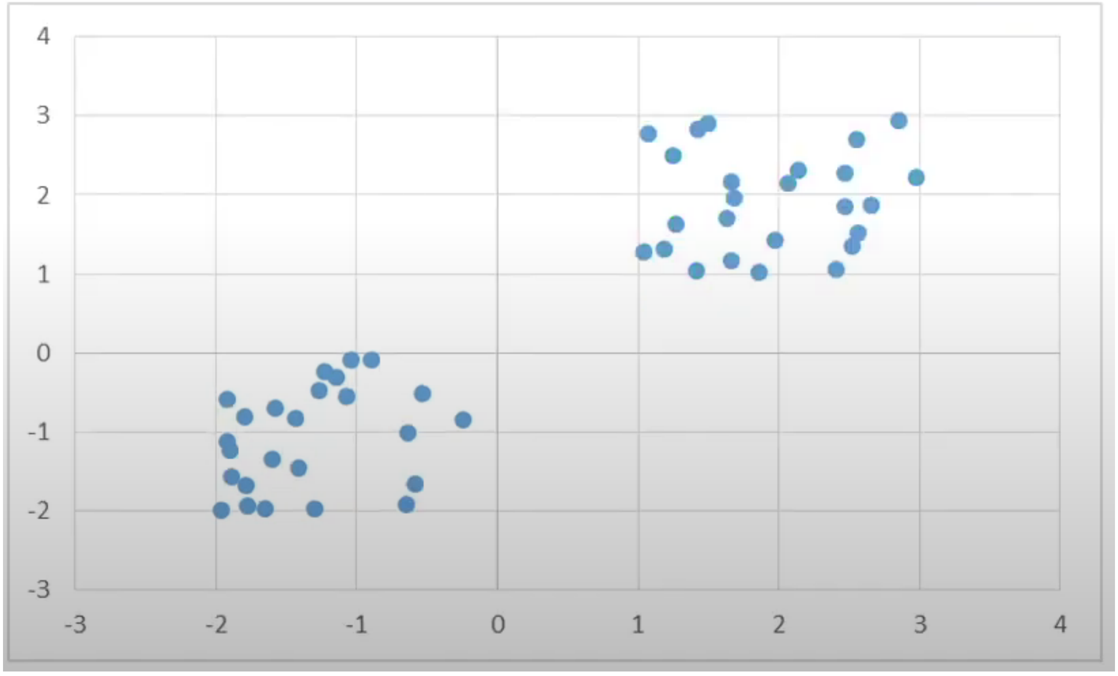
For example, we want to know if our system is functioning normally, or if it is in a faulty state.

We want to classify these events - "Yes, this is a normal state" or "No, this is a problem we need to solve".

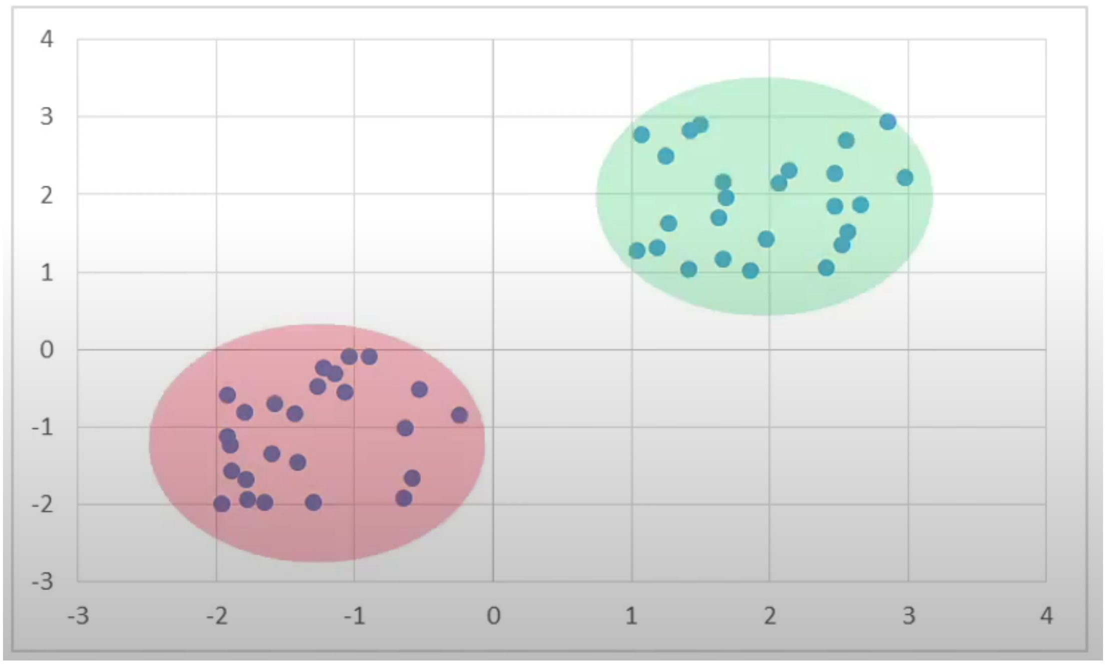


WHAT PROBLEMS CAN IT SOLVE?

3. Finding patterns in the data



For example, we want to detect similar events and group them based on some features.

