SuperDataScience

Machine Learning

Executive Briefing

v 3.0

WELCOME

Instructor Introduction

Kirill Eremenko wants every person he meets to know the extraordinary potential of data. In his role as an online educator and CEO at SuperDataScience, Kirill has helped over 2,000,000 students master the topics of data science, artificial intelligence, and machine learning.

Machine learning will reshape the way you do business. Let Kirill show you how.

SuperDataScience

Purpose of the course & overview of learning outcomes

Amazon and Google rank among the world's most successful corporations. They have reshaped the way we do business, what we learn, how we read, and how we see the world. They have the power to know what we are searching for, even if we can't put it into words. They can tell us what we want before we know it ourselves.

These prosperous organizations both use a special tool that contributes to their success. It requires no significant investment of your time and money. It will not ask you to overhaul your company's organizational structure or way of working. It is something that you can implement into your activities, and that will work alongside you and your team.

You might have already guessed what this powerful tool is: Machine Learning.

So many companies fall for the allure of this profitable discipline. But it's not easy to know how – or where – to start. Maybe you've already tried and failed to implement a data-driven project. You might have advertised for a Machine Learning engineer, only to be underwhelmed by the results that they've produced.

Implement it in the wrong way, and Machine Learning can lead to snags in your development. This course will show you how to avoid these traps and instead develop a resilient strategy for success.

Welcome to Executive Briefing: Machine Learning, a straightforward overview of Machine Learning packed with real case studies and tutorials. These videos will take less than 90 minutes of your time, and they could save you – or make you – a fortune.

Executive Briefing: Machine Learning will show you that there is nothing unapproachable about the discipline. It will:

- ensure you can navigate technical terms and topics related to Machine Learning like a pro,
- help you to develop a robust strategy for implementing the discipline into your organization,
- highlight the tools needed to create a stellar Machine Learning project,
- explain the steps to take to build a reliable team of experts and support staff, and
- show you how Machine Learning can supercharge your company's finances, innovation and marketing strategies.

After this course, you will see how Machine Learning will quickly become your company's most valuable member of the team.

MODULE 1

What, Why & How of Machine Learning

1.1. Purpose of this section

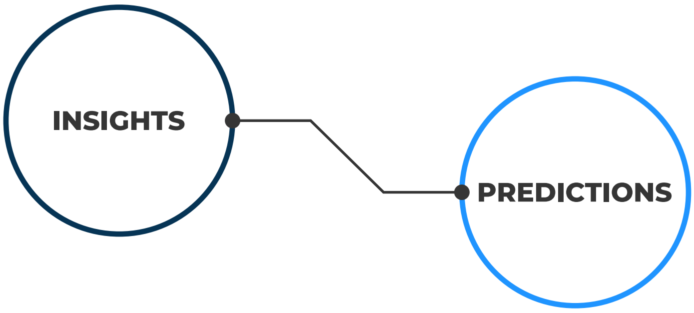
What is Machine Learning? How can it benefit a business?

This section offers a concrete definition of the term and explains its relationship to Artificial Intelligence and Deep Learning. You will also discover how Machine Learning can help you to:

- innovate with new products,
- market the right services to the right customers, and
- reduce overheads

1.2. What is Machine Learning?e

Machine Learning teaches computers to make predictions based on data. The discipline, therefore, (1) teaches machines and (2) makes predictions. For these two components to function, you need to have data.



Let's look at how both components use data:

1. **Learning & Teaching:**

   Humans learn through a combination of theory and practice. Computers are no different. For a machine to understand the benefits and obstacles to your company's success, you must first feed it with past experiences. These experiences are units of information that teach your machines trends and parallels derived from the data, which enable them to forecast future change.

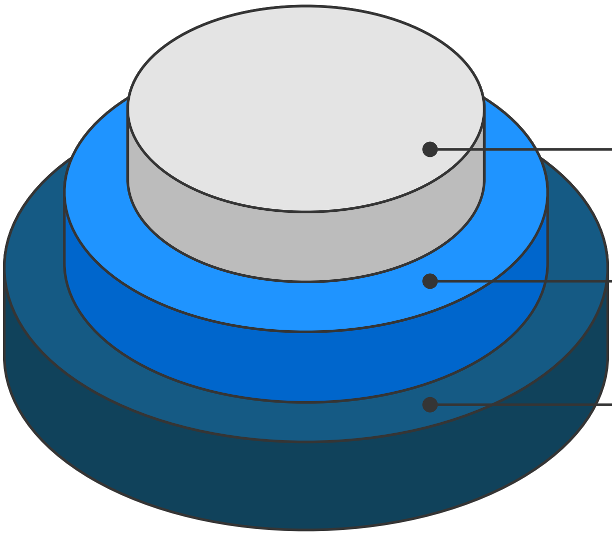
2. **Predictions:**

   Once the algorithm has seen enough past observations to be able to identify patterns confidently, you can make accurate predictions on new data. This is where the most valuable insights are extracted to inform the key business decision makers.

Machine Learning is a discipline based within the broader field of Artificial Intelligence.

The terms Artificial Intelligence, Deep Learning and Machine Learning are often used interchangeably. They all refer to automated processes that rely on company data. So, how can we differentiate between them?

The simplest way to understand how Deep Learning and Artificial Intelligence correlate to our discipline of interest is to consider each term as one part of a Russian doll. Artificial Intelligence is the biggest of them all, and therefore the outermost shell. Machine Learning sits inside of it, and Deep Learning is packed inside both as the smallest 'doll'. Logically speaking, all Machine Learning is a type of Artificial Intelligence, but not all Artificial Intelligence is Machine Learning.



