ManagementSolutions
Making things happen

Machine Learning,
a key component in business model transformation

Design and Layout
Marketing and Communication Department
Management Solutions

Photographs
Photographic archive of Management Solutions
iStock

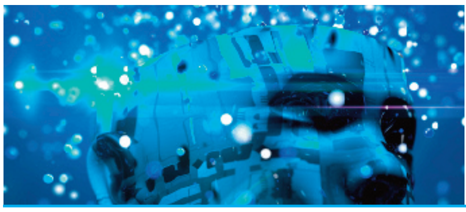
Contents


Introduction 4


Executive summary 8


The digital revolution and the transformation of business models 12


Machine learning concept and trends 18


Quantitative exercise 30


Bibliography 38

Glossary 40

Introduction

I believe that at the end of the century the use of words and general educated opinion will have altered so much that one will be able to speak of machines thinking without expecting to be contradicted

Alan Turing[1]

The digital revolution is resulting in deep changes in consumer habits, caused, among other factors, by greater access to information and increasingly developing new technologies. All this invites us to take an in-depth look into the business models currently being used.

A fundamental driver of business model transformation is data science, which is based on the combined use of machine learning techniques, artificial intelligence, mathematics, statistics, databases and optimization[2], a concept which has already been extensively covered in a previous Management Solutions publication[3].

Several factors, mainly technology related, promote Data Science techniques to be used across many different areas. These factors can be grouped around the following four themes: (i) the unprecedented increase in the volume and type of data available, (ii) data connectivity and access, (iii) improvements in algorithms, and (iv) the increased computational capacity of systems.

Regarding the volume of data, several studies show metrics that raise awareness of the magnitude of such growth. Some of the most relevant are as follows:

- According to recent reports, 90% of all data created in the history of humanity was produced during the last year, while 40% annual growth is expected to occur over the next decade[4]. The volume of available data today is even higher due to developments in both communications, known as Machine to Machine (M2M), and the so-called Internet of Things (IoT).

- Studies published by large telecommunications companies[5] highlight that the number of devices connected to the Internet will be more than 3 times the world population in 2021 and the number of IoT connections will reach 13,700 million that year, compared to 5,800 million in 2016.

- As a result, by 2020 the total volume of existing data will reach 44 trillion gigabytes[6].

- Of these, a large amount of data is directly generated from the digital environment, as it is the case with Google searches (40,000 searches per second), Facebook messages (31 million messages per minute) or data from videos and pictures (300 hours of video uploaded to YouTube every hour).

- By 2020 it is estimated that, all mobile devices will include biometric technology[7]. It is also espected that, by that year, at least a third of the data will be sentthrough the cloud[8].

Secondly, connectivity improvements represent a qualitative jump that leads to new services and business models being developed based on real-time data generation and analysis in order to adapt services and/or price according to usage: data are generated and collected automatically through sensorized and digitalized point-of-sale terminals, which creates a continuous flow of information. Much of this connectivity takes place between machines: once an action is performed, the data generated by the different digital components involved are connected to servers in order to store and analyze the information. This type of M2M connection has increased to reach 1,100 million connections in 2017[9].

Thirdly, improvements in algorithms have made it possible to optimize the processing of large data volumes (through scaling, resampling, etc.) as well as to obtain more efficient and robust methods and to process missing data, non-numerical variables and outliers. Despite the fact that most algorithms were developed before 2000, companies are now making a major effort in implementing these algorithms, achieving better results than those produced by humans. To provide a few examples:

[1]Turing, A.M. (1950). Mathematician considered the father of computer science. He deciphered the Enigma machine during World War II. He was the forerunner of modern computing and artificial intelligence.
[2]Dahr, V. (2013). Professor at Stern School of Business and Director at Center for Digital Economy Research, New York.
[3]Management Solutions (2015).
[4]Ministry of Industry, Energy and Tourism. Government of Spain (2018).
[5]Cisco (2017).
[6]Forbes (2015).
[7]Acuity Market Intelligence (2016).
[8]Forbes (2015).
[9]Statista (2017).

- DeepMind AlphaZero and AlphaGo algorithms play chess and go games at a level beyond what humans are currently capable of.

- An algorithm based on artificial intelligence can detect breast cancer 30 times faster than a doctor and with 99% accuracy[10].

- In the United States, roboadvisors[11] have 25.83 million users, representing a market penetration rate of 1.8% in 2018. This rate is expected to reach 8.3% in 2022[12].

Finally, improvements in computing capacity, which have been huge over the last few decades due to advances in processor technology, are now being driven by other key factors such as the significant development achieved in programming languages (both those of a general nature and those used in data processing, visualization, algorithms, etc.), cloud computing, and especially the design of new computing architectures specifically aimed at machine learning tasks, data analysis and engineering applications (known as GPUs[13]).

In summary, over the last two decades the availability of digital data has increased almost 1,000 times, while algorithms have become 10 times more efficient and computing speed has increased 100 times[14]. All this has led to a renewed interest in these techniques as a formula for adding value to information in the new business environment.

Machine Learning: more than half a century of history

Machine learning techniques are experiencing an unprecedented boom in a number of fields, both in the academic and business world, and are an important lever for transformation. While these techniques were known in both areas, several factors are leading to their use becoming more widespread where it was previously minority based, and are causing it to extend to other fields where they were hardly used before due to both their high implementation costs and the small initial profit expected from their implementation.

Machine learning techniques can be defined as a set of methods capable of automatically detecting patterns among data[15]. Under this definition, the concept of machine learning has existed at least since the 1950s, a period of time in which various statistical methods were developed, redefined and used in machine learning through simple algorithms, though almost exclusively within the academic field.

[10]Forbes (2016).
[11]Automatic algorithms that provide online advice and management with minimal human intervention.
[12]Statista (2018).
[13]Graphics processing unit.
[14]Brynjolfsson, E. y McAfee, A. (2017). Brynjolfsson, professor at the MIT Sloan School of Management. Director of the MIT Initiative on the Digital Economy, Director of the MIT Center of Digital Business. He is known for his contribution to the world of IT Productivity. McAfee, co-director at MIT Initiative on the Digital Economy and Associate Director at Center for Digital Business.
[15]Murphy, K. (2012). Postdoc at MIT, professor in Computer Science at Univ. British Columbia (Canada). This textbook was awarded the 2013 DeGroot Prize. He is currently a research scientist for Google on Artificial Intelligence, Machine Learning, computer vision and natural language understanding.

This concept of machine learning involves since then the use of identified patterns to make predictions, or to make other types of decisions in environments of uncertainty[16].

Machine Learning techniques represent a step ahead from classical statistical techniques in the sense that they enhance the model estimation process, not only because they increase predictive power by using new methodologies and techniques for selecting variables, but also because they lead to improved process efficiency through automation.

Within this context, the present study aims to provide insight into the digital revolution and its impact on the transformation of business, with a special focus on machine learning techniques. For this purpose, the document is structured in three sections, which correspond with three objectives:

- Illustrate the development of the digital revolution and its impact on different fronts.
- Introduce the machine learning discipline, describe different approaches and outline current trends in this field.
- Present a case study to illustrate the use of Machine Learning techniques in the specific case of the financial industry.

[16] Ibid.

Executive summary

The digital revolution and the transformation of business models

1. The paradigm shift introduced by digitalization invites us to review current business models.

2. Mobile and the Internet as a gateway to online services, social networks as a source of data, artificial intelligence and big data architectures, distributed computing infrastructures and the use of cloud applications, the development of blockchain technologies, the use of cryptography and biometrics, the so-called Internet of things, robots and virtual assistants, 3D printing and virtual reality, etc. All of those are useful tools available to companies. They, on one hand, change their environment and, on the other they, drive business model redesign in response to this new environment.

3. In the digital age, data becomes a clear source of value. Its volume increases exponentially due to the digitalization of processes and the growing interaction with customers, employees and suppliers through digital channels.

4. Other areas that are increasingly playing a more prominent role are: the customer experience, product and service customization and the "access versus ownership" model. At the same time, new competitors are entering the market with a technology-based value proposition.

5. The fact is that, the introduction of the digital dimension changes everything for companies: It effecs impacts their strategy definition; governance, organization and work culture; business and operational processes; data access, storage, processing and modeling; risk management and control (with new risks arising in areas such as cybersecurity, personal data protection and ethics in relation to artificial intelligence); and regulations themselves, which become more global.

Machine learning concept and trends

6. One of the cornerstones of this digital transformation is Machine Learning: the use of techniques to build algorithms that will learn from new, different sources of information and improve autonomously with experience. This results in methods capable of automatically detecting data patterns that can be used to predict future data in an environment of uncertainty.

7. The main machine learning components fall under one of four categories:

   - Information sources that can provide both structured and unstructured data, which are the basis of the remaining components.

   - Techniques and algorithms to process unstructured data (text, voice, video, etc.) and obtain patterns from the data.

   - Self-learning capacity, allowing the algorithm to adapt to changes in the data.

   - The use of systems and software as a vehicle for data visualization and programming.