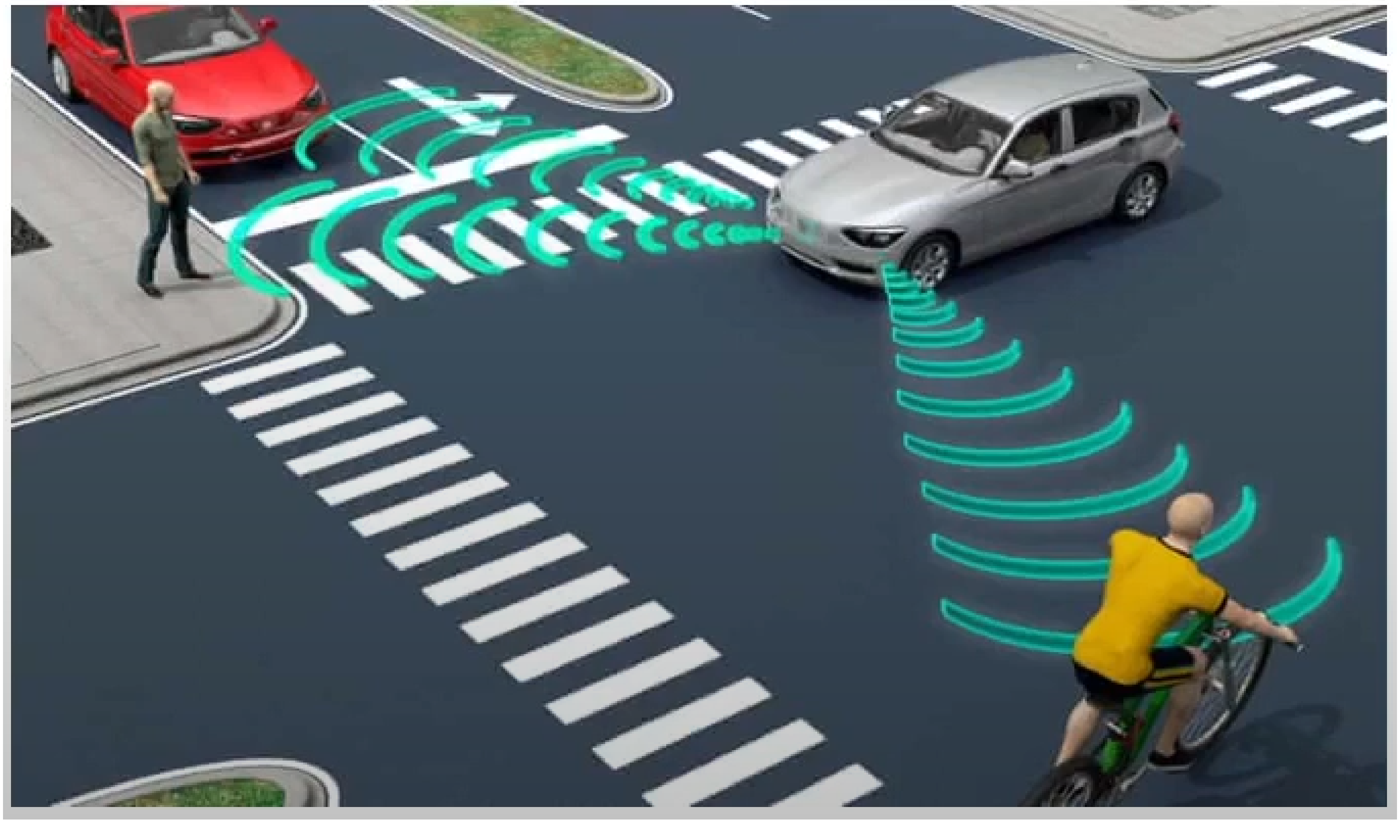


WHAT PROBLEMS CAN IT SOLVE?

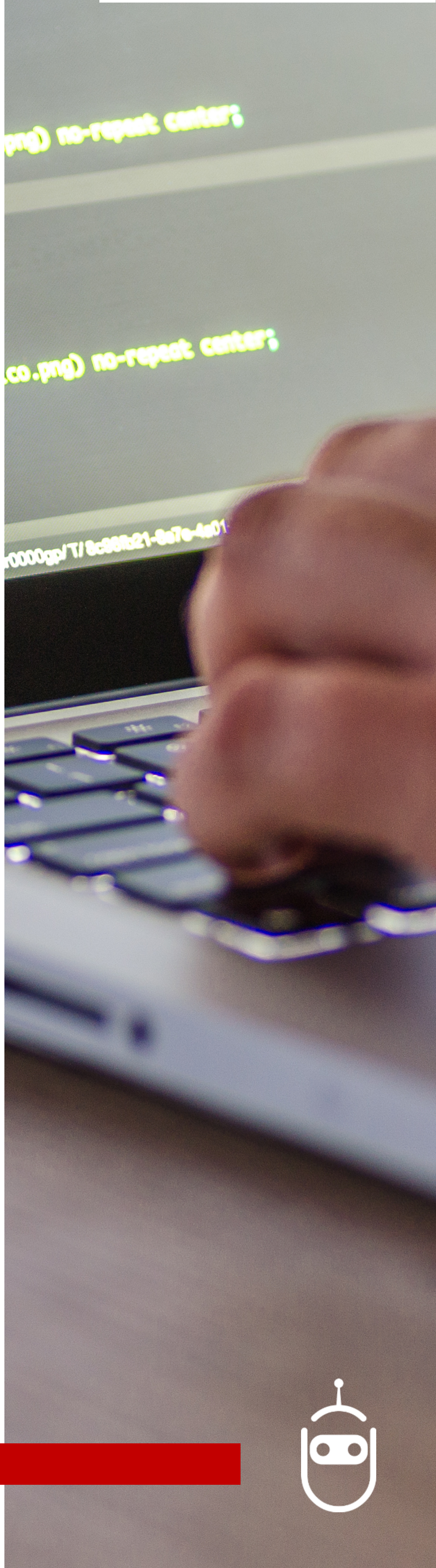
4. Reach the best solution

For example, here we have an automatic vehicle that is finding its way through the environment.