Machine Learning has been around for a long time – it is not a new concept. So, why haven't you heard of it before now? Because technology needed time to catch up with the idea. Computers have only recently become fast enough to crunch numbers at the speeds we need.

Think about how long it took you to download films from the Internet in the late 90s, and how long it takes you now. That's a world of difference. Towards the end of the 20th century, our machines either couldn't manage large datasets or took an agonising length of time to deliver results. Now, you can throw numbers at a computer, and an algorithm will sort through the information, all at the touch of a button.

This is why Machine Learning is so important now – because this is the first time in history when we can work with significant amounts of information within a short timeframe.

1.3. Why do we do Machine Learning?

In the words of the father of business consulting Peter Drucker, "the business enterprise has only two functions: marketing and innovation. Marketing and innovation produce results; all the rest are costs." These functions, he explains, are conducive to what all businesses need: customers.

Both functions – innovation and marketing – can be improved through Machine Learning. As a bonus, you can also use the discipline to knock down company costs.

Innovation

Google usually crops up as the perfect example of an organization that uses technology to innovate. One example of this is its Machine Learning algorithm that detects spam emails and expunges them from your inbox.

This function adds value to their service. Many of us don't even check our spam inbox – half the time we don't even remember it's there. But imagine how your inbox might look if it had no spam filter. Google's function is so useful that it simplifies our life and runs in the background so we can get on with our day.

Marketing

Amazon uses a powerful recommendation engine based on Machine Learning. You will have noticed it in the emails the company sends you about products similar to those you've already bought, and in notifications about items frequently bought together with the product that takes your interest.

By adding a Machine Learning algorithm to online marketing tactics, this little tool uplifts Amazon's sales and revenue by 25%. All the algorithm uses are purchase data to target similar customers. For such a simple idea, Amazon has netted an enormous profit.

Costs

In 2017, investment bank JP Morgan found a way to reduce its overhead significantly. It introduced a Machine Learning algorithm that enabled it to automate the process of interpreting commercial loan agreements.

This move saved the bank 360,000 hours of legal practitioners' work per year – the algorithm could do the work of a human in seconds. Adding the Machine Learning function saved JP Morgan both time and those hefty lawyers' bills.

PRACTICAL ACTIVITY 1.3

What are the most significant challenges for your organization right now? Think about the problems that you want to solve for your customers. Consider the mechanisms that bring operations grinding to a halt. Perhaps you have access to evaluations forms from your employees – what, if any, are their recurring complaints? What opportunities do you envision for your organization, that up until now have been out of reach?

Don't hold back your pen. The more challenges you can identify, the better. Once you have a list of at least five issues, enter them into the chart below.

Make sure to use the correct column! The Innovation column should contain problems related to your company's future activities. Marketing should only include issues that concern branding and reaching out to customers. The final Costs column, following Peter Drucker's philosophy, is for everything else.

WHAT ARE MAJOR CHALLENGES IN YOUR ORGANIZATION?			
	INNOVATION	MARKETING	COSTS
1			
2			
3			
4			
5			
6			
7			

WHAT ARE MAJOR CHALLENGES IN YOUR ORGANIZATION? (CONT'D)			
	INNOVATION	MARKETING	COSTS
8			
9			
10			

MODULE 2

Machine Learning and Your Business

2.1. Purpose of this section

Strategic thinking is important. Too many executives want to dive straight in, but take a moment to sit back and consider the big picture. Module 2 will help you to think through your strategy for implementing Machine Learning into your organization.

Every business professional should recognise the value of strategic thinking – it is the best way to ensure that you and your team are on the path to success. This module outlines the steps you should take to develop a well thought-through Machine Learning strategy, to protect you against project setbacks and snags.

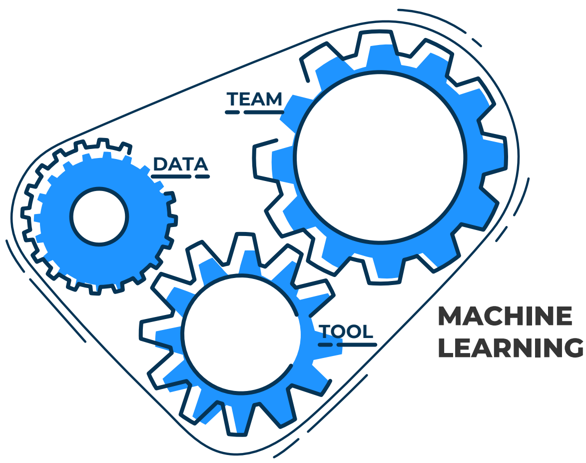
1.3. Why do we do Machine Learning?

Your Machine Learning strategy must always consist of three components:

- Data
- Tools
- Team

Each one of these is essential to the running of your project. Think of them as cogs in a machine – if one of them is missing, the others won't turn.

First, you need data to plug into the tools. Then, the Machine Learning tools must be there to execute analyses. Finally, you need a successful team by your side to ensure that you collect the right data and that you have chosen the most suitable algorithm for the job.



Machine Learning exists at the intersection of these three components.

Think about the problem you want to solve.

An unclear problem will lead you on the road to failure. Think about a real challenge that your company faces. What will its resolution do – improve revenue? Increase customer loyalty?

PRACTICAL ACTIVITY 2.2

Using the chart from Practical Activity (1.3), identify which of those challenges is the most important to you and your company. Which one causes the most day-to-day stress to your workers? Which one makes a dent in your profit?