8. Improvement in these areas represent an step forward with respect to the traditional modeling approach. These developments require, among other things, the use of a greater number of different information sources, the ability to detect patterns hidden in the data using inductive methods, the maintenance of predictive power for longer, as well as the need for greater storage and data processing capacity. Selecting the most appropriate machine learning technique from among the many available will largely depend on these components.

9. Some techniques can be used to transform unstructured information (texts, sounds, images, etc.) into data that can be analyzed and processed by a computer. These techniques include the use of statistics or the classification

of words into categories in order to understand the written text, the use of neural networks for voice or image recognition, the use of Markov chains for the construction of texts in natural language, or the use of unsupervised classification algorithms for image structuring.

10. Modeling techniques that are used with structured information are classified as supervised or unsupervised learning depending on the information used for learning. In supervised learning, there is a target variable observed in the data, while in unsupervised learning the aim is to find patterns or relationships within data.

11. In the modeling process, there is usually a stage involving the organization and preparation of the initial data set in order to facilitate and optimize its subsequent use. This requires descriptive analysis and data preparation tasks (including specific regularization techniques such as the use of elastic net).

12. Some of the most widely used techniques for supervised learning are neural networks (including their deep learning extension), support vector machines, Bayesian classifiers, or regression and classification trees.

13. These techniques can be used together with algorithms, called ensemble methods, which improve predictive capacity by combining models in order to produce a more predictive or more stable model.

14. Unsupervised learning includes clustering techniques and Data Analysis techniques, such as the reduction of dimensionality method.

15. The use of these techniques requires more sophisticated methods for validating results need to be more sophisticated, as it is the case with bootstrapping and cross-validation, which allow the model to be analyzed in more than one validation sample. Also, the dynamic nature of these techniques makes it difficult to trace the models.

16. Machine learning techniques are implemented across many different areas. Examples include their use in education (e.g. smart tutors), finance (e.g. automatic trading, roboadvisors, fraud detection, risk measurement, and development of prospect models for commercial purposes), health (image diagnosis, treatment consultation and recommendation management, collection of medical information and robotic surgery), and also in areas that cut across industries such as efficiency improvement in organizations (through, for instance, improving the IT function).

17. Some of the main trends in the implementation of Machine Learning techniques include the use of information sources that collect data in real time, the heavier value placed on high predictive power to the detriment of model interpretability, incorporating the algorithm's ability to self-modify based on the changes that are taking place in the target population, or the investment in IT architectures and infrastructures that ensure scalability in data storage capacity and greater processing speed, combined with Cloud-based solutions and the use of edge computing infrastructures offering preinstalled functions prepared for direct use (Functions as a Service).

18. All this entails a number of challenges and dificulties. Using new sources of information requires investment in order to identify new sources of relevant data, the incorporation quality techniques, the ensurance that the information is protected, or establishment cybersecurity systems and controls.

19. The use of new techniques and algorithms means reinforcing the model risk management function, selecting the most suitable algorithm in advance from a wide range of possibilities, or monitoring that automation does not result in discriminatory processing. Incorporating the self-learning feature in models can make it difficult to validate them and requires the consideration of new approaches such as controlling the degrees of freedom in automation, more frequent evaluation of the discriminatory power, or other contrast techniques. The solution lies in investing in IT resources and infrastructure.

20. The introduction of machine learning in organizations requires having highly capable and specialized staff, able to understand and work with programming languages, algorithms, mathematics, statistics, and complex IT architectures.

Quantitative exercise: use of machine learning techniques in the construction of a scoring model

21. The objectives of the proposed exercise were to: (i) analyze and illustrate how the use of Machine Learning techniques impacts model development, and (ii) evaluate how the estimation process and the results obtained vary through the use of Machine Learning techniques.

22. To this effect, a behavior scoring model was trained using different Machine Learning algorithms, and the results were compared with those obtained using traditional techniques.

23. The analysis was carried out using a set of over 500,000 loans spanning over 10 years, with a 6% default rate. The sample included information on both the loan and the customer, as well as other information that can be useful in modeling, such as commissions, number of transactions made, etc.

24. During an initial knowledge discovery stage, missing values were processed (using allocation techniques based on methods such as clustering algorithms or regression models), outliers were analyzed and processed, and variables were simplified and grouped together (again using clustering algorithms). This made it possible to reduce

the number of variables in order to improve the efficiency of subsequent processes and to prepare the existing information in order to adapt it to the specific requirements of the different models and possible limitations of the algorithms.

25. A second phase different models have been estimated: a traditional model (logistic) that provides a comparaison, and five Machine Learning techniques: a model with regularization techniques (elastic net), two ensemble methods (a random forest and an adaboost), and two support vector machines (using a linear and a radial function, respectively). Subsequently, two measures of the discriminatory capacity were calculated on a validation sample: the model success rate (based on an estimate of the cut-off point) and the area under the ROC curve.

26. After the comparison with these statistics, it was observed that three of the analyzed techniques improved the model's predictive power in terms of both success rate[17] and discriminatory power: the random forest, the use of the elastic net, and the adaboost.

27. In particular, the random forest was the one that produced the best results compared to the other methods. In terms of success rate, it achieved 80.3% against the 74.7% obtained from the traditional model - a percentage increase of 7.5%; and in terms of discriminatory power, the area under the ROC curve improved from 81.5% in the traditional model to 88.2% - a percentage increase of 8.2%. The second best method was the use of the elastic net, which resulted in success-rate and area under the ROC curve values of 79% and 86.4% respectively - a percentage increase of 6% for both indicators.

28. This increase in predictive power implies that, business volume being equal, the non-performing loan rate would be reduced by 48% if the random forest method were used (using an optimal cut-off point with the traditional model would result in a non-performing loan rate of 2.1%, reduced to 1.1% by using the random forest method provided the number of approved loans was maintained), and by 30% if the elastic net were used (down to 1.4 %). Likewise, the non-performing loan rate being equal, business volume would increase by 16% with the random forest, and by 13% with the elastic net.

29. These improvements, however, were achieved through greater complexity with respect to the traditional model, having used a total of 80 variables between all 50 trees making up the forest in the case of the random forest, and a total of 45 variables after using the elastic net, compared to 11 variables in the traditional model.

30. Therefore, as it was observed in the analysis, Machine Learning techniques improve the discriminatory power of models, which leads to improvements in the business while at the expense of greater complexity both from a statistical viewpoint and in terms of the volume of information used. Analizing the interpretation of results is therefore more difficult and requires. This a poses greater difficulty when analyzing the interpretation of results and requires strengthening the validation procedures (for instance by using replicable models that challenge Machine Learning models and explaining the differences in the outputs of both models).

[17] Percentage of loans correctly classified by the model, defined a cut-off point.

The digital revolution and the transformation of business models

Technology doesn't provide value to a business. Instead, technology's value comes from doing business differently because technology makes it possible

George Westerman[18]

For some years now, there have been many initiatives to transform the processes and systems used by companies, which ultimately impacts the way work is done. These initiatives have taken place in both the private and public sectors, and are based on a combination of physical infrastructures and novel elements from digital technology and biotechnology developments.

These changes are not just a more sophisticated continuation of the current production systems, but also represent a new paradigm, the so-called fourth industrial revolution. This paradigm is based on radical changes in production linked to the speed of implementation, which has no historical precedent, and on the scope of the transformation, which affects a large number of elements in the value chain[19].

Within this industrial revolution, the concept of digital transformation brings together those initiatives directly related to digital technologies. According to some authors[20], digital transformation can be defined as:

> *"The use of new digital technologies (social media, mobile, analytics or embedded devices) to enable major business improvements (such as enhancing customer experience, streamlining operations or creating new business models)"*

These digital technologies have a direct impact on processes in terms of speed, security and transparency, which allows the development of new services that did not exist before due to lack of markets or profitability, since digital technologies usually have few barriers to entry, and their marginal costs tend to zero.

They also affect service intermediaries by correcting, at least partially, some market imperfections such as asymmetric information, transaction costs or asymmetries in the matching of supply and demand, through the use of B2B or P2P[21] platforms with minimum involvement of third parties.

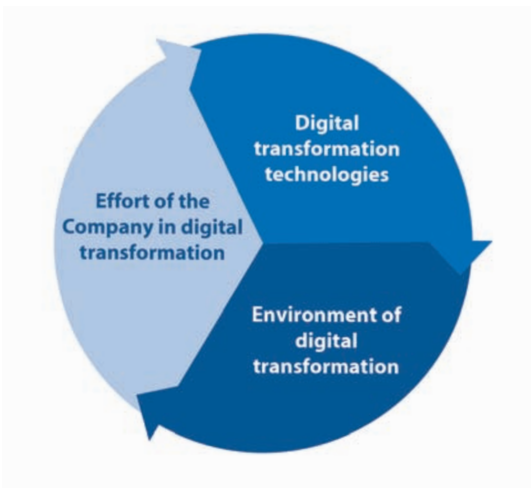
[18]Westerman, G. (2017). Researcher at MIT Initiative on the Digital Economy and Co-author of the best seller *"Leading Digital: Turning Technology Into Business Transformation"*
[19]Schwab, K. (2016). Economist and German businessman, known mainly as the founder of the World Economic Forum
[20]Fitzgerald et al. (2013). Editor and contributor on digital transformation at MIT Sloan Management Review.
[21]B2B: Business to Business. P2P: Peer to Peer.