Or it can be searching for the best route to get from point A to point B, and other similar actions.



WHAT ARE THE DIFFERENT TYPES OF ML MODELS?



When talking about machine learning, it's very important to understand what are the different types of machine learning models that we can use for our problems.

Those are:

- Supervised learning
- Unsupervised learning
- Semi-supervised learning
- Reinforcement learning

The type of model you use will depend on your requirements and the data you have available (its quality, the amount etc).

SUPERVISED LEARNING

The first type of machine learning is supervised learning, and it's definitely the one that is most often used.

It's a type of machine learning where our model learns from examples with some known outputs. That means that, for every data sample that we give to our model, **we will also give it the output or the desired value or category that we want our model to predict for those inputs.**

So, for each sample, our model will know what is expected of it to get in the end and it will adjust its parameters and its features so that it can reach the desired output as well as possible from the given inputs.

This way, in the future, when we give some new samples or some new data examples for the inputs, it will be able to predict the outputs that are as good as possible for the given model.

UNSUPERVISED LEARNING

In unsupervised learning, our model learns from examples without known output. **That means that we give our model examples, data samples, and our model doesn't know what we want it to get.**

These models are usually used in different types of applications. In some applications, we need to find similarities between the data.

Let's say, for example, that we have a group of people and we want to divide them into two categories. We can divide them based on different things.

For example, the tall ones and the short ones, older ones and the children, those who are standing and those who are sitting in a chair, and so on.

In fact, our unsupervised learning model will find the best division there is for our data set, so that it can easily put every sample in one group or the other. It'll find the best division line between the two groups so that it can, in the future, with new samples, easily fit them in one group or the other.

SEMI-SUPERVISED LEARNING

Semi supervised learning is kind of trying to take the best of the supervised learning and unsupervised learning.

Our model can learn from examples **without known output**, which is like unsupervised learning, and which is actually easier to obtain. It's much easier to get the data without known outputs than the data with known outputs because the pool of data is larger in that regard.

But we also have some examples, like a smaller set of data, **which has known outputs.**

So, we use that in combination with different examples, a lot more examples, where we don't know the output.

And this way we can train our model.

REINFORCEMENT LEARNING

Reinforcement learning is different from the previous types of machine learning models. **It learns how to react to an environment.**

These models start with almost zero knowledge and they are put into the environment where they need to work.

Depending on the situations they encounter, they can get some awards or some punishments, so to say, based on the outcome that's the result of them trying to perform a certain task.

One of the best examples of reinforcement learning are autonomous vehicles. We can train them by putting them into traffic and then they learn what is the pedestrian, what is the traffic light, and so on.

They learn through getting awards if it avoids them, or getting punished if it hits them.

WHAT DO YOU NEED TO GET STARTED?

So far, so good, don't you think? It sounds pretty interesting. But what do you actually need to get started with it?

Look, we need to say it again before we dive into that: **it doesn't have to be complex and difficult.**

There's a common misconception that machine learning is a thing of the future, you couldn't possibly get started with it in your organisation right now. That's simply not true.

Sure, you do need a lot of data (although it depends, sometimes you need a few hundred data points, sometimes a million) to build a machine learning solution, but that's completely possible.

At the end of the day, technology is here to make our lives easier, not harder, so don't be afraid of the initiative to advance your systems.

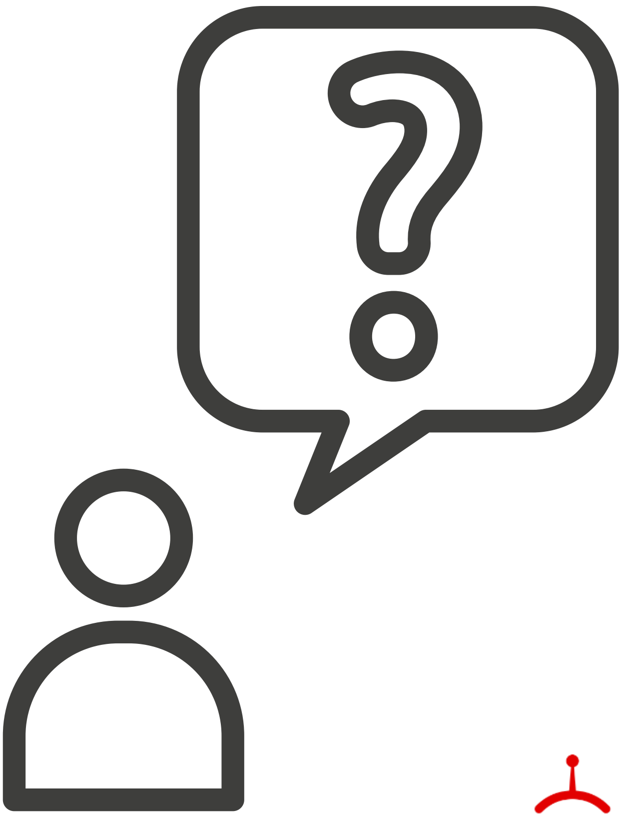
DEFINE THE PROBLEM

You need to know what you want to solve and for what problem you want to apply a machine learning model.

