Tuidi's course on Machine Learning

INTRODUCTION TO AI IN THE FOOD SUPPLY CHAIN

TUIDI

This course aims to introduce the concept of Artificial Intelligence (AI) by addressing three fundamental questions:

- What is AI?
- What technologies govern it?
- What can AI do for us?

The goal is to provide all the knowledge and tools needed to understand what is meant by artificial intelligence, how data processing is carried out to develop predictive models through machine learning, and practical examples within the retail and production sectors.

The course is divided into four lessons:

- **Lesson 1** - Artificial Intelligence: Understanding what this discipline is, what it comprises, its applications in the retail industry, and the main differences from traditional statistical models used so far.

- **Lesson 2** - Data Mining: This lesson explains the data extraction process, emphasizing the importance of accurate information to achieve predefined objectives.

- **Lesson 3** - Machine Learning: Automatic learning techniques, including Classification, Regression, and Clustering.

- **Lesson 4** - Predictive vs Prescriptive Analysis: Explaining the distinction between these two models, detailing their approaches, advantages, and disadvantages.

Lesson 1

ARTIFICIAL INTELLIGENCE

INDEX

1.1 Introduction to the lesson

In the first lesson, we will become familiar with Artificial Intelligence by defining the concept in broad terms and then exploring it in detail through some examples. AI is a vast and constantly evolving topic, characterized by a multitude of possible applications.

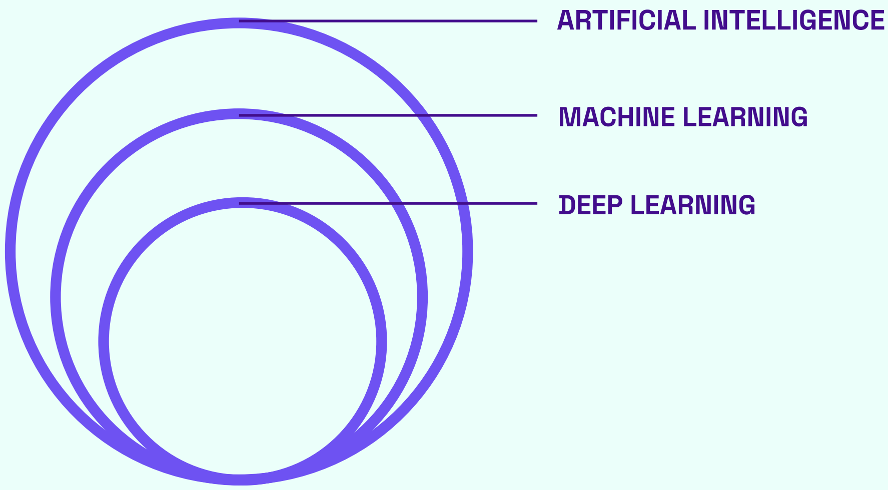
For some, AI refers to artificial life forms capable of surpassing human intelligence, while others define it as a set of data processing technologies.

This lesson will explain what Artificial Intelligence is, its uses (with a particular focus on the food sector), and its advantages compared to traditional statistical systems.

1.2 What is Artificial Intelligence?

Artificial Intelligence is defined as the ability of a machine to demonstrate human-like capabilities such as learning, planning, and creativity. What makes this technology “intelligent” is its ability to learn from manually provided data and use the acquired information to achieve predefined objectives.

The concept of AI is often confused with and mistakenly replaced by the terms “Machine Learning” (ML) and “Deep Learning” (DL), as if they were interchangeable. To clarify the difference between these terms, it can be useful to visualize them as concentric circles.



In fact, ML and DL are not synonyms but rather two subcategories of Artificial Intelligence. Machine Learning, or automated learning, is a subset of AI that relies on learning from data without requiring predefined manual rules.

All Machine Learning systems are forms of Artificial Intelligence, but not all Artificial Intelligence systems are based on Machine Learning.

The decision-making process can therefore be divided into the following stages:

1. The machine analyzes and learns from all the provided data;
2. The machine makes autonomous decisions based on the analyzed data;
3. Finally, the machine optimizes previous decisions by automatically learning from its own mistakes.

Machine Learning (ML) requires large amounts of data to operate effectively and adapt to various situations. One application of ML in e-commerce is Dynamic Pricing, the practice of setting a product's price based on current market conditions.

ML can analyze data in real time and suggest different prices depending on the day of the week or the time of day, with the goal of maximizing profit. For example, the system might predict that raising the price of wine bottles on Friday evening and lowering it on Tuesday at noon could lead to higher overall profits (which, in this case, is our objective function).

How is this possible? Machine Learning analyzes customer data—such as demographic characteristics, purchasing behavior, and website traffic—allowing it to determine an optimal price that takes into account all the variables affecting profit maximization.

1.3 Deep Learning

When we talk about Deep Learning (DL), we refer to autonomous learning (note: no longer automatic learning). In addition to interpreting the variables introduced by humans into the system's input (as in Machine Learning), these algorithms can independently identify distinctive features without external categorization — that is, without any human intervention.
For this reason, DL requires significantly more data than ML and demands greater computing resources.

An example of Deep Learning is Natural Language Processing (NLP), which refers to algorithms capable of analyzing, representing, and understanding text as sequences of words that convey one or more messages in a language. NLP combines linguistic techniques with computational methods to enable machines to perform human-like tasks such as reading and understanding word meanings.

This technique can be applied in the food distribution sector to quantify the effect of product cannibalization. NLP can automatically assign mathematical values to each product by analyzing product descriptions and determining how contextually similar or dissimilar they are.

For instance:
A product described as “tuna in olive oil” would be contextually similar to “natural tuna”, less similar to “tuna sauce”, and completely dissimilar to “chocolate spread”.

By scientifically identifying these relationships, it becomes possible to determine whether a promotion on one product will have a positive, negative, or neutral impact on others.

1.4 Is a machine equipped with Artificial Intelligence conscious?

To answer this question, it’s essential to first define the concept of consciousness. Consciousness can be described as:

“The awareness individuals have of themselves and the external world they interact with, their own identity, and the complex set of their inner experiences.”