Don't worry at this stage if you are not yet sure how Machine Learning can resolve the problem. Throughout the course, as you get to know more about Machine Learning, you'll come to understand its breadth of possibilities. You can return to this exercise at any time and change your goal. For now, pick something to use as a case study, so that you have a tangible example in front of you.

Now, answer the three questions below:

Which challenge have you identified?

How does your organization currently respond to the identified challenge?

Imagine your ideal outcome. What would your organization look like if you could overcome this challenge?

2.3. Having your Data in Check

Data is generated all the time. It tracks stock on product lines, it notifies recipients when parcels arrive, it logs customer footfall – there are so many events in our life that generate data. But how can you capture it?

Every successful Machine Learning project requires an initial three stages related to data. Be mindful of them when you're developing your Machine Learning strategy. They are:

- Capture (through sensors, data points on your website, tracking devices on equipment)
- Store (in storage facilities such as SQL or data lakes)
- Process & Analyse (with relevant Machine Learning algorithms)

Stage 1
CAPTURE THE DATA

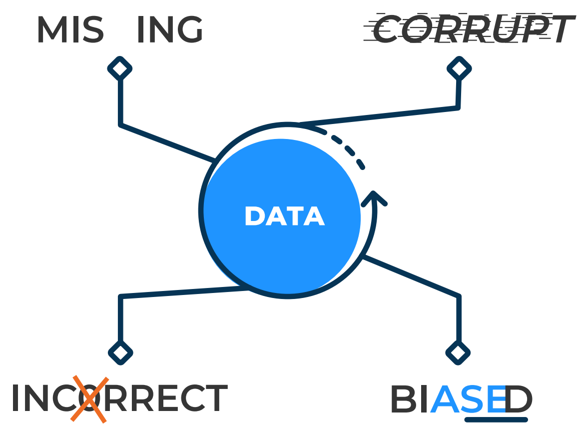
Stage 2
STORE THE DATA

Stage 3
PROCESS & ANALYZE THE DATA

There are many potential pitfalls at this early point that are necessary for you to consider, to ensure the smooth running of the project. Most problems that arise at the third stage are the result of unresolved issues in steps one and two. They are:

Your data is a mess

Some of the most common problems for companies that do not have data science strategies in place are that the information you have stored is missing, corrupt, or just incorrect. Your team might need time to find, gather, and clean your prior data, for it to function alongside your new units of information.



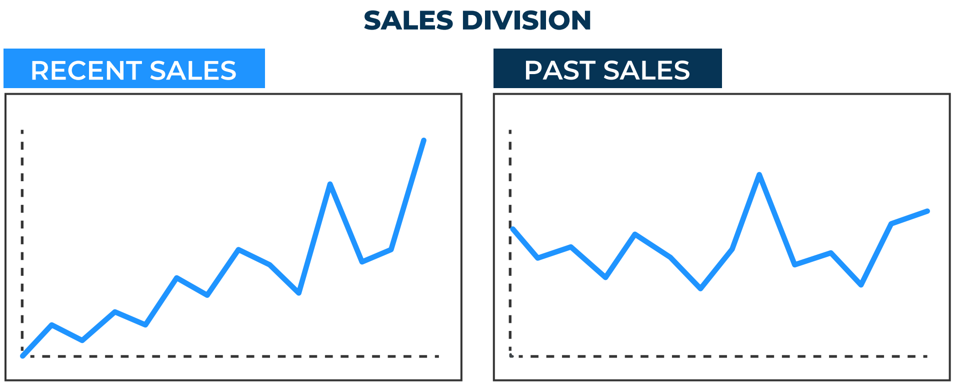
Your data only tells you what you think is true

Biased data is one of the most challenging issues that you might come across. If a mortgage company is using a Machine Learning algorithm to replace its manual decision-making about whether or not to offer a loan, it will use the company's historical data to predict the outcome for each new client. But the company's prior data is biased. Although it contains information about the loans that were and were not successfully repaid it does not offer any detail about the loans which weren't approved – but that might have been successful. Thus, the mortgage brokers continue to make the same errors in judgment because the machine has learned the same bad habits.

Your data is in the wrong place

Organizations might have grown in such a way that data is kept in different departments, making them not easily accessible across the company. This disparity leads to data silos – information archives that exist separate to the rest of the organization. Silos are a common problem, and it can stop your project dead. Here are four possible reasons for silos:

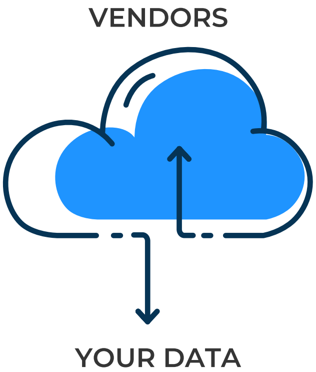
1. **Structural/functional:**
   A sales division might choose to store their recent months' sales in one location for easy access and to archive all previous months in another. This division makes it hard to analyse all sales data together.



2. **Political:** Data is power, and many managers know it. Some departments become data hogs, holding onto their information and unwilling to relinquish it to other staff.

3. **Growth:** Management changes, new tools enter the market, and operational strategies get updated. These upheavals can lead to disparate, unmanaged data.

4. **Software vendor lock-in:** This is potentially the most problematic issue. When you use external services to log your information, you leave your data at the mercy of that company. They also understand that data is power, and many will keep your data in their cloud, meaning access to it might be temperamental.



Consider each of these issues when building your strategy. Leaving one of these areas unresolved may cause roadblocks to your project.