You also need to make sure that this problem that you have fits the problem types that can be solved with machine learning.

Not every type of problem in the modern world can be solved with machine learning. And as we often hear about machine learning everywhere, we want to apply it to anything that comes to our mind. That's not feasible.

Once we know the problem, we need to have enough good quality data.



QUALITY DATA

- Data samples that are collected over a representative time period.

What does that mean? Let's say we want to predict or forecast our clients demand for a certain product in the next year.

We need to take into account that **there are different time periods of the year and probably this demand won't be the same in January or in June, depending on our product.** So, we have to collect previous data for January and for June and for all the other months that we want to predict for the future.

Then, we need to have this data from all the key points of the system. It's not enough if we have the data just from one data point. If it is, for example, a distributed water network, we want to have the values measured in different points of the system that are key for solving our problem.

Then, we need enough data for each important outcome or scenario.

- Balanced data

Often we have a problem with huge amounts of data.

For example, we want to do anomaly detection or fault detection in a system.

For that system, we have very, very big amounts of good data when the system is functioning normally, and everything is in order, but only a few data points where the system was in the state of fault or error or some anomaly happened. **Therefore, we don't have enough of these data points that the model can learn that there was a problem.**

We want to have the balanced data for each scenario or each category that we want to detect.

Of course there are also models that can predict these anomalies, even if we don't have enough data, but it's always better to have enough data for each important outcome or for each category.

QUALITY DATA

- Adequately labeled data

This is especially important if we're doing supervised learning, where we need to know the label or the value or the outcome for each data sample or for each previous situation.

And if this data is not labeled correctly, our model will not be able to learn well because it might learn from the wrong label and it will learn perfectly but it will learn the wrong outcome.

So that is why it is very important to have good and quality labeled data.

HARDWARE

So, we have data, we know the problem, and we know machine learning can help.

We still need some adequate hardware for our solution.

For some simpler applications, maybe a standard computer or server that we already have will do the job.

But if we are implementing some more complex models and algorithms that are developing practically every day in the modern times, then we will need some faster processing units.

Powerful GPUs (Graphical Processing Units) or TPUs (Tensor Processing Units) enable massive parallel execution of the processing tasks and will be able to train our model and to predict in real time.

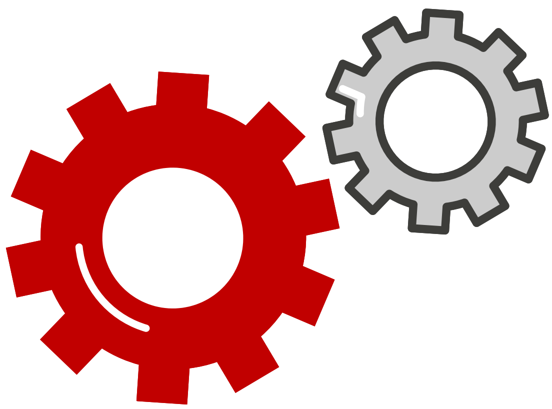
HARDWARE

Or if we don't have that, we can also consider using some of the cloud services that are widely available today. They can provide us with additional processing power when we need it.

For example, for training the model, and then for forecasting or predicting or detecting a situation, our standard computer server architecture, that we're already using, will be able to do it because that's often not so demanding as the training process for the model.

But when the time comes for the model to be trained, cloud services are a big help.

But let's dive deeper into that, and see what other factors we should take into consideration.



WHICH HARDWARE AND SOFTWARE TO CHOOSE?

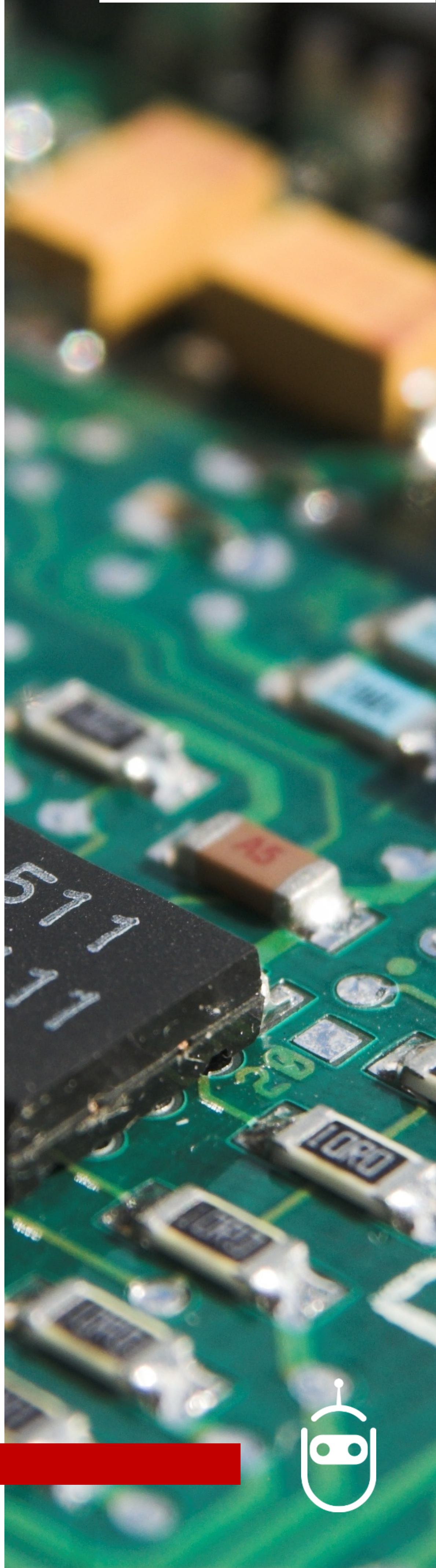
We want to choose machine learning in our solution, but we are not quite sure which hardware or software option to choose and which options exist in the first place.