Thus, when we speak of a machine with consciousness, we mean a machine that is not only intelligent but also self-aware. It’s crucial to distinguish between knowledge and consciousness. The key question is: What makes the human brain — composed of a dense network of neurons — aware that a warning light in their car is on, while the car itself, a sophisticated piece of electronics and engineering, remains unaware?

This topic is highly debated and complex. For example, consider Deep Blue, the machine that defeated world chess champion Garry Kasparov. Some argue that although Deep Blue understood chess rules and knew how to make the right move at the right time, it wasn’t truly conscious. Every move it made was simply the result of programmed rules and algorithms. The machine lacks consciousness because, for example, it cannot autonomously determine what is right or wrong. Consciousness can therefore be seen as an ethical concept. A machine’s sense of right or wrong stems from how it was programmed, rather than from independent thought.

1.5 The use of AI in the food supply chain

The Artificial Intelligence market is growing by nearly 20% annually and is expanding into an increasing number of industries. While autonomous vehicles and voice assistants like Apple's Siri or Amazon's Alexa are well-known examples, there are many less widely recognized applications.

One notable example is the retail and production sector, where AI presents a significant opportunity for businesses thanks to its diverse applications.
A key application in this sector is optimizing supply chain management, which includes:

- Analyzing data from customers and suppliers;
- Warehouse management optimization;
- Demand forecasting.

Regarding the last point, AI can factor in numerous variables, including external factors such as weather conditions, holidays, and competitors' promotional changes — all of which can significantly impact demand fluctuations.

While it may seem intuitive that weather forecasts can influence the sales of certain products, what is less obvious is how to quantify this impact across an entire product range or how it affects order planning for producers.

Machine Learning algorithms can identify such relationships. For instance, it can predict how much beer will be sold if the perceived temperature is 30°C versus 10°C. Furthermore, it can also predict the impact on beer sales if a competitor introduces a promotion on the same brand or a different one.

Retailers, wholesalers, and producers possess vast amounts of data regarding transactions and customer interactions, collected both offline and online.

AI algorithms allow companies to leverage this data effectively, enabling faster and more accurate decision-making — a crucial factor in a sector that must manage millions of product information flows to align supply with demand.

On average, even a small company must handle over 400 million daily transactions from various sources, making it impossible for humans to analyze this data manually.

According to a McKinsey report, Artificial Intelligence — particularly Machine Learning — can reduce forecasting errors by 30% to 50% across the supply chain. Forecast accuracy is improved by combining hundreds of variables that

influence a single product's sales.
Since supply chain efficiency directly impacts revenue and profits, this technology offers considerable benefits for both retail and production companies.

For retailers, accurate demand forecasting enables:

- Increased revenue through improved product availability, reducing out-of-stock situations for both standard and promotional items;

- Warehouse management optimization, reducing handling costs in distribution centers and stores;

- Optimization of safety stock for fresh products, minimizing food waste while maximizing sales;

- Better use of shelf and warehouse space, allowing the introduction of new products by eliminating low-performing items previously used as substitutes;

- Product transfers between stores, resolving issues related to bulk purchasing and individual unit sales;

- Document reconciliation, aligning delivery notes with orders to identify inefficiencies such as unfulfilled orders. .

For production companies, implementing a demand forecasting system offers:

- Increased revenue by reducing unfulfilled orders and overstock in both the short and long term;

- Reduced uncertainty in customer requests by analyzing product trends to anticipate sudden sales spikes;

- Better raw material procurement, suggesting the optimal amount of ingredients required to fulfill orders;

- Better raw material procurement, suggesting the optimal amount of ingredients required to fulfill orders;

- Production optimization, improving efficiency by better organizing production lines and workforce;

- Warehouse management optimization, ensuring optimal stock levels for efficient deliveries;

• Improved inventory turnover, anticipating peaks in demand and reducing inventory holding costs.

An example of how AI can be transformative in retail is Walmart's Intelligent Retail Lab (IRL), a "smart store" designed to collect real-time information about in-store activities through multiple sensors, cameras, and processors.

In this setup, AI — combined with Machine Learning and Deep Learning — has introduced a new inventory management approach that ensures product availability on shelves in real-time.

Out-of-stock issues are a major concern for retailers, often resulting in lost sales, reduced annual revenue, and fewer returning customers.

To address this, Walmart's IRL is equipped with 1,500 ceiling-mounted cameras that use image recognition technology to monitor product availability.

Image Recognition, a Deep Learning technique, allows machines to recognize specific shapes through image analysis.

To achieve this, the system is trained with extensive image datasets featuring supermarket products, such as canned tuna. Initially, the system may misidentify items, but after repeated corrections, it achieves a highly accurate ability to distinguish what is a can of tuna and what is not. This ensures Walmart's IRL can effectively monitor product availability on shelves.

1.6 Why invest in Artificial Intelligence?

The rapid advancement of Artificial Intelligence is opening new frontiers in areas that were once considered futuristic. This progress is driven by:

• Increased computing power compared to previous years;

• Access to vast data repositories, which are essential for developing AI systems.

AI can perform complex analyses that would be nearly impossible to carry out manually. By processing massive volumes of data, AI can explore patterns, identify correlations, and extract useful insights to inform future predictions.

Due to these numerous benefits, companies are investing heavily in AI solutions.

The International Data Corporation (IDC) has forecasted that global spending on AI will reach $632 billion by 2028. This figure includes investments in AI-enabled applications, infrastructure, and related IT and business services.

1.7 How to identify an artificial Intelligence application?

Broadly speaking, Artificial Intelligence refers to machines capable of learning, reasoning, and acting autonomously. Currently, most AI applications in the spotlight fall under the category of Machine Learning. Therefore, the more accurate question is not “How to identify an AI application?” (which refers to a computer’s ability to perform tasks typically associated with intelligent beings) but rather: “How to identify a Machine Learning application?”

1.8 Difference from traditional statistical models

It is important to highlight the differences between Artificial Intelligence techniques and traditional statistical models. While both approaches aim to improve data understanding, they follow distinct methodologies.

Traditional statistical models have been in use for many years, predating the rise of computers. These models rely on small data samples and make assumptions based on prior knowledge. For instance, demand forecasting in traditional models often relies on past sales data, assuming that historical demand patterns are a reliable indicator of future trends.