Finally, ensure that you leave enough time for quality assurance. Successful Machine Learning projects take time. As you can see from this section, data preparation, processing, and compilation take a significant chunk (roughly 80%) of project time. A good rule of thumb is to quadruple the time requested by your team. If your data scientist or Machine Learning engineer tells you that a model will take two weeks to build, multiply that amount of time by four to get the amount of time required for data preparation and quality assurance, and you should expect your project to be complete in a total of 10 weeks.

PRACTICAL ACTIVITY 2.3

To see if Machine Learning can solve a particular problem, we first need to establish if the required data is free from any of the issues outlined above. Consider the challenge you selected in Practical Activity (2.2) and identify:

- the data that might be required to resolve it,
- if there are any possible issues with it, and
- (if 'yes' to the above question), how those issues can be fixed.

It takes practice to know what type of data you need. The best way to identify required information is to remember that any action, any event, generates data – the onus is on your organization to capture it. Look at the processes that relate to your chosen challenge. What steps are required to complete those processes? If you're looking at online purchases, consider that a transaction first needs a customer to put an item into their virtual cart. If you want to track employee attendance, remember that workers have to leave work by clocking out. If you are looking to make production more efficient, take into account that a factory conveyer belt needs information about each of the points where a new part is added to a product. Jot down the relevant actions for your selected challenge.

Consider the problems concerning information that you already learned about, such as incorrect data and data silos. Always double-check your historical data, or your analysis will only return biased results.

How can these identified data issues be addressed? If data silos exist, how can they be broken down? If you need to gather more data, can a budget be allocated for creating systems to capture it? Can missing data be recovered, or can they be purchased from external sources?

Use the chart below to record your findings.

IS YOUR DATA IN CHECK?			
	REQUIRED DATA	POSSIBLE ISSUES WITH THE DATA	HOW CAN THE ISSUES BE RESOLVED?
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			

IS YOUR DATA IN CHECK? (CONT'D)			
	REQUIRED DATA	POSSIBLE ISSUES WITH THE DATA	HOW CAN THE ISSUES BE RESOLVED?
11			
12			
13			
14			
15			
16			
17			
18			
19			
20			

2.4. The Tools

There are lots of tools available to a data scientist – you might have heard of for-profit software such as SAS, SPSS, Matlab. These will help you to carry out your Machine Learning experiments.

Many organizations prefer to go with commercial tools because it gives them a feeling of security. What they're creating is their intellectual property, and it cannot be easily shared or taken by other people online.

But there is also a strong case for using open source tools and software.

You might wince at the idea. Open source. Won't that leave my data accessible to hackers? Will others be able to steal my information? If it's free-to-use, surely it can't be very good?

These are common misconceptions of what open source actually means. You might be surprised to hear that Facebook and Google both use open source tools for their Machine Learning practices and even their AI algorithms. These tech giants have plenty of money to buy software if they wished to do so, but they use open source (specifically, Python). Why?

Think about the problem you want to solve.

Don't be misled by the thought of creating the world's most cutting-edge Machine Learning tools. Static software is not valuable – the value is sitting in your data. Thousands of researchers at Google and Facebook are developing open source software and publishing them online. Google published 454 research papers in 2018 alone – that's more than one per day. So, why waste time inventing the wheel?

You are not going to outfox Google in the tech game. Rather than tax your Machine Learning team with developing software that will go obsolete in months, get them to read Google's research papers and consider how to apply their approaches to your business.

Open source tools allow a Machine Learning team to start from an already developed base, to find solutions online, and to source talent.

Another benefit to using open source software is that they have massive online networks. Software that is not open source often doesn't account for the speed at which technology develops. With so many people across the world contributing to the development of tools, if your Machine Learning team has any questions or needs guidance, online help is readily available. The latter benefit also makes it easier to find talent, directly from the pool of people assisting you.

Two industry-preferred open source tools dominate Machine Learning: R and Python. My recommendation is Python. This tool allows you not only to carry out Machine Learning but also develop applications, and it facilitates the productisation of your algorithms. Google uses it for TensorFlow; Facebook uses it for PyTorch. There are many manuals and resources for both these tools, to help you in your search for the perfect solution.

2.5. Hiring a Machine Learning Team

When you're hiring a team of people to help you with Machine Learning, avoid looking for the 'unicorn' who will magically solve your problems. Too many organizations want to hire someone to fix all their problems without really thinking through the specifics of what they need. Career sites are flooded with completely unrealistic job descriptions, looking to find candidates with ten years' experience, a PhD, and knowledge of 50 different tools and algorithms. These are red flags to the best data scientists – they signal desperate managers who don't understand what Machine Learning does and how it can help to solve problems.

UNICORN

HAS NEVER BEEN FOUND

Select and structure your team the smart way.

Tech entrepreneur Lukas Biewald outlines two approaches, one for small companies, and another for medium-to-large organizations. The former may have budget constraints. If this is the case for you, look for a Machine Learning engineer who is passionate about your company's mission. When they start work, ensure they are not focused on multiple projects – instead, ask them to prioritise the most critical initiatives. Concentrating on one project at a time will ensure they deliver maximum value to that specific issue.

For medium-to-large enterprises where the budget is not such a constraint, seek to build a team of Machine Learning engineers. Hire someone who already has experience in the problems you want to solve. Then surround them with a team of support staff who can support them in preparing the data. These could be analysts or business intelligence experts, or they could be people already working in the company, who are passionate about the topic and want to learn more. These support staff should assist the expert with individual components of the project's lifecycle.