So, if we are talking about hardware for machine learning, the first thing we should discuss is **whether we will use the local hardware or cloud services.**

The Local Hardware

The local hardware is the hardware we have at our premises in our local system.

The first big bonus of this option is that we don't need to leave our internal environment, we have everything and we control it completely.



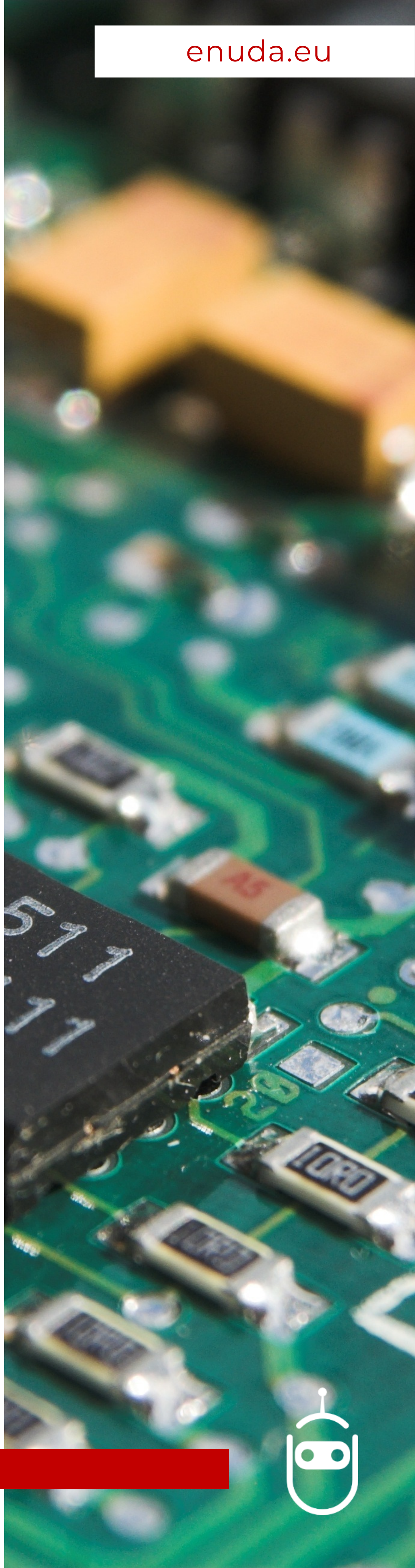
WHICH HARDWARE AND SOFTWARE TO CHOOSE?

It also doesn't depend on the connectivity or functioning of any other system, it's completely independent. And we can always count on it that it'll work or we are there and we can fix whatever problem arises.

The disadvantage of this approach is definitely the possible need for some expensive equipment, expensive processing units that we might need, depending on the complexity of the model that we are using.

And these processing units evolve very fast and might also become obsolete soon, therefore we might have to upgrade it, buy some more equipment and so on.

On the other hand, if we discuss cloud services as an option, they are basically inverse.



WHICH HARDWARE AND SOFTWARE TO CHOOSE?

The Cloud Services

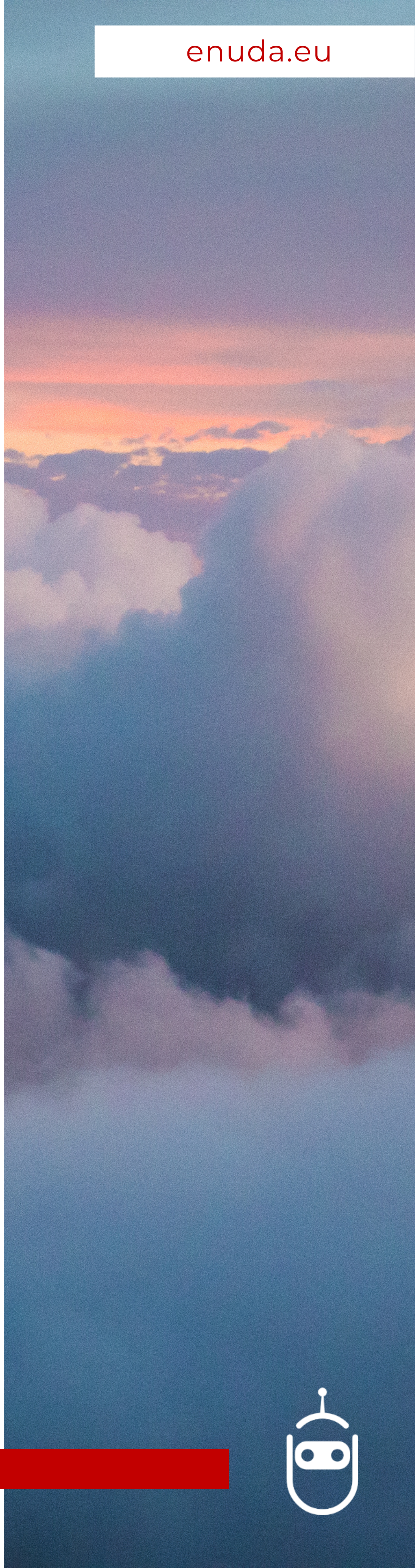
There's no need to buy expensive equipment and we always have access to the most recent hardware upon request.