However, in recent years, several additional factors have emerged that must be considered when predicting future product demand. These include:

New data sources such as holiday schedules, e-commerce data, and competitor information;
Improved computing power that is now more accessible and efficient, thanks to cloud services from major providers like Microsoft, Amazon, and Google.
These advancements enable businesses to produce continually updated sales forecasts that account for the increasing volatility in demand caused by factors such as the introduction of new products or promotional campaigns.

This is where Machine Learning plays a crucial role in supply chain planning. In environments where demand volatility is the primary challenge, creating an accurate forecast can be extremely difficult.

Unlike traditional models, ML can predict demand not only based on historical sales data but also by incorporating numerous other variables, such as:

- Weather conditions;
- Seasonal trends;
- Holidays and special events.

Ultimately, the foundation of an accurate demand forecast lies in the quality of the data. Therefore, ensuring data is precise and correctly extracted is essential. This is achieved through a structured set of techniques known as Data Mining.

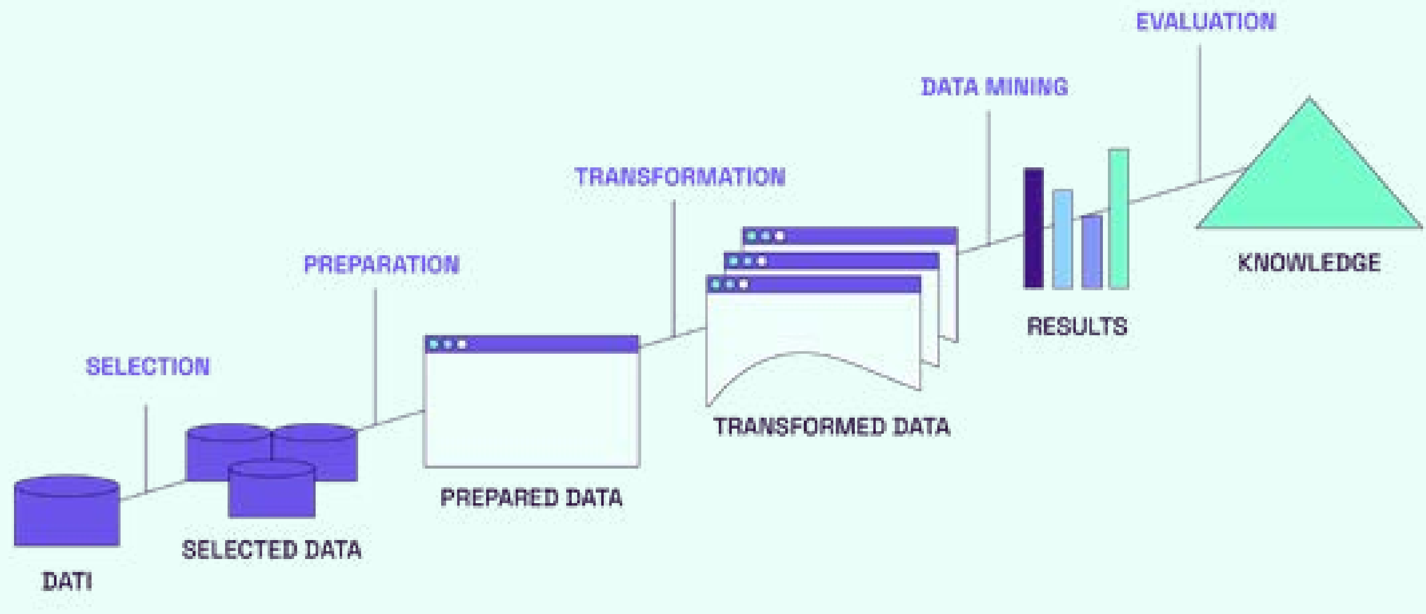
2.1 Data Mining: what it is and why it's important

Data Mining refers to a set of techniques and methodologies that allow the extraction of useful information from large databases. The extracted and processed data is then used to achieve predefined objectives, such as maximizing or minimizing a given target function.

In the retail sector, this process can help supermarket owners understand customer behavior. Information such as: purchase history of one or more consumers, demographic characteristics, time of purchase for certain products, are just a few data sources that can reveal correlations and uncover customer purchasing preferences.

In the food production sector, Data Mining can help producers improve product efficiency. For instance, production lines can be customized based on customer preferences by identifying popular products and adjusting supply accordingly. If the data shows that a particular type of bread is especially popular during a certain season, production can be increased during that period to meet demand.
Additionally, Data Mining allows businesses to group customers with similar characteristics into categories known as clusters. This enables tailored promotions for specific customer segments.

For example: A customer who purchases infrequently but in large quantities may receive different promotions compared to a frequent shopper who buys smaller amounts. This process of Data Mining, which ultimately leads to knowledge extraction, is a powerful tool for enhancing decision-making and optimizing business strategies.



The selection, preparation, transformation, and evaluation of data are essential phases for achieving maximum insight and being able to opt for the most optimal solution to a problem. As in the previous example, a solution to the problem of "most effective promotions" could be to understand which targeted promotions to carry out for a certain customer cluster in order to increase their loyalty. In the previous iteration, this represents the goal to be achieved.

When we talk about "raw data," we refer to information that has not been processed and may have little meaning if considered as such. Binary code is an example of this; taken only as a sequence of 0s and 1s, it holds little significance for a computer user, but when processed by a machine, it can provide more understandable information. For instance, the binary code "01000001" might seem meaningless to us, but for a machine, it corresponds to the uppercase letter "A."

Raw data, therefore, needs reworking, particularly because there may be missing information that needs to be integrated, or erroneous data that requires correction. Along the way, there is also a significant reduction in the volume of information, but an increase in its value, as it becomes more accurate and less prone to errors. In fact, initially, one must select raw data from enormous datasets, often generic, and after the Data Mining phase, useful patterns of information emerge that are essential for reaching the final decision—information that would have been impossible to obtain without this process.

2.2 The CRISP-DM methodology

As we've seen, Data Mining is a phase in the knowledge extraction process, and when it needs to be applied, there are several ways to do so. The Cross-Industry Standard Process for Data Mining (CRISP-DM) is the most commonly used analytical model for this purpose.