However, data management has become a fundamental asset for companies, and it is a raw material and a business generator at the same time. The digitalization of processes and the interaction with customers, employees and suppliers through digital channels, provides a huge amount of information that can be used for new opportunities. By way of illustration, according to the European Commission the value of the data economy in the European Union could reach 4% of GDP in 2020 (more than double with respect to the current situation).

It also involves a paradigm shift in the labor market, since it requires more qualified tasks, which implies the need to make a greater investment both in the educational system and in the ongoing training plans of companies.

However, as will be discussed later, digital transformation does not refer to the new applications and uses of new technologies, but to how these technologies change the environment in which companies operate and how they must adapt to this new environment (Figure 1), using precisely these technologies in processes and systems, but also changing their organization and strategy.

Figure 1. Digital transformation elements.



Digital transformation technologies

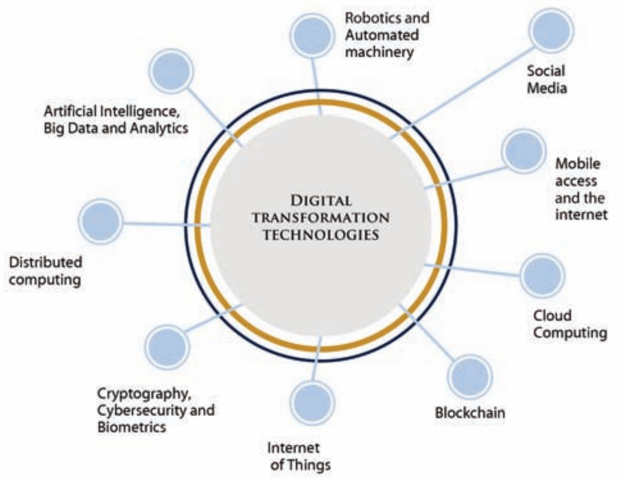
Business transformation, which encompasses a number of areas and proposals, is based on technology developments that, while having different origins and nature, share some common elements: their strong leverage in investment, the production and use of large data volumes, and the goal of improving the customer experience and operational efficiency. Some of the main technologies that drive transformation are the following: (Figure 2):

- **Mobile and internet access:** the prevalence of telecommunications and access to a wide range of online services is leading companies to develop application programming interfaces (APIs), which enables companies to have new channels of dialogue with customers as well as a broad data set. This leads to the appearance of new relevant forms of information, such as geolocation. Likewise, mobile access has allowed universal access to information and communication, regardless of physical location. As an example, mobile broadband communications have multiplied fivefold in the developed economies over the past ten years[22].

- **Social Media:** social networks are a source of data that provides a better understanding of customer behavior and greater productivity through improved internal communications. Trends for Social Media suggest a much more segmented and directed marketing as well as new forms of communication in real time.

- **Artificial intelligence, Big Data and Analytics:** pattern detection tools and behavior prediction models based on the use of huge databases (containing both structured and unstructured information[23] - internet connection logs, messages on social networks, etc.) and on the use of advanced modeling techniques and algorithms. All this provides better customer knowledge, which in turn improves segmentation, product customization and price, and allows a more efficient marketing.

Figure 2. Digital transformation technologies.



- **Distributed computing[24]:** technology resources (storage, processing, data sources and supercomputers) can be geographically distributed, but their interconnection can be taken advantage of by users anywhere in the world.

- **Cloud computing:** technology resources (storage, processing, data sources and supercomputers) can be geographically distributed, but their interconnection can be taken advantage of by users anywhere in the world. Also, the use of cloud-based applications enables access to information from anywhere, facilitates business operations and allows for greater speed, security and lower costs. This leads to a new business model where these resources are offered as utilities and the use of the service is billed.

- **Distributed Ledger Technology (Blockchain):** a data structure that implements a distributed registration system (ledger), that is, a cryptographic record of all operations performed and previously validated by a network of independent nodes through a consensus algorithm. This allows the recording of any digitizable asset, such as cryptocurrencies, financial instruments or "smart contracts" (programmable contracts that implement business rules and whose code is registered and can be executed in a distributed manner by the different nodes on the network). All this brings immediacy of operations, security and privacy through cryptographic rules that allow the unbreachable registration of operations, transparency (since all operations are stored in the register and can be audited by any network member), removal of a single point of failure, and cost reduction (intermediation to validate and register operations is eliminated).

- **Cryptography, cybersecurity and biometrics:** new cryptographic tools, and initiatives that seek the improvement of the information security and encryption processes, as well as the creation of more robust security systems in sensors and biometrics.

- **Internet of Things:** unlike traditional communication systems, the Internet of Things concept refers to an open-network interconnection of computer devices that send and receive data without human intervention, which allows direct, mass collection of data, as well as the remote, real-time operation of the Internet-connected devices. This technology improves the customer experience and at the same time allows the data to be used in some business processes, as is the case with pricing or the pursuit of efficiency.

[22] International Telecommunication Union (2017).
[23] Structured information is taken to mean information that is processed and prepared to be used directly through statistical and programming languages. Unstructured information is taken to mean the information that does not meet these characteristics, such as natural language information (text, audio), images, video, etc.
[24] International Telecommunication Union (2009).

- **Robotics and automated machinery:** RPA[25] tools and virtual assistants based on the interpretation of natural language allow the automation of repetitive, reduced added value tasks traditionally carried out manually, so that this available capacity can be focused on greater added-value tasks. End-user industries are rapidly adapting this new technology in order to improve product quality and reduce manufacturing costs.

- **3D printing**: aimed at the delocalized and decentralized manufacturing of items based on the remote, digitized reception of industrial design, and is widely used in industry as well as directly by consumers (in areas such as aeronautics, automotive, electronics, medicine, etc.).

- **Augmented reality and virtual reality:** used in many areas, such as in the video games or multimedia-content industries, employee training in highly specialized industries, repairs and maintenance support in the energy industry, support in off-the-plan property sales, property search engines, bank branches, ATMs, etc.

The new digital transformation environment

This new digital environment is significantly modifying the conditions in which markets reach equilibrium by changing both the supplying companies and the expectations and behavior of consumers, as well as the different conditions in which operations are performed (sale of products, services, etc.).

As will be discussed in this section, digital transformation is taking place in an environment where different drivers come into play. While new technological capabilities are the root cause of this transformation, the fundamental component making it possible are the changes in customers and markets brought about by technology, which enable, encourage and require the implementation of this technology in companies so they can adapt to this new paradigm.

These changes are magnified as a result of the disappearance of previous barriers to entry in the different industry sectors (Figure 3).

The fact is that this disruption in the supply of goods and services is motivated by various factors. A key factor is the entry of new competitors whose value proposition is fundamentally based on a technology component, which determines the business model and its related elements (high growth capacity, low costs and decreasing marginal cost, strong leverage in mobile technology, data analysis, cloud technologies and cybersecurity).

In addition to entering the market, new competitors sometimes replace the product or service offered with a totally digital or hybrid service, which results in lower costs and price and makes traditional services easily and quickly replaceable. This is what happens with new service models based on the use of algorithms that interpret data from social networks and other unconventional sources to provide relevant, real time information to their clients, thus replacing traditional information sources.

Finally, suppliers can also influence the adaptation of specific processes by generating communication platforms and digital business approaches that make it easier for their clients to migrate their product or service provisioning processes.

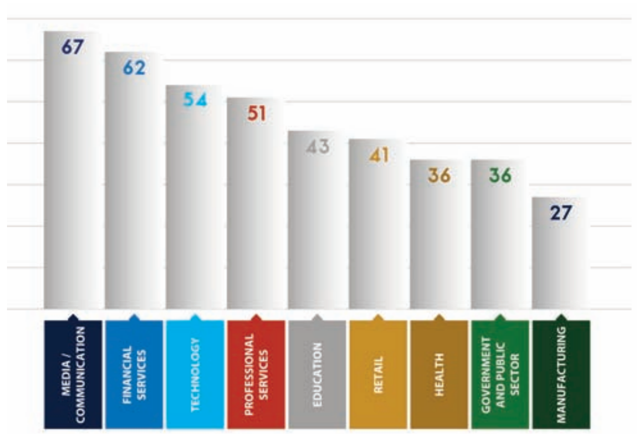
In response to all this, market incumbents are also adapting their business models despite the various inhibitors to digital transformation they have to deal with, such as resistance to change, culture or lack of training.

Indirectly, technology also acts as a driver of change in terms of how customers understand the provision of services.

Changes in consumption patterns are synthesized by the World Economic Forum (WEF) as[26]:

1. Greater focus on the customer experience: the purchasing experience becomes more important, to the point that people and organizations choose product and services not only based on their quality and price, but also on the related purchasing experience (e.g. delivery times or after-sales service).