We just go to the cloud service, and we request that we train a model.

We express our requirements for the hardware, and we can pay for that and use its processing power as a service.

The disadvantage of this approach, one that is a lot talked about, is the need to leave the internal environment that maybe we don't want to do.

We want to have everything in our local system and we don't want to depend on external services or environments.



OPEN VS. PROPRIETARY

The next thing to discuss is whether we use the open machine learning frameworks or proprietary solutions.

- Open frameworks

Open machine learning frameworks are free to use and that is definitely one of their biggest advantages, but they probably require more knowledge and skills, programming, and more hands-on coding that we need to perform ourselves.

A good thing about the open machine learning frameworks is that we can always follow the trends that exist in the machine learning community.

We can also change models, adjust them so that they fit completely the need that our system has.

- Proprietary solutions

Proprietary solutions can give us out of the box solutions for different models, but they definitely come for a price. Sure, they could help us develop our model faster because of the already done solution, but we should question whether or not it's worth it.

DATA PRIVACY

Data privacy is definitely an important aspect that we need to consider when choosing hardware and software, as data privacy characteristics are very closely related to the options that we choose.

It's something that is very much talked about today in both the scientific and the business community.

We value our data highly and we don't want to expose it or have any risks on the data privacy from our system.

So this can affect both the option between the local hardware or the cloud and the open framework or the proprietary solution.

Often because of data privacy, we want to have our system closed locally and not let any of our data leave the system no matter how secure the connection is. But then again, that comes with the limitations we've talked about above.

EXPLAINABILITY

Are we able to explain what our model did or not?

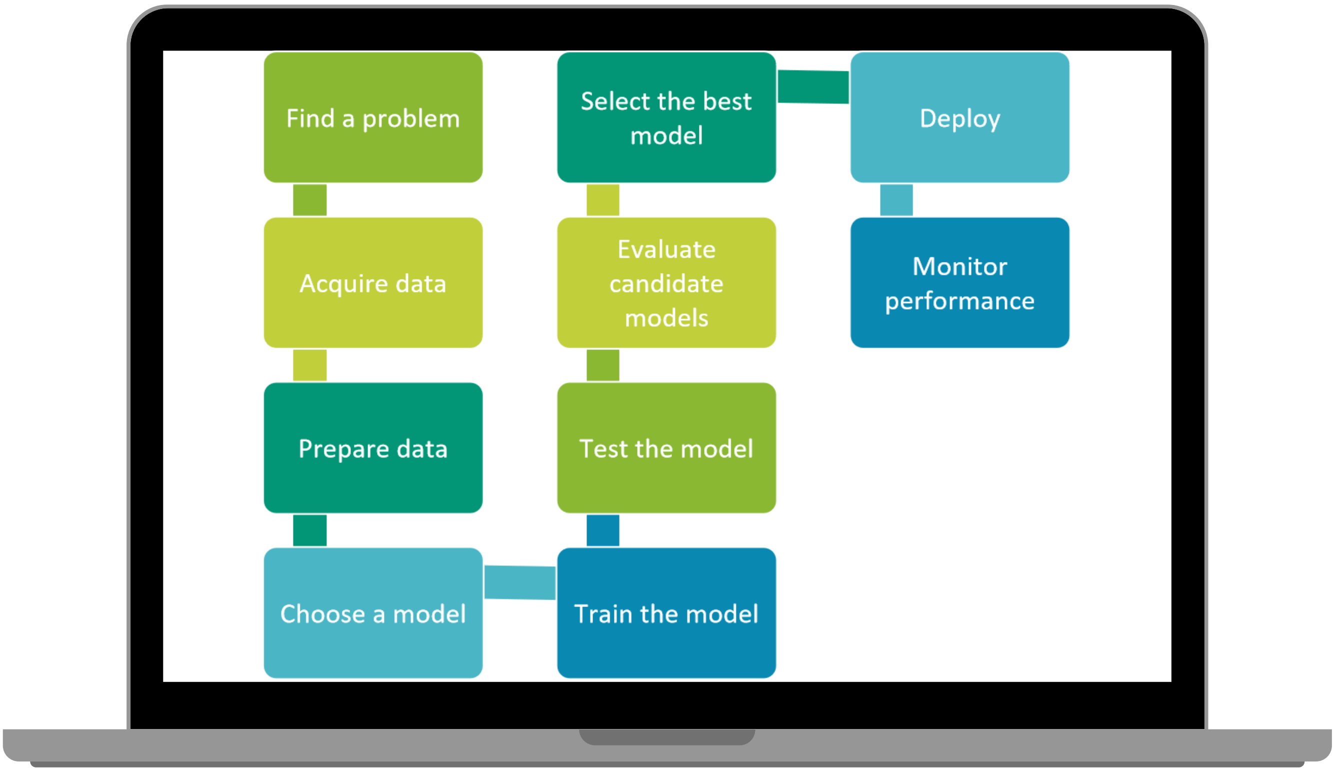
Do we know why our model made a certain decision or prediction or how it reached its output?