The CRISP-DM methodology consists of six main phases, which will be explained along with examples related to demand forecasting in the retail sector:

Business Understanding: This initial phase focuses primarily on defining the business problem or identifying the type of information required. The goal is to understand the problem and how its resolution will impact business decisions.

The question answered in this phase is: "What is the problem I want to solve?" For example: predicting the sales of promoted products more accurately through demand forecasting algorithms.

Data Understanding: This phase is crucial for the collection, comprehension, and evaluation of data quality. Specifically, it involves assessing the integrity of databases by identifying anomalies (also known as outliers) and gaps that need to be filled with external sources of data.
In this phase, relationships and trends between key variables are also identified (for example, the relationship between price and quantity sold).

The question answered during this stage is: “What data do we need?”
For a supermarket chain, we would consider data from receipts, which include the purchased products, the time of purchase, and promotional details, both current and past. For a food manufacturing company, the necessary data might include quantities of raw materials purchased, production times, and customer feedback. These data points allow for the analysis of production efficiency and help improve quality control.

Data Preparation: This phase encompasses all activities related to constructing the final dataset, such as data selection, data cleaning, the creation of new information, and any transformations. It is the most time-consuming stage of the process because data preparation operations are not immediate; they require careful study and in-depth navigation through the data. The primitive variables in the initial dataset may not be sufficient.
Through data engineering, these variables are combined and/or manipulated to create more elaborate variables that provide additional insights into the artificial intelligence model.

This phase answers the question: “Are the data sufficient, consistent, and accurate?” Data preparation is essential before providing the data to the model that will lead to the desired solution. If the data is insufficient, new data may need to be created, such as identifying the time of day when a particular product is usually sold or determining which day or weather conditions influence certain purchases. Clusters of similar customers can also be created, or correlations between products can be studied, for instance, if most customers who buy strawberries also purchase whipped cream.
This process begins the data preparation necessary for demand forecasting.

Modeling: This phase involves selecting and applying machine learning models that will allow the achievement of a predefined objective. Once identified once the type of models to be used has been identified, in order to maximize results, a parameter tuning phase is required for the model’s training.

During this phase, the model is trained with different parameters, yielding more or less satisfactory performance depending on the configurations. If, after modifying the parameters, the results are not satisfactory, it will be necessary to return to the phase of researching and creating new variables to feed into the model during the training phase.

This answers the question: “Which technique best fits the available data and will lead to the predefined results?”

In the retail and manufacturing sectors, the variables and data from the previous phase will be used as input for a demand forecasting algorithm that will rely on historical data as well as external variables such as weather conditions, holidays, and seasonality.

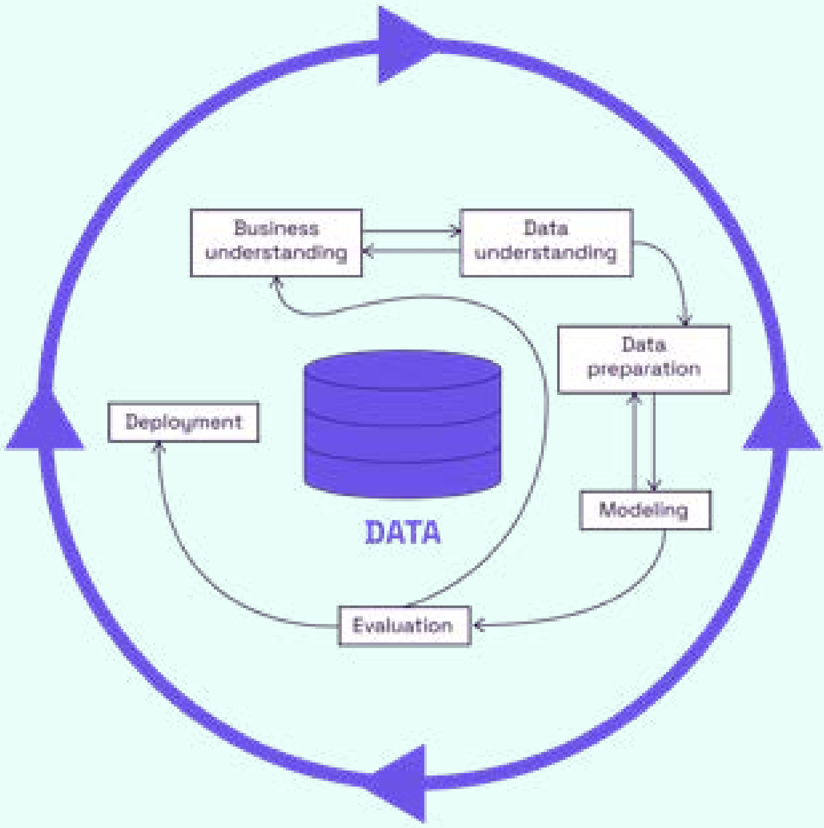
Evaluation: During the evaluation phase, the models are assessed and compared in order to select the one that best meets the business objectives and to determine whether there are any reasons why the model could be considered weak.

The question to ask in this phase is: “Do I get a meaningful response to my initial problem?” To determine if the model is appropriate, it is necessary to analyze the alignment of the data mining results with the business success criteria.

Deployment: The correct implementation of the model is essential to ensure its effective integration into business processes.

This answers the question: “How will the model be deployed?” In this phase, planning and monitoring of the results’ dissemination take place first, followed by the preparation of a final report for project review.

The difference between this model and other typically linear and unidirectional processes is that the phases characterizing CRISP-DM are designed to be repeated cyclically.