Figure 3. Percentage of respondents who think it very likely that an industry would be affected by digital trends. Source: Harvard Business Review (2017).



Source: Harvard Business Review (2017).

[25]Robotic Process Automation.
[26]Digital Consumption (2016).

2. Hyper-personalization of the customer offer: customers have higher expectations as to how much products and services suit their preferences and lifestyle. Digital technology allows companies to meet these expectations without dramatically increasing costs.

3. Access against ownership: the concept of access as opposed to ownership is becoming widespread. Customers prefer on-demand access, thus optimizing product consumption.

The above explains the success of initiatives such as Amazon, where the customer experience during the purchase process is enriched with a personalized selection of related products, and the value proposition to the customer includes the product and some personalization in terms of delivery times and methods; or that of Netflix or Spotify platforms, where customers access content without owning it, at a lower cost.

At the same time, international regulations are adapting to the new environment, becoming more global and incisive, especially in areas relating to the protection of personal data.

Finally, the labor market itself is also being affected, with growing demand for IT and quantitative profiles, as is the Human Resources management area, with greater work flexibility and fragmentation being promoted and changes taking place in areas such as performance measurement, hiring strategies or training needs.

Implications for organizations and areas of action

Many organizations are incorporating new technologies into their processes (with a greater or lesser degree of integration), but not always with the same goal. Some companies intend to take advantage of new technologies in order to be more efficient and reduce their costs, others as a way to reach new markets, but only some use them to generate new business models.

According to a European Commission survey[27], companies tend to use new technologies, such as movile services, cloud technology, social media, etc. (Figure 4), to improve specific business functions rather than as a disruptive tool to transform business models.

Digital transformation in companies, understood as disruptive change, has implications not only in processes and systems, as they become adapted to new tools and work approaches, but also in governance, organization and in strategic definition itself.

It is therefore important for companies to take stock of their position in each of the main areas of action affected by digital transformation (Figure 5).

[27]EC (2017).

Strategy and mobilization. A strategic approach to the digital challenge implies questioning the current business and support models. There is no universal recipe for all companies, but digital transformation is not an option but a necessity whose timing and scope will depend on the context of each company.

Several formulas can be used to undertake this transformation, both in terms of a company's position vis-à-vis new technologies (early adopter, follower, challenger, etc.) and in relation to third party involvement (joint ventures, partnerships with start-ups, universities, research centers, venture capital funds, or other companies, or even through platforms based on open ecosystems).

A very important first step is to properly document the starting position of each company and its industry sector, to then try and deduce the opportunities and threats that arise with digitalization.

Another fundamental issue is the leadership for transformation that needs to be adopted by the company's CEO. The impact of transformation on people and on the company's work culture cannot be forgotten.

Since costs and risks are high, and implementation times uncertain, it is important to synthesize and prioritize the transformation goals, and to have specific indicators (such as the economic impact, improved customer experience or employee engagement) that help measure company´s progress towards compliance with such goals.

Organization, culture and governance. The way companies work and organize work is also changing. Organizational structures are becoming more horizontal, with multidisciplinary teams that are organized into projects, adopt principles and develop agile organizations. In addition, technology is changing interaction, communication and decision making.

New functions are also emerging, such as those dealing with the strategic approach to technology, the governance of data and

Figure 4. Percentage of adoption of key technologies. Source: EC (2017).

Source: EC (2017)

models, the protection of personal data, and computer security (cybersecurity).

Commercial processes. Companies are reviewing their production and distribution models, redesigning the use of their digital and traditional channels. In this context, the mobility lever stands out, as the mobile channel has become a centralizing element of a company's relationship with its customers.

Companies also seek to further personalize their value proposition (interpreting the information available through models), thus trying to improve the customer's experience.

Operational processes. The fundamental aim of (end-to-end) operational process transformation is to improve efficiency and customer service quality, as well as to strengthen operational control.

Many initiatives are used to achieve these goals, such as paperless back-office design, contact center digitalization, production process sensorization, processing center robotization, etc.

Data and modeling. Greater storage and data processing capacity leads to more efficient use of the information available; but there are also new challenges, such as the use of unstructured information, the management of big data or the analysis of information in real time.

In addition, modeling and machine learning techniques are incorporated (such as neural networks, deep networks, support vector machines, Bayesian classifiers, classification and regression trees, etc.), which contribute to improving decision making.

Data Protection. In the current environment, data has become a strategic asset. Therefore, data confidentiality and security is now a fundamental feature for companies, especially in relation to personal data.

This requires proactively managing user rights in order to comply with regulations and adequately use the potential of available data. As a result, consent and purpose management policies are needed, as are cross-border processing policies, identifying and maintaining personal data repositories, linking business actions to permitted uses, etc., which is an opportunity for organizations to rethink their data structure and governance.

Cybersecurity. In recent years, cyber risk has increased due to various factors, such as the existence of bigger and more complex IT ecosystems in companies, the integration of companies from different industry sectors or the professionalization of attacks. Damages for cyber-attacks have already reached 3 trillion dollars a year, and it is estimated that by 2021 they will have reached 6 trillion per year[29]. This is huge in a context of digital transformation, as a potential incident could impact on the continuity of business operations.

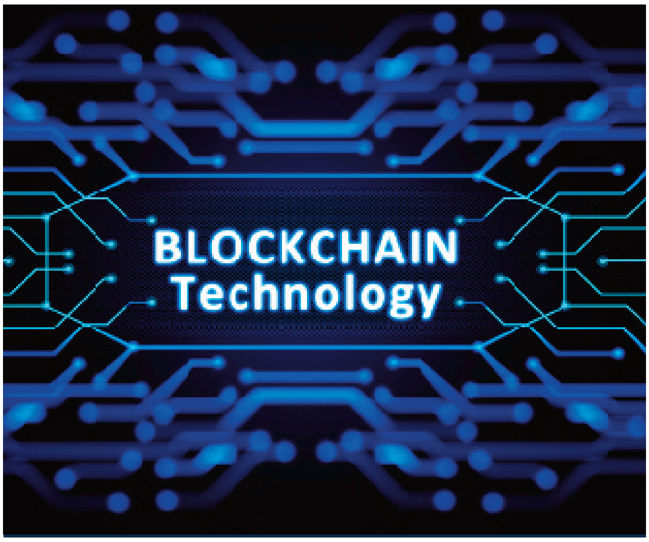
Within the digitalization strategy, cybersecurity involves taking initiatives to avoid risks during and after transformation processes. Some of the main initiatives are implementing cybersecurity frameworks, reviewing the organizational structure relating to cybersecurity, identifying and reviewing critical systems and services, measuring risk, and developing an effective response to incidents.

New technologies. Investment in new technologies is a key factor in any digital transformation process. It requires having sufficient knowledge of information and communication technologies.

The fact is that the evolution of IT platforms towards cloud environments, the definition of software architectures that incorporate open solutions and the use of disruptive technologies (big data, AI, block-chain, IoT, biometrics, robotics, etc.) have become key strategic issues for organizations.

[28]Círculo de Empresarios (2018).
[29]Cybersecurity Ventures (2017).

Figure 5. Areas of action of digital transformation[28].



Machine Learning concept and trends

All models are wrong, but some models are useful
George Box[30]

As it has already been outlined, the volume of data and the variety of data sources available are growing exponentially at a rate that exceeds the human capacity for analysis. The use of new information sources as well as techniques and algorithms capable of learning from the new information[31] leads to a number of advantages, some of which are the following:

- More types of data (both structured and unstructured) and sources of information to be incorporated into the modeling process.
- Efficient use of large data volumes in the decision-making process.
- Detection of sophisticated or non-obvious patterns, not based on a priori assumptions.
- Greater automation of modeling and self-learning leads to an increase in model predictive power. Once designed and implemented, their calibration and maintenance requirements are lower compared to traditional models, thus reducing modeling time and cost.

Machine Learning concept and components

The classical definition of machine learning is based on the first studies conducted to develop learning techniques:

"Area of study that gives computers the ability to learn without being explicitly programmed".[32]

Thus, the Machine Learning area of study deals with how to build algorithms that improve autonomously from experience[33]. From the point of view of computer science, experience is materialized in the information that is produced through data storage processes. The models and algorithms that constitute the body of this discipline are therefore based on the extraction of information from different data sources. All this leads to the formal definition of Machine Learning[34]:

"A computer program is said to learn from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with E".

Less formally, machine learning can be defined[35] as a set of methods capable of automatically detecting patterns in the data, and using them to make predictions on future data or take other types of decisions in an environment of uncertainty.

[30]Box, G. y Draper, N. (1987). Box, british statistician who worked in the areas of quality control, time series analysis, design of experiments, and Bayesian inference, is considered one of the brightest minds of 20th century statistics. Draper, professor emeritus in the statistics department of the University of Wisconsin, Madison.