ORGANIZATION TYPE	
SMALL	**MEDIUM/LARGE**
IS BUDGET A CONSTRAINT?	
YES	**NO**
RECOMMENDATION:	
One M.L. Engineer passionate about your mission	One experienced or talented M.L. Engineer
MANAGEMENT KEY:	
Not focused on multiple projects Prioritize important projects one by one	Build a M.L. Support Team Team helps out on project lifecycle

2.6. Integrating Machine Learning in the Organization

It can be easy to get ahead of yourself. Blue sky thinking is a large part of AI and Machine Learning. But when you're just starting, you also need to keep your feet firmly on the ground. Prioritise your projects, and focus on specific problems and applications for them.

One useful point of consideration that you should have at this stage is where your team will sit in the company. Which department will be the 'parent' for your team – marketing? IT? Operations? Sales? There are two approaches for an executive to take. Both methods have their merits, and it is essential to consider both before settling on one:

1. **Create a separate division:**

   For large companies, many Machine Learning teams begin in IT and eventually branch out into a division in their own right. When this happens to your company, head up this division with a Chief Data Scientist. Ask them to sit in on executive meetings and advise how and where their team can add value to the organization. Enabling the division's head to sit in on board meetings will improve communication across departments and prevent data siloing by giving your Data Science team authority. The only limit here is the cost of developing a Machine Learning division, but if your scientists are working to improve company operations, you should consider appraising this value against the staff costs.

2. **Install experts into relevant departments:**

   Rather than building a standalone Machine Learning team, you could consider dropping an engineer into every division that could utilise Machine Learning. This approach enables each scientist to help their target division, not to mention having an expert in each department will facilitate cross-company communication related to information archives. One disadvantage of this approach is a lack of assistance. Without support staff, your experts might become overwhelmed with their projects.

PRACTICAL ACTIVITY 2.6

Which of the outlined approaches will you take? Using the table below, list the benefits of both options for your organization. If your organization is small and for the time being you are looking to hire only one Machine Learning expert, then perform this exercise for a future point in your business. What will your company look like in 5-10 years? How will it change, and how could a Machine Learning team support or even bring about that change?

WHICH APPROACH?		
	BENEFITS OF A STAND-ALONE MACHINE LEARNING TEAM	BENEFITS OF M.L. EXPERTS INTEGRATED WITHIN DIVISIONS
1		
2		
3		
4		
5		
6		

WHICH APPROACH? (CONT'D)		
	BENEFITS OF A STAND-ALONE MACHINE LEARNING TEAM	BENEFITS OF M.L. EXPERTS INTEGRATED WITHIN DIVISIONS
7		
8		
9		
10		
11		
12		
13		
14		
15		

2.7. Recap

In this unit, you learned that a Machine Learning strategy comprises three components: data, tools, and team. Identify the questions that you want your Machine Learning team to answer. Rank them in order of importance. Ensure your team does not disperse its efforts but instead focuses on one project.

1. **Data:**

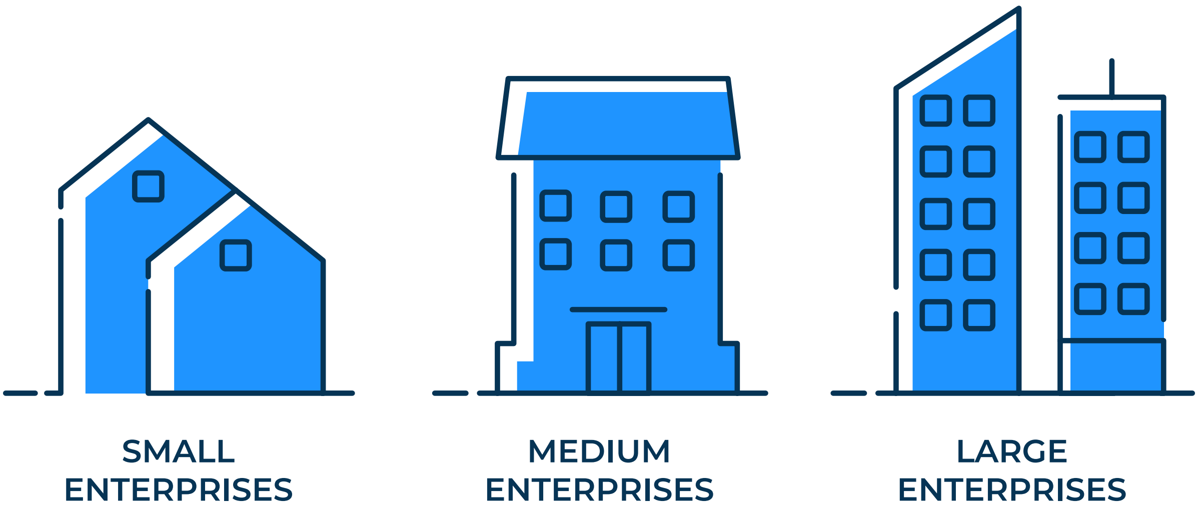
   Every business generates data. You need to find a way to collect, store and process it.

2. **Tools:**

   Try out open source software such as R and Python, and leverage these publicly available tools to gain momentum in your Machine Learning practice.

3. **Team:**

   Build a Machine Learning team that fits your type of organization and the problems that you want to solve.