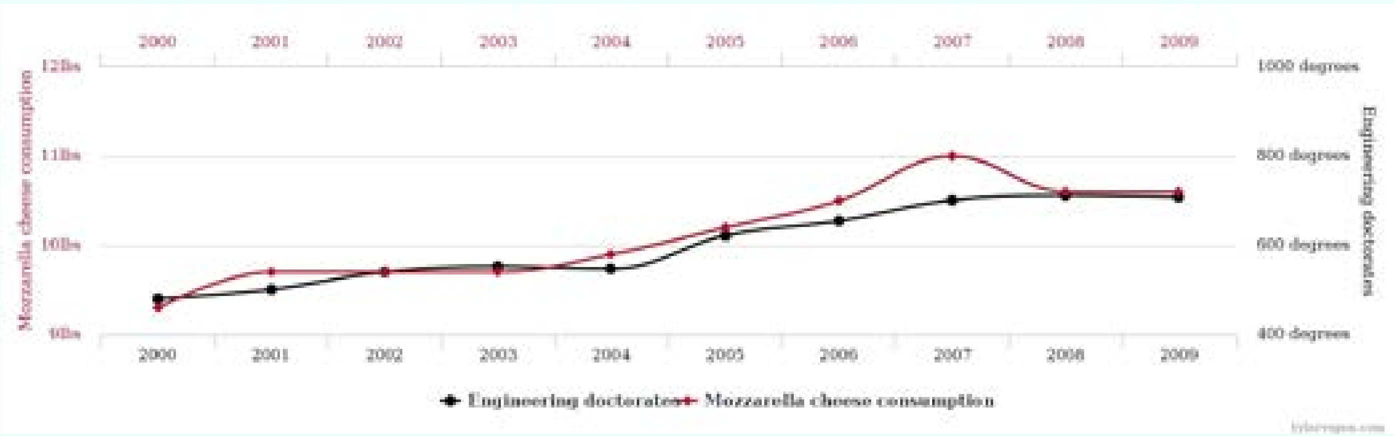
In this way, for example, if it becomes clear that there are no valid models in the "Modeling" phase, it is possible to go back to the "Data Preparation" phase and check whether it is necessary to select a different subset of data or if there is missing information. In a linear or unidirectional model, on the other hand, the phases move linearly in one direction. For example, constructing a building follows very specific steps. It is not possible to build the structure and then redo the foundation. Mining projects are inherently uncertain, as they are closely dependent on the quality of the provided data.

Going back and integrating are fundamental elements of a mining process, in stark contrast to a waterfall engineering process.

2.3 Data Mining and Artificial Intelligence

It is important to emphasize the significance of Data Mining when discussing AI. The data obtained through this process serve as input to the machine to achieve predefined goals.
Therefore, it is essential that the selected data be as accurate as possible, as incorrect data can only lead to incorrect results. This phenomenon to avoid is referred to as a “garbage in, garbage out” scenario, where incorrect input data (garbage in) leads to incorrect output results (garbage out).

At times, it is also necessary to distinguish between causality and coincidence. Not every piece of information that seems related to the issue at hand necessarily makes sense to be so. An example of this phenomenon is spurious correlations, where the machine, relying on inaccurate data, identifies correlations where there are none, referring to meaningless information. Let’s take a look at the example in the figure:

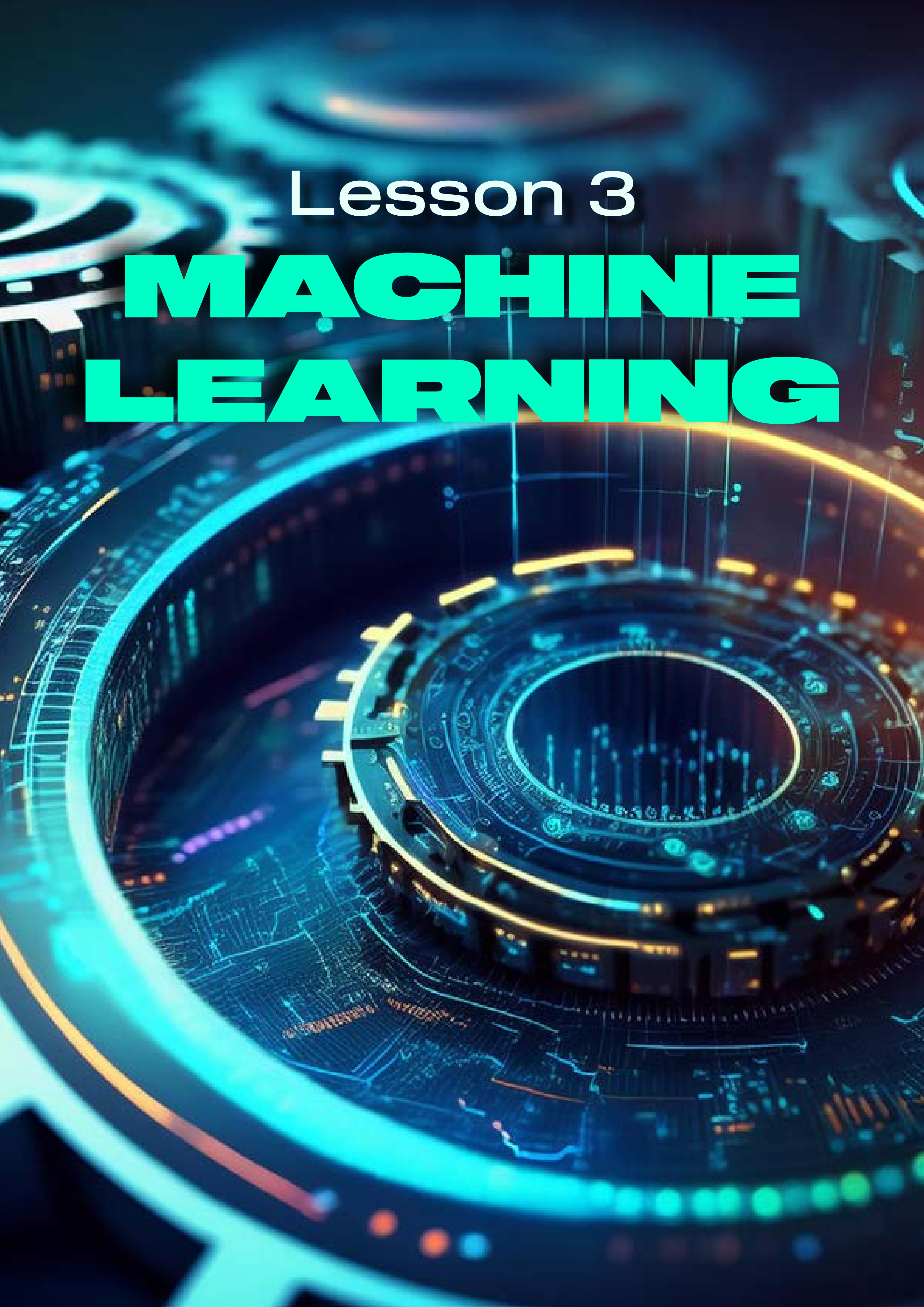


In this graph, a correlation is identified suggesting that people with a PhD in engineering buy more mozzarella.
It is evident that this is a misleading association that highlights the importance of having accurate data, as if the goal is to predict future sales demand, it is crucial not to rely on erroneous correlations like this one. Such correlations could lead to reordering certain products— in this case, mozzarella—based on baseless information.