[31]Shalev-Shwartz, S. y Ben-David, S. (2014). Shalev-Shwartz, professor at the Centre of Computer Science and Engineering at the University of Jerusalem, Israel. Ben-David, professor in the School of Computer Science at the University of Waterloo, Canada.

[32]Paraphrased from Samuel, A. (1959). The actual quote is: *"Programming computers to learn from experience should eventually eliminate the need for much of this detailed programming effort"*. Samuel, Master's degree in Electric Engineering from MIT, ensambled the Bell Laboratories, worked in the first comercial computer at IBM, and was a professor at Standford University.

[33]Mitchell, T. (1997). PhD from Stanford and is E. Fredkin Professor in the Machine Learning department of the School of Computer Science at Carnegie Mellon University.

[34]Mitchell, T. (1997).

[35]Murphy, K. (2012). Postdoc at MIT, professor in Computer Science at Univ. British Columbia (Canada), this textbook was awarded the 2013 DeGroot Prize. He is currently research scientist at Google on Artificial Intelligence, Machine Learning, computer vision and natural language understanding.

Machine learning techniques start with a set of observed data[36], from which classification rules or behavior patterns are obtained to be used on data different from those used for the analysis[37].

This definition encompasses the different machine learning components:

- **Information sources,** which reflect the "experience E" from which learning takes place:
  - Structured data: relational databases, file systems, etc.
  - Unstructured data: transactional, mailing, CRM, voice, image, etc.

- **Techniques and algorithms,** which are related to the "tasks T" to be executed:
  - Techniques for processing unstructured information: tf-idf, parsing, self-organized maps, etc.
  - Supervised and unsupervised models: classification models, stochastic models, simulation, optimization, boosting, etc.

- **Self-learning ability,** which improves "performance measurement P".
  - Automatic retraining based on new information.
  - Combination of models and reweighting / calibration.

- Use of **systems and software** for the visualization of information and programming[38]:
  - Visualization: QlikView, Tableau, SAS Visual Analytics, Pentaho, TIBCO Spotfire, Power BI.
  - Programming: R, Python, Scala, Ruby, SAS, Java, SQL, Matlab, C, Google, AWS, Azure.

The development of these components represents a step forward with respect to the traditional modeling approach.

There are some differences between both approaches[39] (Figure 6).

Machine Learning techniques

Machine Learning based on information and the learning paradigm

Different Machine Learning techniques are used depending on the type of information (structured or unstructured information), and learning paradigm (Figure 7). The choice of technique will depend on the aim of the model to be built and on the type of information available, as well as on other factors.

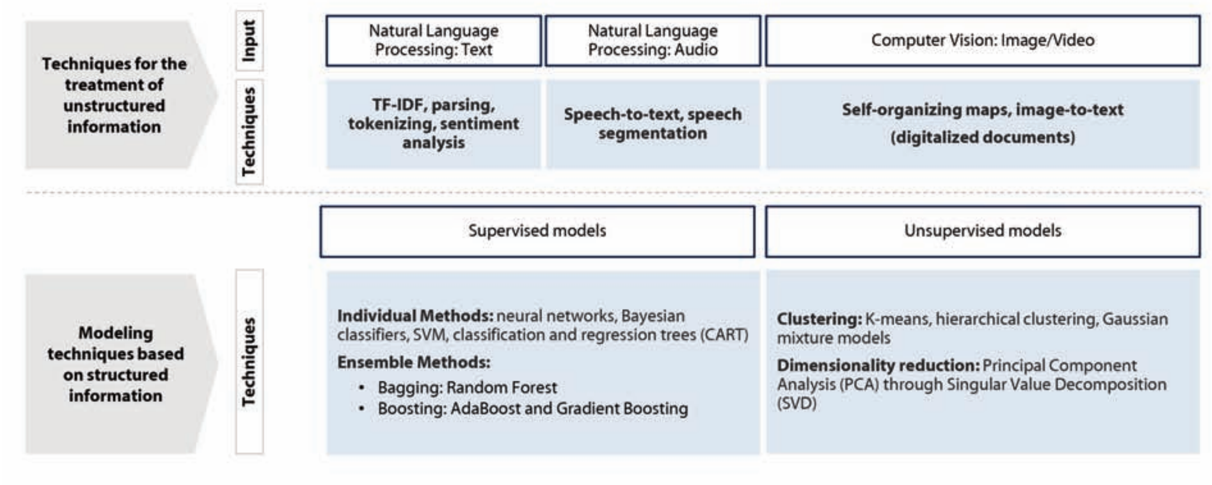
There are some specific techniques that can be used for transforming unstructured information (texts, sounds, images, etc.) into data that can be analyzed and processed by a computer. These techniques are used in areas such as natural language processing or image identification (Figure 8). Some examples are the use of statistics for assessing the relevance of words in written texts, the classification of words into categories for the understanding of written text, the use of

[36]This data set is often referred to as a "training sample".
[37]However, the model construction and validation process itself uses data which, although observed, are not included in the training sample, and simulate the data to be used. These data usually make up what are called the "test sample" and the "validation sample".
[38]The massive use of data also entails the implementation of big data tools and the use of computational efficiency techniques, although these areas have not been included in the Machine Learning concept in this paper.
[39]Domingos, P. Professor at University of Washington. He is a researcher in machine learning and known for markov logic networkenabling uncertain inference. (2012).

Figure 6. Differences between Machine Learning and the traditional approach.

	Traditional modelling	Machine Learning
1 Sources of information	Structured data Reduced number of data sources Limited by the initial design, structure, and internal processes of obtaining information Non-continuous and planned updates of the data	Structured and unstructured data, and more granular Multiple data sources Extensive, easily accessible and continuously growing: social networks, public databases or private data pools, Int. Of Things, etc Continuous and real-time update of the data
2 Techniques and algorithms	Statistical and mathematical basis Limitation in the patterns and relationships identified due to the assumption of previous hypotheses Use of deductive methods	Statistical and mathematical basis + computer science and artificial intelligence Identification of patterns hidden in the data without assuming previous hypotheses Use of inductive methods
3 Learning	Manual and planned update of the model Predefined hypotheses or prior knowledge of the relationships between variables Reduction of predictive power over time by anchoring to a time window Traceability available	Automatic update and self-learning Search for patterns and relationships without restrictions Maintenance of predictive power over time by adapting the time window Traceability: not assured
4 Systems and software	Lower computational requirements Traditional tools, sometimes used stand-alone Unique use of repositories of structured information Unstructured data seen as worthless files	Higher computational requirements and processing speed and data management capacity Combination of tools Storage and use of Data Lakes that combines structured and unstructured data

Figure 7. Machine Learning types and techniques.



neural networks for voice or image recognition, the use of Markov chains for the construction of texts in natural language, or the use of unsupervised classification algorithms for organizing images.

Modeling techniques that are used with structured information can be classified according to the information used for learning[40,41]:

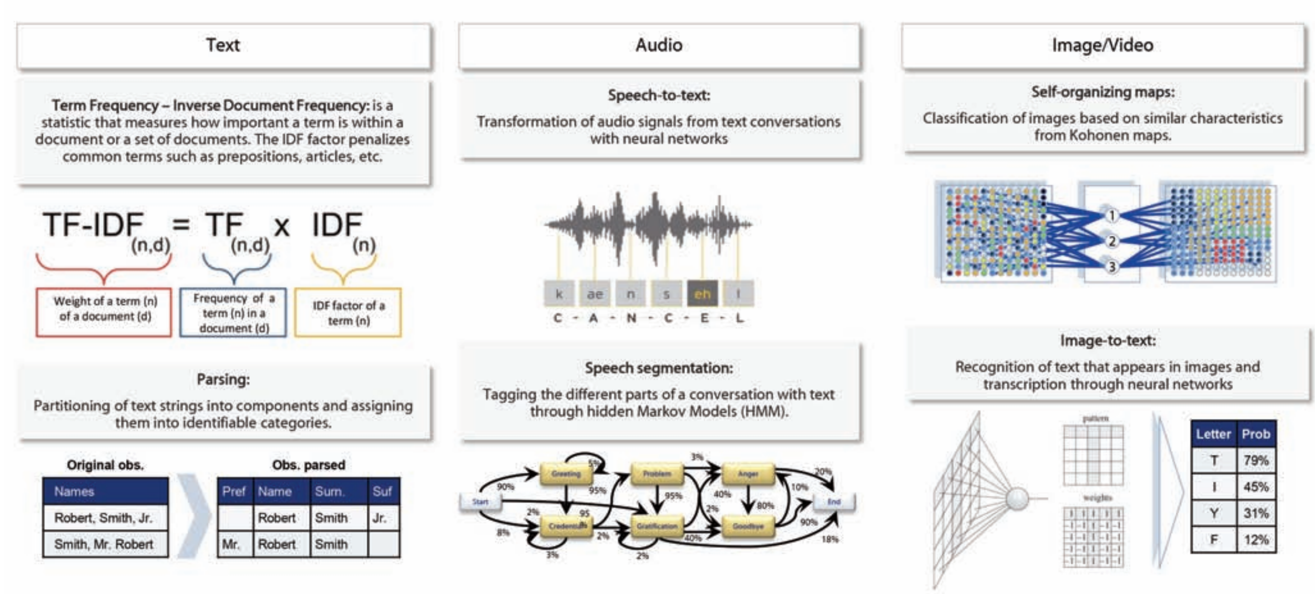
Supervised learning: the data used to build the algorithm contains information about the characteristic under study, which is not present in the future data. The information that is to be predicted or used for classifying a population is present in the data chosen to build the model. More formally, the aim of supervised learning is to train a set of variables (called explanatory, characteristics or factors) "x" in an output variable "y", from a data set (called training sample) of pairs $\Delta=\{(x_i, y_i), i\in 1, \ldots, N\}$, where N is the sample size.