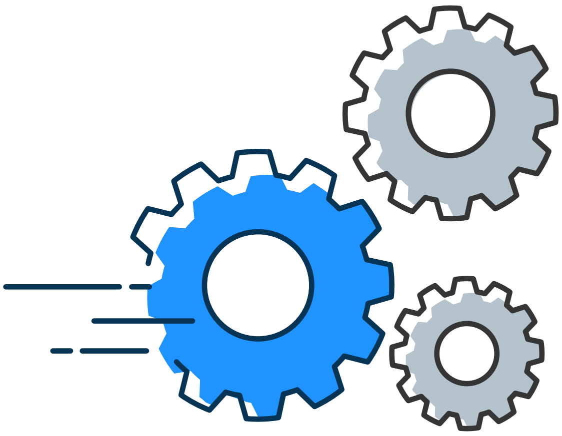
MODULE 3

Technical Case Studies

3.1. Purpose of this section

This module provides real-life business examples of Machine Learning. You won't find a single equation here – everything you learn in Module 3 will be from an executive viewpoint.

You will come away from this module with a clear idea of what each algorithm does, and which Machine Learning algorithms could benefit your business' individual challenges.



3.2. Regression Models: Intuition + Case Study

Regression is a branch of Machine Learning, and it is used to predict continuous variables: such values as revenue, expenses and mileage. Continuous variables are numeric variables that have an infinite number of values in a range between any two values. A value for mileage, for example, could be anywhere from 5 to 160 mph: it could be 15.2 mph, or 15.21 mph, or 15.217 mph, and so on.

NOTE Not all numeric variables are continuous. For instance, 'Birth Year', while numeric, is actually a categorical variable. We will touch more on categorical variables in further tutorials.

The Problem

You are the CEO of a luxury travel company and you want to know if the happiness of your customers affects your revenue.

The Data

You will need to gather your data for revenue and customer satisfaction. Each point on the plot represents a calendar month and customer satisfaction score, as well as revenue for that month.

The Terms

Revenue is the most important factor for this study because it is the value that you want to examine. For this reason, it is known as the dependent variable. Customer satisfaction is therefore the independent variable, because its change is not affected by your revenue.

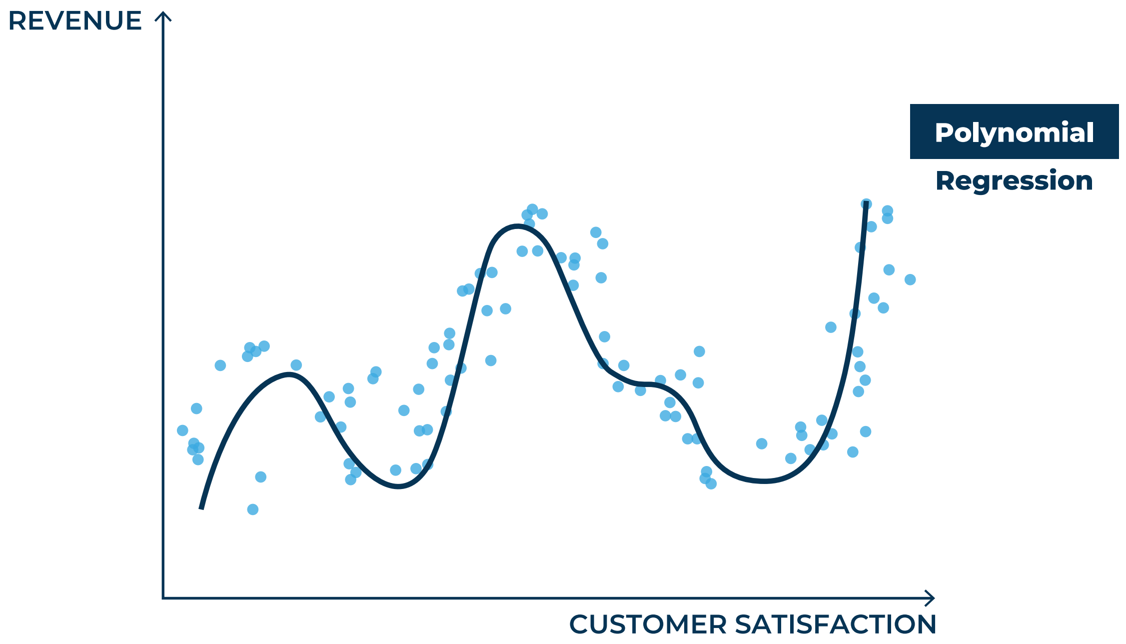
The Analysis

Let's say you have been collecting relevant data for many months. By entering this information into an algorithm for regression, you might see a trend emerge (each point represents a separate month). This is known as a simple linear regression, and you may remember it from high school.

The result in the image is clear: The more the customers are satisfied, the higher the revenue of the business will be.

Regressions will not always form straight lines – sometimes, the data will show something more scattered. These are called polynomial regressions.

his is a very basic example of regression, which plots one variable against another. But regressions don't have to be restricted to two variables. You could input several variables, if you feel that they are relevant to answering your project's question. Using the above example, there may be factors that affect your revenue additional to customer satisfaction, such as the weather, or the revenue of your competitor. In these cases, where you use more than one independent variable, you need a multiple linear regression.



PRACTICAL ACTIVITY 3.2

The best way to commit something to memory is through practice. Think of a continuous numeric variable in your organization that you might be interested in analysing with simple linear regression. This will be the dependent variable for the purposes of this exercise. Next, think of several independent variables that could relevant to the variable you identified, for a multiple linear regression.

Brainstorm your thoughts using the chart below

REGRESSION IN YOUR BUSINESS		
DEPENDENT VARIABLE		INDEPENDENT VARIABLES
	1	
	2	
	3	
	4	
	5	
	6	

3.3. Classification Models: Intuition + Case Study

Classification algorithms group together similar data. Here's an example of how they might work for your company:

The Problem

You're the marketing director of an ecommerce company and you want to analyse which customers are likely to churn.