We have therefore seen that, in order to discuss Artificial Intelligence applied to predictive models, accurate data is essential (quality over quantity), achieved through the proper use of the Data Mining process. We also saw in the first lesson that, in retail and manufacturing companies , Machine Learning is necessary to make accurate demand forecasts, but what exactly is it?

Lesson 3

MACHINE LEARNING



INDEX

3.1 Machine Learning

Machine Learning (ML) is a subset of AI that focuses on creating systems that learn or improve their performance based on the data they use. Each of us uses Machine Learning on a daily basis. One example of its application is speech recognition, which is featured on many smartphones, allowing users to activate commands via voice. Another use is enabling companies to create targeted advertisements through user profiling. This profiling allows them to collect interests and needs expressed by each user in their internet searches and while visiting web pages. If a user spends a lot of time on a webpage about soccer, it is highly likely that the next ads they will see on social media will be related to soccer shoes or sports items in general.

Data Mining and Machine Learning are closely related, as the first is essential for extracting accurate information and data that will serve as input to the latter to gain deeper knowledge and make more accurate predictions.
Depending on the type of algorithm used to enable machine learning—i.e., the ways in which the machine learns and accumulates data and information—different ML systems can be distinguished.
Two of the main types of algorithms currently used are those characterized by supervised learning on one hand, and unsupervised learning on the other.

- **Supervised ML:** Supervised learning is a technique that generates predictions using models trained on “supervised” data, where each input is associated with an ideal output that the model will try to replicate. The algorithm learns from a pre-labeled dataset with a predefined output. When we talk about labeled data, we refer to data that has been given a kind of virtual label, indicating certain properties or characteristics of that data. We talk about supervised ML when, for example, it is necessary to train an algorithm to predict whether a customer will stop purchasing or remain loyal based on purchasing behavior. The AI system trains on a dataset where the input is the customer’s behavior (recurrent purchases, number of products bought, satisfaction, etc.).
Each training row is based on real behavior, manually labeled (stop purchasing/loyal customer).

- **Unsupervised ML:** This type of algorithm uses a more independent approach, where a computer learns to identify processes and complex patterns without the constant and attentive guidance of a person. Unsupervised ML involves a training phase based on data that lacks labels and for which no specific output has been defined.

For example, we might input into the system a training dataset containing the characteristics of 5 plant species without classifying the species. In this case, the algorithm automatically generates a classification model based on the common characteristics it identifies and groups the plants based solely on the information it extracts from the data, without human intervention in the grouping process.

3.2 Classification, Regression, and Clustering

In supervised learning, the types of problems that can be solved are divided into two main branches: classification and regression.

Classification aims to solve problems whose outcome is finite and limited to a set of responses. Examples of classification problems include:

• Based on a set of historical banking data, will the customer be insolvent? -> (yes/no)

• Based on a series of aesthetic and shape-related properties, is this plant a carnivorous plant? -> (yes/no)

• Based on a set of climate, surrounding habitat, flora, and fauna information, which continent does a certain person live on? -> (Europe, Asia, Africa, America, Australia)

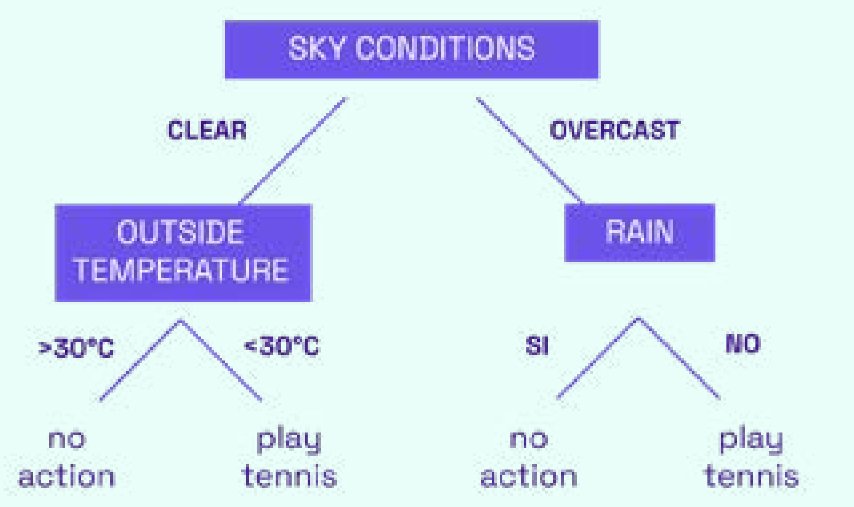
One classification method is the “Decision Tree,” which is common in supervised Machine Learning because it is fast and easy to interpret. In this algorithm, the input data is continuously divided into groups based on certain criteria. It is based on:

• **Nodes:** where the division occurs.
• **Leaves:** These are the intermediate or final results, where the data ends up after being separated.

The goal is to divide the initial dataset into groups, which should be as homogeneous as possible internally (similar characteristics among the components of each group) and homogeneous within each group (similar characteristics among the components of each group) and heterogeneous between the groups (dissimilar characteristics among the various groups).

To better understand this process, it’s helpful to consider an example.

Suppose we need to decide whether to play tennis based on the weather and the outside temperature. This problem, naturally, has only 2 possible outcomes, but it's necessary to evaluate the combinations of the considered variables to determine whether to play tennis or not.



The nodes are the points where the divisions occur, considering situations such as whether it is raining or not, while the leaves are the results where we find the data after a division, as seen at the final level. Looking at the situation on the left, we can notice that if it is sunny, with a temperature below 30 degrees Celsius, it's okay to play tennis. However, if the temperature is above 30 degrees Celsius, it is not recommended to play tennis. This is clearly a very simple example of a decision tree, while Machine Learning-based procedures involve hundreds of variables and outcomes derived from thousands of divisions of the tree and different levels of depth. This highlights how making accurate decisions on much more complex problems is impossible if relying solely on human reasoning, which is why it is useful to rely on advanced technologies.