If the output variable "y" is continuous, it is referred to as a regression problem, whereas if it is nominal or discrete, a classification problem[42].

Unsupervised learning: as opposed to supervised learning, the data training sample does not contain a variable to be predicted. Since in this case the output variable is not available, the data set has the form $\Delta=\{x_i, i \in 1, \ldots, N\}$, where N is the sample size[43].

[40] Shalev-Shwartz, S. and Ben-David, S. (2014).
[41] Murphy, K. (2012).
[42] E.g. in the case of classification models based on credit quality, the target variable is dichotomous (default / no default), or discrete (rating level), in which case it is referred to as a classification problem.
[43] Different groups in the data are intended to be found without having a sample where these groups have previously been observed.

Figure 8. Natural Language Processing and Computer Vision methods.



This type of problem aims to find patterns or relationships in the data, which is why it is also known as knowledge discovery (identifying patterns in data that are valid, novel, potentially useful and understandable[44]).

Machine Learning techniques

The modeling process usually includes an initial knowledge discovery phase. Some of the tasks carried out as part of this process are as follows (Figure 9):

- **Data understanding:** initial data preparation and descriptive analysis, data quality analysis, etc.

- **Data preparation:** data cleansing or processing (including the processing of missing data, outliers and erroneous or inconsistent records), multivariate analysis, combination or creation of new variables from existing ones, reduction of the number of variables (through the removal of redundant variables, projection onto lower-dimensional subspace), etc.

- **Selection of the appropriate technique and application of regularization processes,** where information is transformed and prepared for modeling. These processes include the following:

  - Unifying the range of variables, for example by subtracting the mean and dividing by the standard deviation (normalization), or dividing by the range of the variable (scaling) etc.

  - Identifying those variables that are most relevant to the model to be built. As a step forward from methods previously used for selecting variables, such as stepwise[45], techniques such as the use of elastic nets can now be used: the function used to estimate model parameters (called target function or cost function, whose value is to be minimized) is modified by adding an additional element to detect which variables do not provide information when comparing the model in training and test samples, which therefore allows the automatic selection of variables: if L is the cost function used to obtain the models' estimates, $\beta^t=(\beta_1,\ldots,\beta_n)$ the estimates, and $\lambda_1 \in \mathbb{R}$, $\lambda_2 \in \mathbb{R}^+$, then the function can be transformed using the following expression[46]:

  $$L' = L + \lambda_2 \|\boldsymbol{\beta}\|_2^2 + \lambda_1 \|\boldsymbol{\beta}\|_1$$

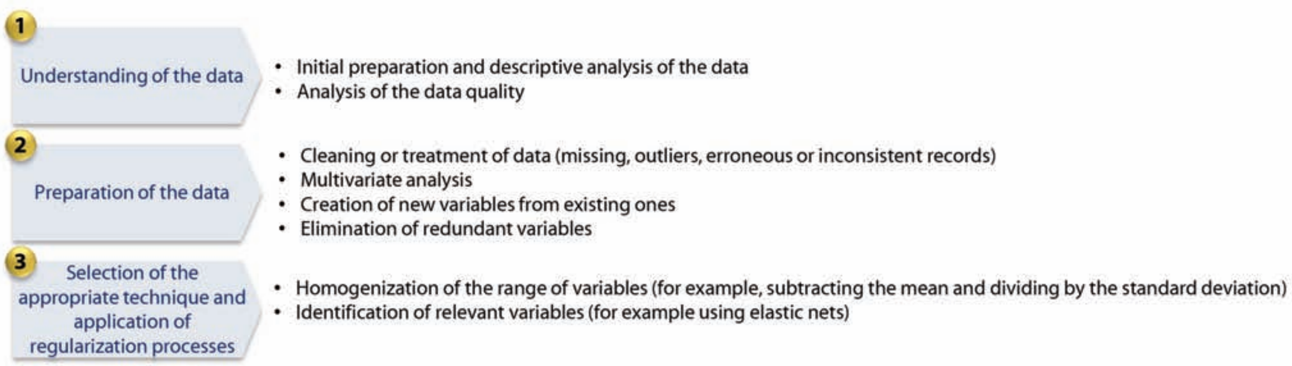
  In the specific case where $\lambda_2=0$, the LASSO method is obtained, and where $\lambda_1=0$ the ridge regression method is obtained[47].

Supervised Machine Learning techniques include the so-called individual methods (Figure 8), referred to in this manner because they can be used in isolation. Some of the best known are neural networks, support vector machines, Bayesian classifiers, and classification and regression trees:

- **Neural networks** are non-linear, multivariate mathematical models that use iterative procedures with the aim of minimizing a specific error function and thus classifying the observations. Neural networks are composed of neurons connected to each other through nodes and layers. These connections emulate the dendrites and axons in biological nervous systems, through which information is transferred. They are used in both supervised and unsupervised classification

[44]Cios, K.J. and other (2007). With a PhD in Computer Science from the University of Science and Technology in Krakow, he worked for the International Atomic Energy Agency, and is currently Professor and Head of the Computer Science department at Virginia Commonwealth University. His research is focused on machine learning, data mining, and biomedical informatics.
[45]An iterative model construction method based on the automatic selection of variables.
[46]Where $\mathbb{R}$ represents a set of real numbers, $\mathbb{R}^+$ a set of positive real numbers $\|\boldsymbol{\beta}\|_2^2 = \sum_j \beta_j^2$ y $\|\boldsymbol{\beta}\|_1 = \sum_j |\beta_j|$
[47]For a detailed discussion on each method, advantages and disadvantages, see Murphy, K. (2012).

Figure 9. Tasks in the modelization process.



problems, with the advantage that they can separate regions non-linearly. Their biggest disadvantage is their "black box" nature, i.e. the difficulty in interpretation of the results and the limitations encountered when incorporating business sense into their complex weights structure.

An important extension of neural networks is the so called deep networks, consisting of the use of multiple-layer neural networks. These deep learning[48] models can have millions of parameters, depending on the complexity of the problem to be tackled. However, given the estimation difficulties, there are multiple approaches to this type of method[49] (e.g. the use of optimization algorithms to adjust deep network parameters based on output error; the use of greedy algorithms to train specific networks such as directed networks; the use of auto-encoders to reduce dimensionality, etc.). New technologies have made it possible to incorporate these methods into, for example, automatic text recognition and generation processes, or in computer vision.

- **Support vector machines (SVM)** are classification models that aim to solve the difficulties that can be posed by complex data samples, where relationships are not necessarily linear. The aim is therefore to classify the observations into several groups or classes, but these cannot be separated using a hyperplane in the dimensional space defined by the data. For this, the data set is embedded in higher dimension space through a function[50] that allows the data to be separated in the new space

[48]The algorithms used in deep learning can be of the supervised or unsupervised learning type.
[49]Murphy, K. (2012).
[50]This function is called kernel.

Differences between artificial intelligence, Machine Learning and deep learning

Artificial Intelligence (AI), Machine Learning and Deep Learning are related concepts and the difference between them is often obscure. The following definitions are provided in order to shed some light on these differences:

- Artificial intelligence is the broadest of these three concepts. It aims to make machines capable of performing tasks in the same way that a human being would. In most cases, this is achieved through programming the machines in advance. It was defined in the 1956 Dartmouth Artificial Intelligence Conference as "every aspect of learning or any other feature of intelligence that can in principle be so precisely described that a machine can be made to simulate it". There were a few examples of this already in the early twentieth century, with breakthroughs such as Alan Turing's cracking of the Enigma machine using what nowadays is called neural networks.

- Machine Learning (ML) can be considered a branch of AI, and is defined as[51] "a set of methods that can automatically detect patterns in a data set and use them to predict future data or to implement other types of decisions in environments of uncertainty".

- Deep Learning is a subset of Machine Learning techniques that, defined at its most basic level, can be explained as a probability system that allows computational models made up of multiple processing layers to learn about data using multiple levels of abstraction[52].

[51]Murphy, K. (2012).
[52]Bengio, Y. and others (2015). Is a computer scientist in Canada. He is mainly known for his work in neural networks and Deep Learning.

through a hyperplane in this new space. A hyperplane is then searched for that is equidistant to the closest points for each class (i.e. the aim is to find a hyperplane that separates the classes while being furthest away from the related observations).