This is very important when choosing the exact model that we are going to use and whether we will program this solution from scratch or we will use, for example, a cloud service or a proprietary solution.

Sometimes we are able to use a model like a black box. We give it input, it gives us output and the outputs are good. The error rate is very small, and we are happy with that.

But for some situations, the risk is just too high. And we have to be able, for insurance reasons or other reasons, to be able to explain why our model made this decision or why it predicted this and that, and how it made us decide that we will take one path or the other.

WHAT'S A TYPICAL WORKFLOW?

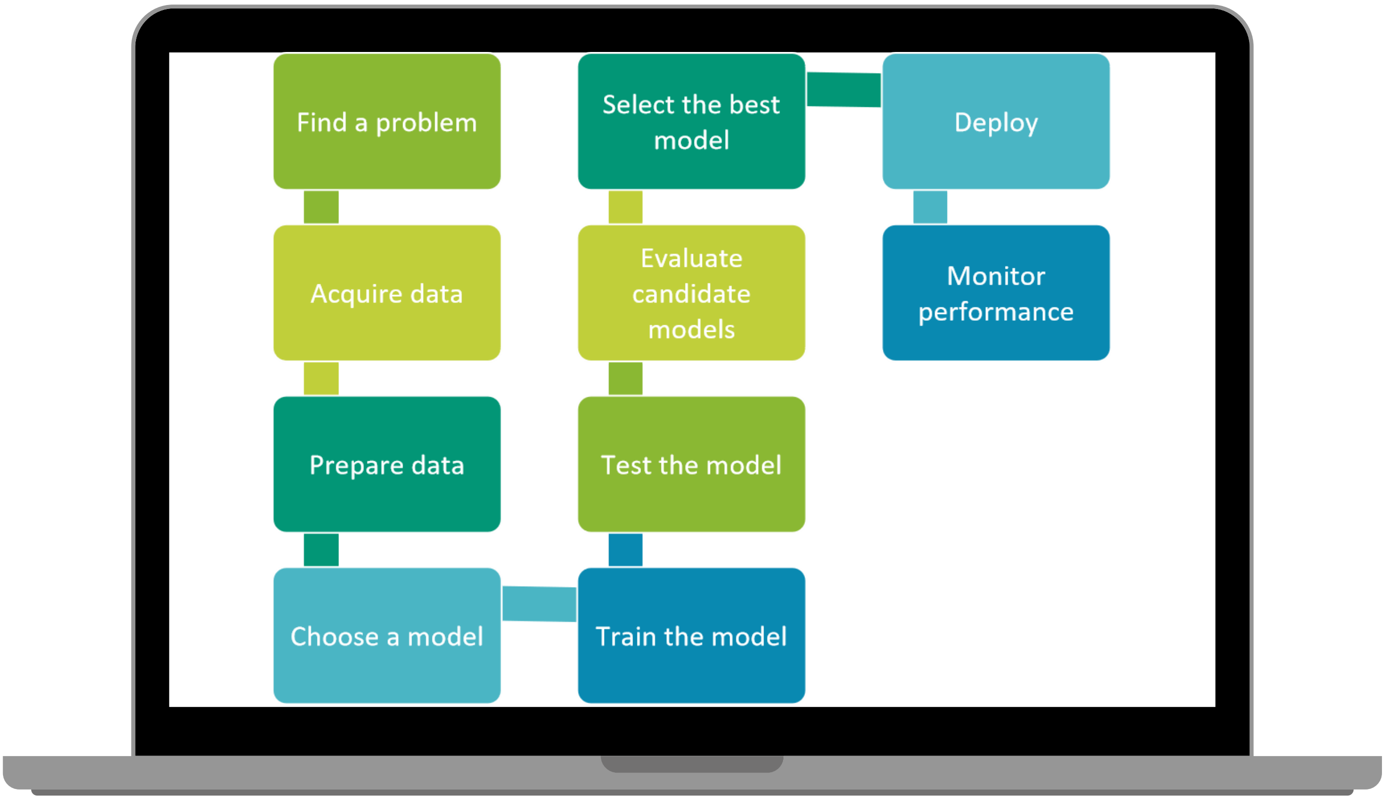


Usually, we start by finding a problem. As previously discussed, we need to make sure that this problem can actually be solved by a ML model.

The next step is to acquire data from the system. Usually, we need historical data, and we want to acquire enough quality data for our model.

Then we need to prepare this data because, often, there are some problems such as missing data or not every set of data from every part of the system is in the same format. Basically, we need to turn it into quality data.

WHAT'S A TYPICAL WORKFLOW?