Regression algorithms, on the other hand, answer questions that do not have a finite value as a response. Examples of regression problems include:

- In relation to inflation and market prices of wheat, how much will pasta cost per kilogram next month? -> (1€, 2€, 3€, etc.)
- What is the maximum speed of a car with specific characteristics? -> (220 km/h, 220.5 km/h, 220.566 km/h, etc.)

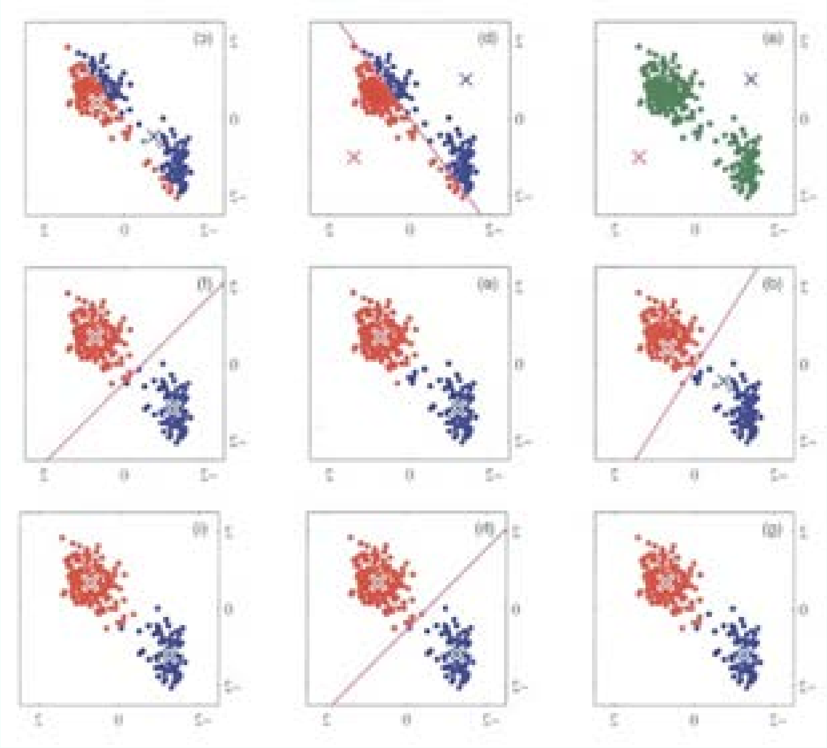
In this sense, sales forecasting is also a regression problem, and the problem can be formulated as follows:

• In relation to prices, historical sales, competitor prices, exogenous events, and relationships with other products, what is the sales forecast for tomorrow for a specific product? -> (1 unit, 2 units, 3 units, etc.).

Some of the most popular unsupervised techniques used in retail are definitely **clustering** algorithms, which automatically identify similarities between observations and divide them based on common characteristics. These groupings are called "clusters."

For example, consider the online customers of a specific e-commerce platform selling sports items. Each customer has personal characteristics and purchasing behaviors such as gender, age, frequency of purchase, etc. Using a clustering algorithm, it is possible to group customers in order to offer them special promotions based on their purchasing behavior. A priori, the purchasing behavior is unknown, and it is not possible to know if they are customers who practice sports at a professional or amateur level, or whether they are occasional or loyal customers. It will be up to the algorithm itself to discover whether there are patterns in customer purchasing behaviors that allow the definition of consumer groups.

K-means is one of the most widely used and efficient clustering algorithms. The goal of this algorithm is to group similar observations, discovering hidden patterns, and dividing the dataset into a number k of clusters. A cluster is thus a set of observations that share similar characteristics. Once the desired number of k clusters is defined, the algorithm will group the observations it is trained on.



In the image, you can see how during the learning process, the algorithm progressively tries to divide a set of observations more and more distinctly. By the ninth iteration (bottom right), the algorithm has outlined two distinct groups of observations that are easily distinguishable.

An extraordinary feature of Machine Learning is its predictive ability. In the past, business decisions were often made based on historical results, whereas today, as we have seen, this technology allows the use of much more advanced analyses that take into account many more factors, such as external variables.

As analyzed in the second lesson, which presented the CRISP-DM methodology, the problem was predicting sales during a promotion (page 8), and what was being sought was a predictive analysis. However, if we had to solve a problem related to purchase or product restocking decisions, this would not be sufficient. In fact, in order to achieve this goal, it is necessary to introduce the concept of prescriptive analysis.

Lesson 4

DIFFERENCE BETWEEN PRESCRIPTIVE AND PREDITTIVE ANALYSIS

INDEX

4.1 Difference between predictive and prescriptive analysis

A predictive model analyzes current and historical data to make forecasts, such as the probability that an event will repeat in a given situation. For example, it is possible to predict, based on historical data and external variables, how many kilograms of bananas will be sold next week in a specific store depending on the weather conditions. But is this enough to maximize the retailer's profit? No, because there are many other variables that need to be considered to determine the correct amount of product to reorder, such as stock levels and supplier constraints.

For this reason, retailers should adopt prescriptive analytics, which, based on predictive results, allows the system to provide detailed recommendations for decisions that maximize the company's economic performance. An example of this in production: **prescriptive vs predictive analysis**.

Through Machine Learning, as we have seen, it is possible to access a wide range of internal and external data, uncover previously unknown patterns and correlations at the level of individual stores, products, or shelves. Then, it is possible to optimize the results to provide weekly or even daily recommendations, such as adjusting prices, promotions, or assortments in each store to maximize profits. According to McKinsey data, prescriptive analytics can increase sales by 2-5%.

But what is missing in Machine Learning to achieve this result? Many other constraints need to be considered, such as lead time, minimum order quantities, coverage stock, pallet constraints, value-based discounts, etc., in the case of a retailer or wholesaler.