- **Bayesian classifiers** are models based on Bayes' conditional probability theorem, which uses the known information from explanatory variables, the so-called priors, to classify the observations. In other words, a Bayesian classifier assumes that the presence or absence of certain characteristics allows a certain probability to be assigned to the absence or presence of another characteristic, or a target variable to be defined based on the relationship existing in a sample between these characteristics and the target variable. It is a simple but robust technique to classify observations into a set of classes. The Bayesian classifier takes different forms depending on the distribution assumed to be followed by the explanatory variables (normal, multinomial, etc.).

- Finally, **classification trees** (when the target variable is categorical) **and regression trees** (when the target variable is continuous) are analysis techniques used to predict the assignment of samples to predefined groups based on a series of predictive variables. Decision trees are simple and easily interpretable models, which makes them highly valued by analysts. However, their predictive power may be more limited than that of other models, because they perform orthogonal[53] space partitioning, which turns the sample into silos and limits predictive capacity due to the fact that this type of algorithm is prone to overfitting.

All these individual methods can be combined with techniques and algorithms that improve predictive capacity into ensemble methods (Figure 10). These methods consist of the aggregation of individual models in order to generate a more predictive or more stable model that takes advantage of collective knowledge. For example, models can be combined by using results from independent models (as in the case of the bagging technique) or models can try to correct errors in each new iteration (as in the case of boosting).

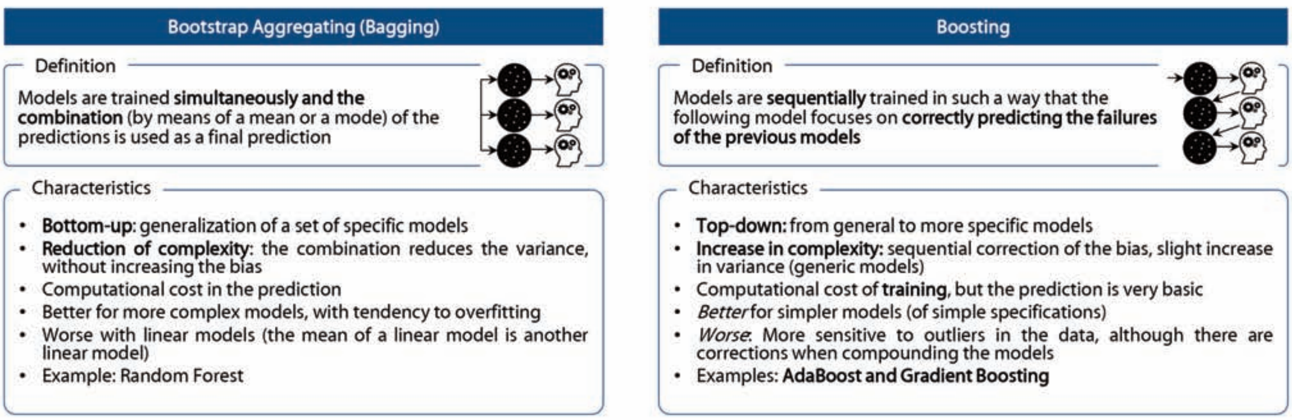
Finally, unsupervised models include clustering techniques or Data Analysis techniques such as those for dimensionality reduction:

- Clustering is an unsupervised model used to identify groups (clusters) or similar observation patterns in a data set. One of the most widely used techniques is the k-means method, which consists of defining a central point of reference for each cluster (called a centroid), and assigning each individual to the nearest centroid cluster based on the distances between the input attributes. The algorithm is based on the random placement of k centroids and, through an iterative process, assigning each point to the cluster with the closest centroid to then update the value of the centroids. This process ends when a certain convergence criterion is reached.

- Dimensionality reduction methods aim to reduce the number of dimensions in the analysis space, determined by the set of explanatory variables. One of the techniques is Principal Component Analysis (PCA), which converts one set of correlated variables into another uncorrelated set with fewer variables called principal components. The main disadvantage of using PCA on the data set is that they lose their interpretability.

[53] I.e. a partition of a n-dimensional space into regions through hyperplanes perpendicular to each of the axes that define the explanatory variables

Figure 10. Examples of ensemble methods.



The use of these techniques means that methods for results validation need to be more sophisticated. Since in many cases the learning is often continuous and more than one subsample is used in construction, suitable techniques need to be used for these new processes, such as bootstrapping[54] or k-fold cross-validation[55], which allow the model to be evaluated in more than one validation sample. Attention should also be paid to other aspects of the validation, such as potential bias in relation to the frequency with which the data are updated, and spurious or non-causal relationships. Finally, the dynamic nature of these techniques makes traceability difficult. This is particularly relevant in regulated or supervised environments where a specific validation and traceability framework is required.

Machine Learning uses and trends

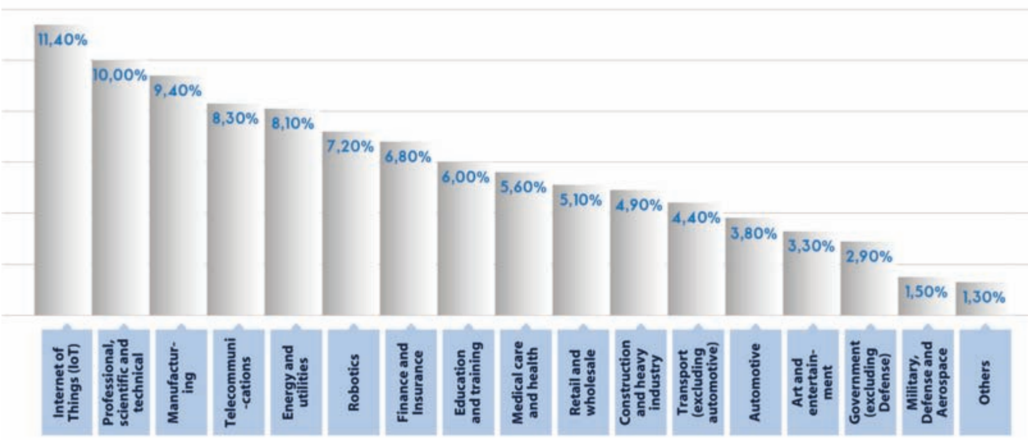
Use of Machine Learning in industry

The use of Machine learning in business is taking place at different rates (Figure 11) depending on both the industry and various factors related to size, management style and the general environment in which individual companies operate.

Machine learning techniques are being implemented with different degrees of speed and depth in different industry sectors. Some relevant examples can be seen in the education, finance, and health industries, or, from a cross-

[54]Development of many validation samples from the random selection with replacement.
[55]The sample is divided into k groups. In the first iteration, k-1 are used for training, and the remainder for validation. This process is repeated, choosing each of the k groups as a validation group.

Figure 11. Machine Learning presence in the different industry sectors in 2016.



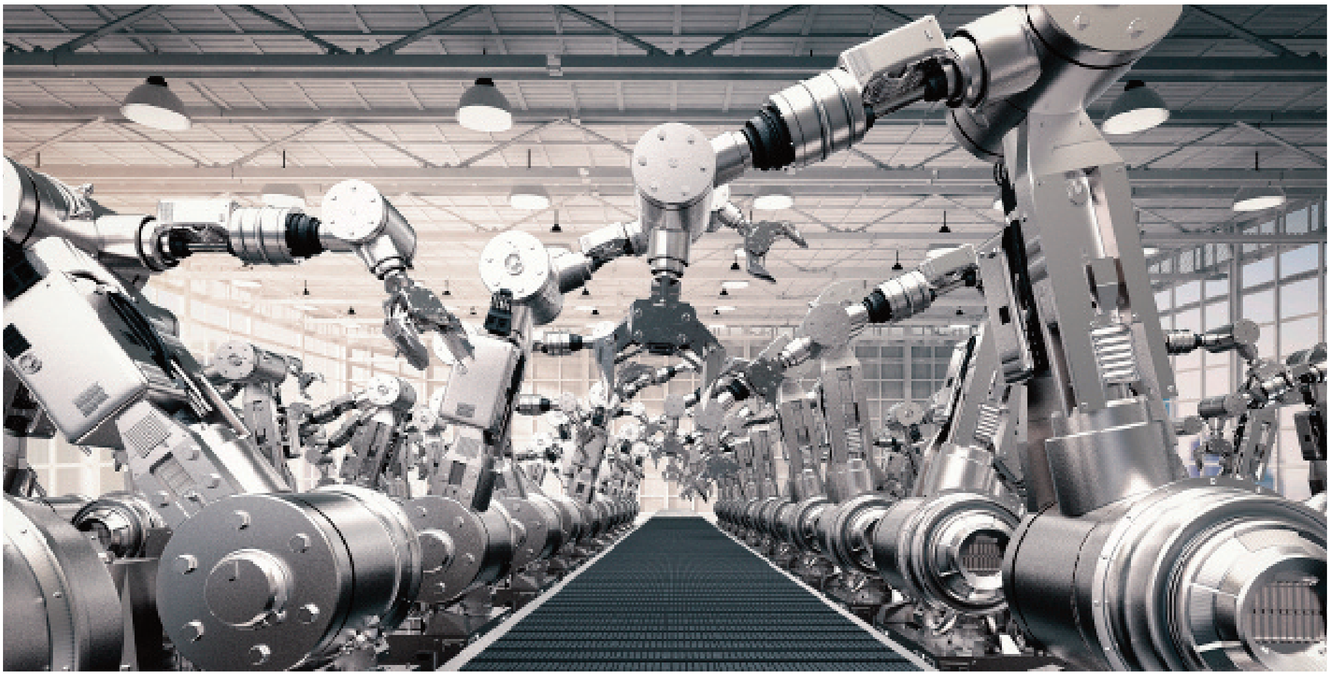
Source: Statista

industry perspective, in the improvement of efficiency within organizations[56,57]:

- In the educational field[58], artificial intelligence will be able to provide systems that function as "learning companions" throughout the student's life and are accessible through multiple channels.