The Terms

Here, 'churn' is the dependent variable and a categorical variable, the latter because it either has a value of 1 or 0. It tells us if a customer is likely to leave to a competitor (1) or not (0) within six months of joining. Your independent variables for this experiment are 'Age' and 'Time on webpage' (time spent on your website per month, on average).

NOTE More broadly speaking, categorical variables have one of a set – usually fixed – number of values. These values assign each unit of information to a group. Categorical variables are the basis of classification because, as the name suggests, you want to classify your observations into groups.

The Data

Every point represents an existing and/or past customer, their age and the average time they spend on your website per month. The colour of each point represents your categorical variable of churn (whether or not a customer left within six months of joining). Since you are training your algorithm with past data, you can extract this information from your database.

The Analysis

To see if a new customer is at risk of churning within the next six months, you need to first collect data for their age and average time per month they've been spending on your website since joining, and then plot them onto this historical data.

In the scatterplot, there is a clearly defined boundary between loyalty and churn. Depending on which side of the boundary the customer under investigation falls, he or she may or may not be at risk of leaving. If they are in the 'remain' group (on the right) they are less likely to churn in the first 6 moths. If a different customer were to end up on the other side, then they are at risk and you could take action (such as notifying your marketing department) to make sure they stay.

This is a trivial example of a basic problem. In this 2D diagram, it is easy to see where the groups lie. You might think that you don't need a Machine Learning algorithm to analyse this. However, in most cases the analysis would involve thousands of customer data points and multiple variables. It would simply be impossible to plot the results in a 2D diagram. That's where Machine Learning comes to the rescue – the algorithm will calculate everything in the background and deliver actionable insights.

PRACTICAL ACTIVITY 3.3

As with Practical Activity (3.2), think of a categorical variable in your business and up to five variables that could count as predictors for it. Use the table below to note down your thoughts. These insights can be used to inform your future Machine Learning projects.

CLASSIFICATION IN YOUR BUSINESS		
CATEGORICAL VARIABLE		PREDICTOR VARIABLES
	1	
	2	
	3	
	4	
	5	
	6	

3.4. Clustering Models: Intuition + Case Study

Clusters reveal patterns. In the earlier example on classification, you looked for known groups – customer loyalty vs customer churn. Clustering is similar to classification in the sense that it is also designed to identify groups within your data. The difference is that, with clustering, you don't yet know what you're looking for. This is called unsupervised learning. Clustering finds groups for you, but it is your job to interpret why these patterns emerge and how they're relevant to your business.

The Problem

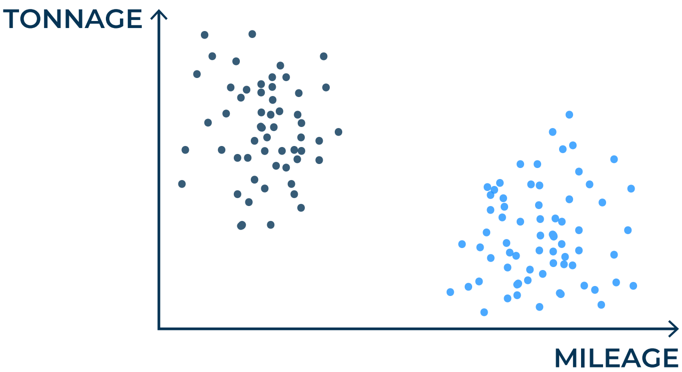
You're the operations director of a mining company, and you want to explore if you can optimise the servicing needed for your hauling trucks.

The Data

You would need to collect data that you feel is relevant to the organization. For the purpose of this exercise, either due to your domain knowledge or based on intuition, you have chosen to analyse data for hauled tonnage and travelled mileage of the trucks.

The Terms

Supervised learning refers to algorithms where you know what categories to expect prior to the analysis. Classification is an example of supervised learning. Unsupervised learning is the opposite, where the algorithm returns groups previously unknown to you. Clustering is an example of unsupervised learning.



The Analysis

By plotting these data points on a chart, two clusters emerge. How you choose to interpret these groups is a question for your team. What do you know about mileage and tonnage and their effect on hauling trucks? What could be the reason for these clusters? Perhaps the group containing trucks with high tonnage and low mileage require additional servicing. Use your domain knowledge to answer this question.

NOTE Once the clusters have been identified and explained, they can be used in a classification model similar to the one in the previous tutorial. Any new data point (in this instance, a truck) can be analysed across the same parameters ('Mileage' and 'Tonnage') by the algorithm, to see which group it falls into. This will then enable you to make appropriate business decisions about that truck.

PRACTICAL ACTIVITY 3.4

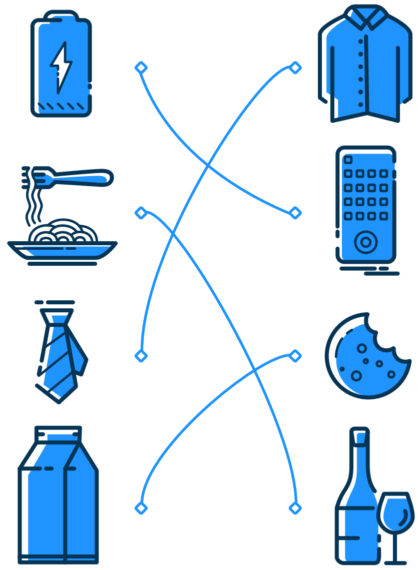
Clustering can reveal patterns in your data, which makes it one of the most exciting branches of Machine Learning. But this also requires your expertise, to interpret the results.