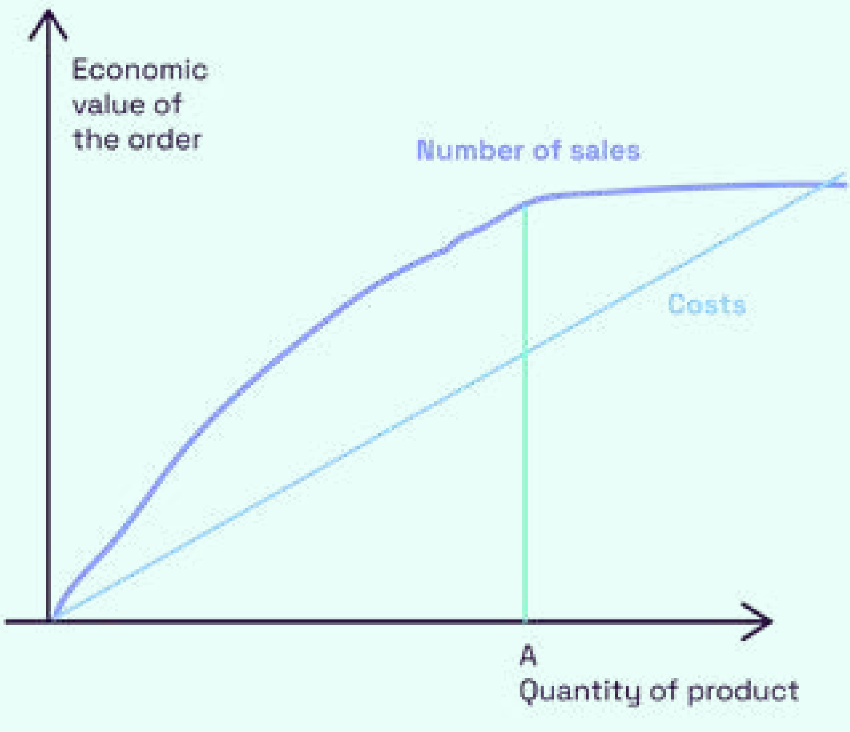
To account for all these variables simultaneously, mathematical optimization algorithms are needed. These algorithms cross-reference sales forecasts with the constraints listed above, selecting the option that maximizes the difference between revenues and costs, i.e., the margin.

This is the goal of prescriptive analytics: to analyze various situations and suggest the most economically profitable one.

The main outcome of prescriptive analytics in the retail and production sectors is undoubtedly the ability to make more economically sound decisions. However, the consequences of adopting this approach are numerous. For instance, in retail, it can help determine whether to keep low-selling products on the shelf because they attract the best buyers, or remove them to introduce new items that could increase the profitability of specific store sectors. This is something that predictive analytics alone cannot do, as it is limited to forecasting only the future sales of products.

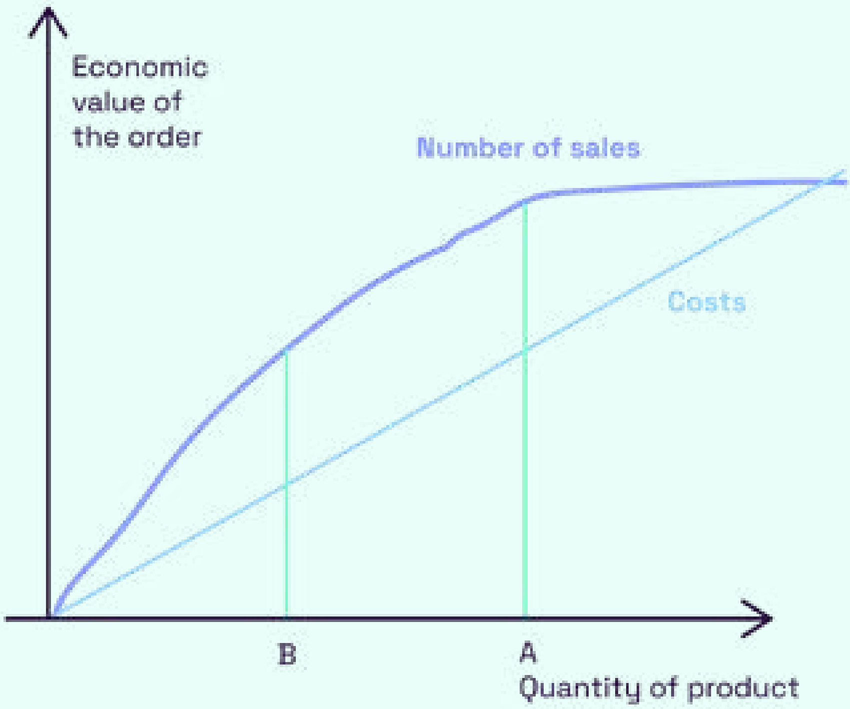
Another application of this approach is the recognition of various patterns to recommend the timing of promotions and price changes for products. These correlations, however, are not easy to identify, and the variables to consider are always complex and varied. For example, in some cities, avocado and potato chip sales might increase by 5% on the Friday before a typical sports event, but a prescriptive tool might recommend stocking 20% more of these two products in stores where local teams are more likely to reach extra time.

This analysis can thus recommend specific assortment changes and estimate the monetary value of these changes. In the supply chain, it becomes clear how important prescriptive analytics is for food industry companies, as the recommendations provided aim to determine the reorder quantity that maximizes profit by finding a balance between the maximum achievable sales and the minimum sustainable costs. To better explain this concept, it may be helpful to refer to the following graph:



LES. 4: DIFFERENCE BETWEEN PREDICTIVE AND PRESCRIPTIVE ANALYSIS

One might be tempted to think that the optimal quantity of products to order/produce would correspond to point A, as it represents the maximum sales. However, a model based on prescriptive analytics might suggest that the ideal quantity is actually at point B, because it corresponds to a lower amount of product but also to proportionally lower costs. This means that what is being sought is not the maximum sales quantity, but rather the quantity of product that maximizes the difference between revenues and costs.



This happens because sales are not the only determining factor; other factors must also be considered, such as the increase in product quantity, which is not directly proportional to the increase in sales but can lead to a higher risk of overstock. Therefore, the goal for companies is not simply to improve forecast accuracy, but rather to aim for more economically efficient decisions. Some of the key performance indicators (KPIs) that can be achieved include:

- Reduction in inventory surplus by up to 46%
- Increase in sales efficiency by 18%
- Reduction in warehouse inventory holding costs by -8%

Because, as we always say, the future belongs to those who can anticipate it!