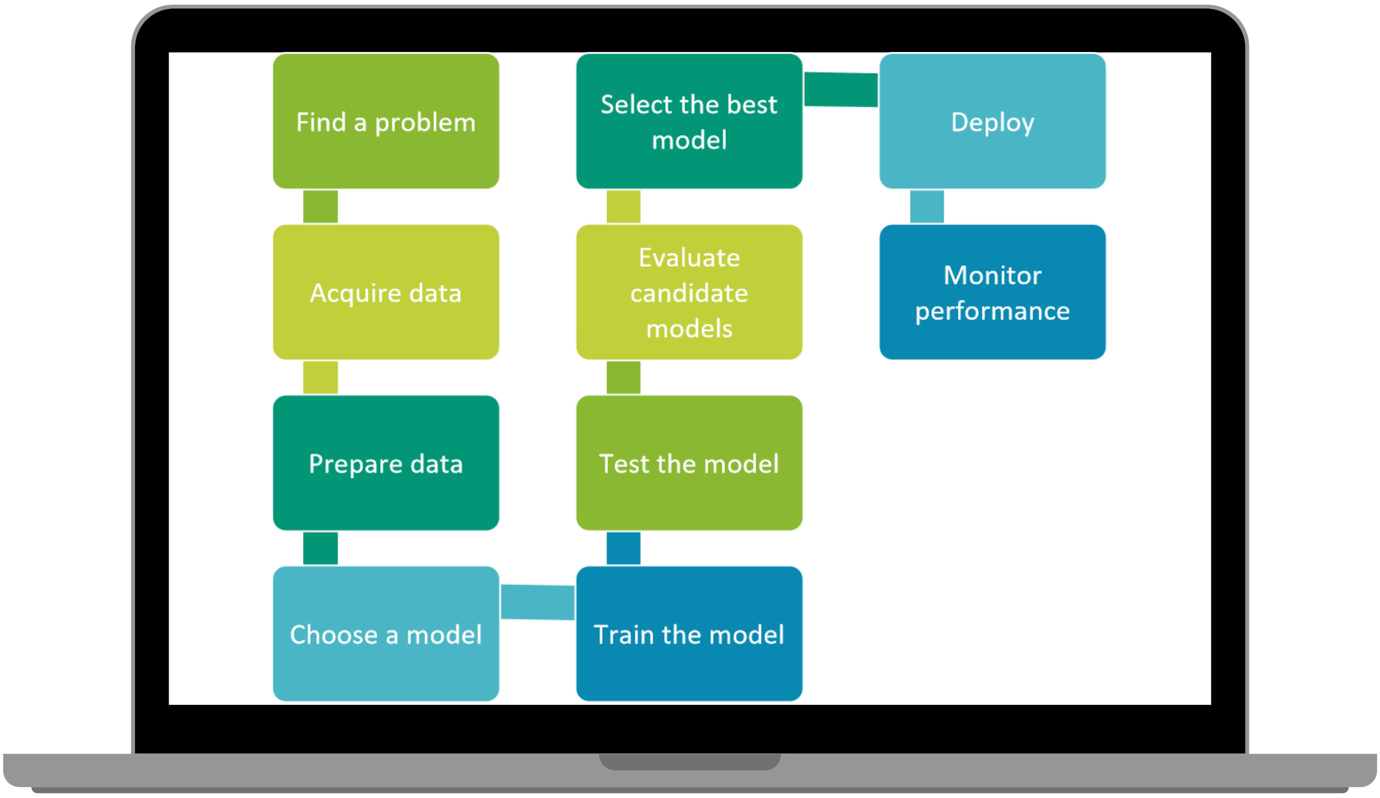


Next, we need to choose the model that we will apply and then train this model.

At this point, we can test the model. Usually, we will split our data into two or three data sets: the training data set, the validation data set, and the test data set.

For fun, we can always reiterate the steps of choosing the model, training it, and testing it. That way, we can choose different models and try to train and test each one, and see what results we can get with different machine learning models.

WHAT'S A TYPICAL WORKFLOW?



When we have tried several of these models and we're happy with the results we're getting, we'll evaluate these candidate models and see which one fits our application the best.

Then, we'll select the best model and deploy it for some time and test in a real working environment or in production. Also, we'll monitor the performance. If our model continues to perform well and to do what we expect it to do, we may say that our machine learning solution is developed and we're happy with it.

Now, we can use it in the production environment in our system. ☺

HOW CAN YOU USE ML IN INDUSTRY 4.0?

As we know, in Industry 4.0 and the Industrial Internet of Things, there are more and more devices that are getting interconnected and that are able to collect the data from their environment, or from the system. They can communicate this data among each other, and probably collect all this data into one place.

Of course, every time we have these big amounts of data, we can think about using machine learning to get some value for our business application.

But as we've established throughout this e-book, we can't solve all problems with machine learning. Although, those that we can solve, are going to be a fun journey.

HOW CAN YOU USE ML IN INDUSTRY 4.0?

Here are some examples of what we can do with machine learning:

- Early fault detection

We could predict when the fault of the system is going to happen.

Therefore we're able to suggest predictive maintenance actions of our systems or our equipment in order to timely repair or change what needs to be adjusted or fixed before failure.

- Process and output optimisation
- Adaptive control
- Decision support for better business decisions
- Production forecasting
- Demand management
- System anomalies
- Intelligent system supervision

STEP BY STEP...

Machine Learning models analyse HUGE datasets to deliver faster, more accurate results.

But at the end of the day, what you actually unlock through machine learning is **knowledge.**

Knowledge enables you to avoid risks, make better decisions, and empower your people to work more efficiently and effectively towards their goals.

To remain competitive in your market, consider starting your machine learning journey as soon as possible.

But remember: don’t get carried away by all the nearly limitless possibilities.

Start small, learn, and build along the way.

Want to see a real-world example of a machine learning solution, with some help from Ignition?

CLICK HERE FOR A VIDEO WALK-THROUGH!

Enuda is an Ignition Premier Integrator.

We build customised industrial solutions (SCADA, MES, IIoT, Data Analytics) for companies in the EU who embrace new technology *to improve efficiency and work smarter in the era of Digital Transformation.*

Sign up to The Enuda Insider to get insightful information like this right in your inbox.

SIGN ME UP!

www.enuda.eu