Using your domain knowledge and business experience, list five variables that might be of interest to your Machine Learning team when they are searching for hidden clusters using the organization's data.

CLUSTERING IN YOUR BUSINESS	
	VARIABLES
1	
2	
3	
4	
5	
6	

3.5. Association Rule Learning Models: Intuition + Case Study

Associations between different types of data can give you a deeper understanding of how customers, processes, products, or even ads are all likely to behave in the future. You might have heard the infamous example that the sales of diapers and beer are strongly correlated. Whether or not this is true, this example serves a point, that people tend to make purchases of similar combinations. Crackers, for example, might be bought in combination with humous or another dip. Electrical appliances are likely to be bought together with batteries.



These trends can be found in any marketplace, from Walmart to Netflix. Organizations make smart use of association rule learning to better serve their customers and tailor their sales and marketing. If you watch Stranger Things on Netflix, the platform might later recommend to you a Stephen King movie, even though you've never seen or considered watching one. This is because Netflix accumulates prior data from other users and the content tags regarding the series to associate like products.

This is why you may get the impression that conglomerates know you better than you know yourself. Machine Learning, especially association rule learning, can be used to derive benefits for customers. Extract these associations with Machine Learning and discover the hidden potential of recommendations.

PRACTICAL ACTIVITY 3.5

Using the space below, record any associative components that you can find for your business. What do your customers frequently buy together? Which of your services are reliant on each other?

3.5. Association Rule Learning Models: Intuition + Case Study

Deep Learning is a subdiscipline of Machine Learning. Within a Deep Learning tool, a deep neural network 'mimics' the human brain, to learn processes and categories on its own.

Deep Learning has been around since the 1980s. Over this time, its theoretical concepts have been subject to rigorous attention and adjustment. Now, we can input thousands of data units and obtain insights within minutes.

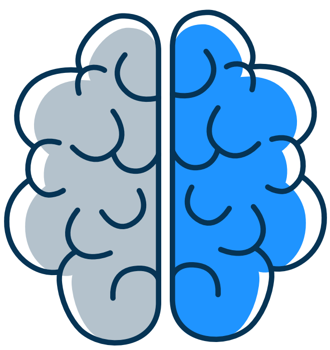
What is so attractive about Deep Learning, as opposed to other branches of Machine Learning, is that it has no predefined framework. There are no groups or structures to fit onto our data. What we do have, however, is a neural network – a deep neural network comprising multiple layers.

A neural network is an artificial construct designed to mimic how the human brain works.

A Deep Learning model will learn about your dataset in the same manner as a baby learns a language. The child has a brain, but it has no prior understanding of vocabulary or the rules of grammar. It has no predefined framework. By interacting with its parents, it will over time learn to speak.

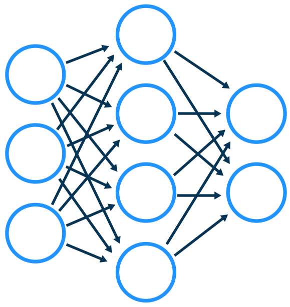
A Deep Learning neural network uses the same principles. We feed it the necessary information, so that it can learn. Let's say you wanted a machine to distinguish between dogs and cats. By uploading thousands of labeled photos of these two animals, you can teach a neural network to distinguish between them. The more photos you have, the more accurate it will become. Again, the principles are similar to the way that humans process information. The more we know about an animal's distinguishing characteristics, the more likely we are to correctly identify it.

HUMAN BRAIN



Although Deep Learning algorithms will increase in accuracy over time, the drawback is that a great deal of data is needed for it to be effective – far more than other Machine Learning algorithms. The benefits, however, can be game-changing. A Machine Learning tool that tries to recognise handwritten text might get an 87% accuracy rate today. With a Deep Learning tool that has enough data, it's entirely possible to get 99% accuracy – it's that powerful.

DEEP LEARNING



PRACTICAL ACTIVITY 3.6

Having lots of data is one of the prerequisites for Deep Learning. Which area of your business has a surplus of data? Anything to do with rich media (audio/image/video) is a particularly good option. Write down three potential areas of your business in the table below.

DEEP LEARNING IN YOUR BUSINESS	
	AREA OF BUSINESS
1	
2	
3	

MODULE 4

Conclusion

4.1. Review Learning Outcomes

This course advocates taking the time to develop a **Machine Learning strategy**, the necessary components of which include data, tools, and a team, to ensure that its implementation will add value to your business. To recap the process:

- It is essential to **identify a question** before jumping into a project, as Module 2 highlights. You cannot run a Machine Learning project if there is no data, and understanding what your question is at the beginning of the process will help you to **identify the data you need** to collect.
- Hire **the right people** in the right number, depending on the size of your business and what it hopes to do with Machine Learning. Use Module 2 to aid you in making this decision.
- Select an **open source tool** such as R or Python to kick-start your first Machine Learning project. Choose among the common **algorithms** of Machine Learning –Module 3 tells you which one to use for which project.

BONUS

5.1. Next Steps

Now that you've completed Executive Briefing: Machine Learning, what can you do next?

Why not get together a team in your company and educate them about Machine Learning as well? Our comprehensive Machine Learning A-Z course contains 40+ hours of content, 300+ hours of tutorials, and it runs the gamut from R to Python across all branches of the discipline from regression, association rule learning, clustering and classification, to natural language processing and deep learning. This course is the quickest and easiest way to get everyone to understand the business and practice of Machine Learning, and there are plenty of frameworks and templates for your team to download and use. You'll even convert the technophobes.