- In finance, Machine Learning algorithms target functions such as automatic trading, the creation of roboadvisors for automatic portfolio management, fraud detection or risk measurement. One of the fastest growing areas is known as RegTech[59], where Machine Learning techniques are used to comply with regulation and supervision. The RegTech market is expected to reach 6.5 billion dollars by 2020[60]. These techniques can also be used for the analysis of portfolios where not as much structured information is stored in the databases of financial institutions, as is the case with non-client prospect models. Here, prospect models are intended to classify potential customers according to their probability of default, which is useful in marketing processes such as campaign launches or new product releases. These models are also useful in customer segments for which not much information is available, as is the case with self-employed workers and micro-enterprises, or in individuals or self-employed segments that do not use banking services. For this, information on annual accounts can be used (including business activity-related information such as inventory or supplier churn and leverage, liquidity, solvency or profitability ratios), supplemented by information on products and services, or with unstructured external information.

[56]Kaggle (2017).
[57]Forbes (2017).
[58]Pearson (2016).
[59]Regtech results from combining the terms "regulatory technology" and consists of a group of companies that use technology to help companies comply with regulations efficiently and cost effectively.
[60]Frost & Sullivan (2017).



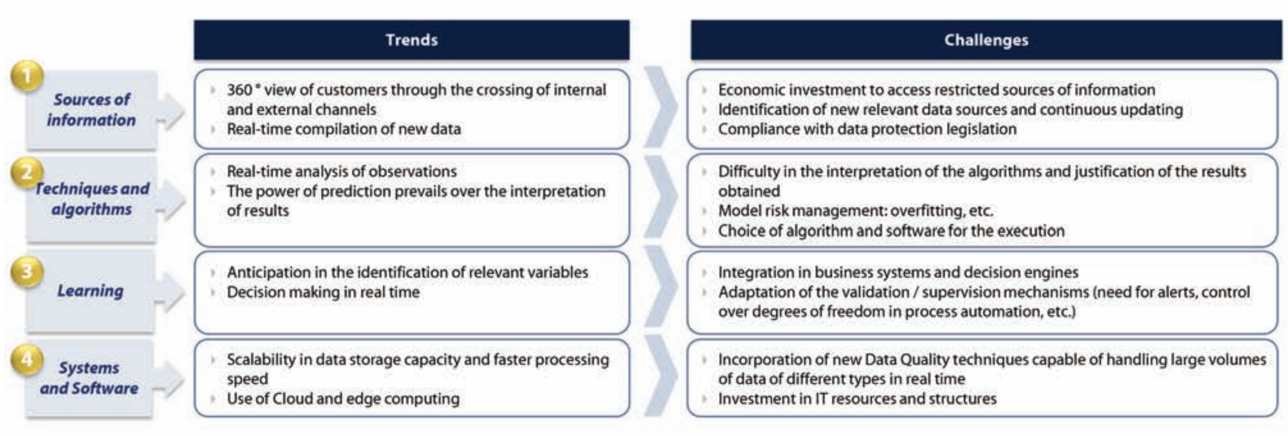
- In the health sector, efforts are aimed at improving diagnostic imaging, the management of consultations for treatment and recommendations, as well as the collection of medical information for study or robotic surgery. It is estimated that the market for artificial intelligence in the field of health can reach 6.6 billion dollars by 2021 and potentially save 150 billion dollars by 2026 in the US health sector[61].

- In areas such as manufacturing or logistics, solutions are proposed to improve the maintenance of machinery (through the use of predictive maintenance models and sensorization) or distribution efficiency (for example, by optimizing the correlation between the transportation needs of multiple enterprises)[62].

- As for efficiency improvements in organizations, the IT and operations function can be more proactive in companies where IT systems generate big data (logs, status reports, error files, etc.). From this information, Machine Learning algorithms can detect the root cause of some problems, and system failure can be mitigated significantly using predictive analytics. A personalized IT experience can also be incorporated: the availability of a virtual assistant for each employee will help companies maximize productivity in the workplace. This assistant will be able to connect information on the employee's use of applications with other stored databases, which will enable it to detect patterns in order to anticipate the next actions and access to document databases required by the employee.

[61] Techemergence (2017).
[62] Círculo de empresarios (2018).

Figure 12. Machine Learning component trends and challenges.



An example of how these techniques are being used in a traditional industry is Rolls-Royce, which has reached an agreement with Google to use Google's machine learning engines in the Cloud in order to improve short-term navigation efficiency and safety, ultimately aimed at achieving remote control without a crew. This case includes all the typical elements of machine learning: collection of unstructured information based on sensors, information processing using new techniques, full automation, etc.

Machine Learning component trends

The use of different techniques and the proliferation of different approaches and tools change the modeling process all around: the data used, new calculation techniques and methodologies, and support systems and architectures (Figure 12). However, this new approach entails a set of challenges from a technical, methodological and business point of view[63]:

1. **Depending on the sources of information:** the aim is to obtain a 360 ° customer view by cross-referencing internal and external channels, using data collected in real time. This implies challenges such as the economic investment required to access restricted sources of information, the identification of new sources of relevant data and their continuous updating, as well as the incorporation of Data Quality techniques capable of dealing with large data volumes of different types in real time. Another major challenge is implementing data protection regulations such as the GDPR in Europe, which implies in-depth analysis of information sources that may be used in machine learning models, their treatment and storage, and the establishment of cybersecurity systems and controls.

2. **Depending on the techniques and algorithms:** Although it makes it more difficult to interpret the algorithms and justify the results obtained, prediction power prevails over interpretation in some areas. This makes it all the more relevant to properly manage model risk throughout the life cycle (development, validation, implementation, use, maintenance and withdrawal). The nature and potential sources of model risk, such as shortcomings in the data, estimation uncertainty or model error, as well as inadequate use of the model[64], are equally present when using machine learning techniques. Other challenges in solving problems through the use of Machine Learning techniques include the proper choice of algorithm from among a wide range of possibilities (since this can depend on both the purpose of the model to be built and the information available), the existence of potential overfitting[65] and data dredging[66] issues, and the need to establish mechanisms so that the use of automatic results does not result in discriminatory processing (this risk is dealt with through what is known as algorithmic ethics).

[63] Jordan, M. I., Mitchell, T. M. (2015) and Kaggle (2017). Jordan, after graduating from the University of California with a PhD in Cognitive Science, he was Professor at MIT, and is currently Professor in the Department of Electrical Engineering and Computer Science and the Department of Statistics at the University of California.
[64] Management Solutions (2014).
[65] *Overfitting*: characteristic that occurs when the model has been fit to the training sample too closely, so that it does not achieve satisfactory results on samples other than this one (e.g. on the validation sample).
[66] Data dredging occurs when relationships are found that are not supported by hypotheses or causes that really explain those relationships.

3. **Depending on self-learning:** Incorporating the self-learning feature in the models implies the algorithm's ability to modify autonomously based on the changes that occur in the target population. This can lead to modifications in the parameters or variables that are relevant for building the algorithm, as well as the need to make modeling decisions in real time. However, adjusting the models in real time makes it more difficult to validate them, as the validation has to move away from the traditional approach and incorporate new elements such as the control of the degrees of freedom in automation, more frequent evaluation of the discriminatory power, or other contrast techniques (e.g. through simpler alternative models that will make it possible to challenge the results even though they might lose their ability to fit the data). In the specific case of models that require supervisory approval, such as regulatory models in the financial industry, the difficulty of obtaining supervisory approval poses an additional challenge to validation.

4. **Depending to the systems and the software used:** an investment in IT architecture and infrastructure that ensure scalability in data storage capacity and a greater processing speed is required. Also, Cloud-based solutions are developed as an alternative to the in-house implementations traditionally used by companies, and edge computing infrastructures are incorporated, offering pre-installed functions (FaaS[67]) ready for direct use, which simplifies the developer's experience since it reduces both the programming effort and code management. All this requires investment in IT structures and resources.

A further implication is the need for human resources with a high degree of preparation and specialization, capable of understanding and working with programming languages, algorithms, mathematics, statistics, and complex IT architectures.

[67] *Functions as a Service,* a server-less model.

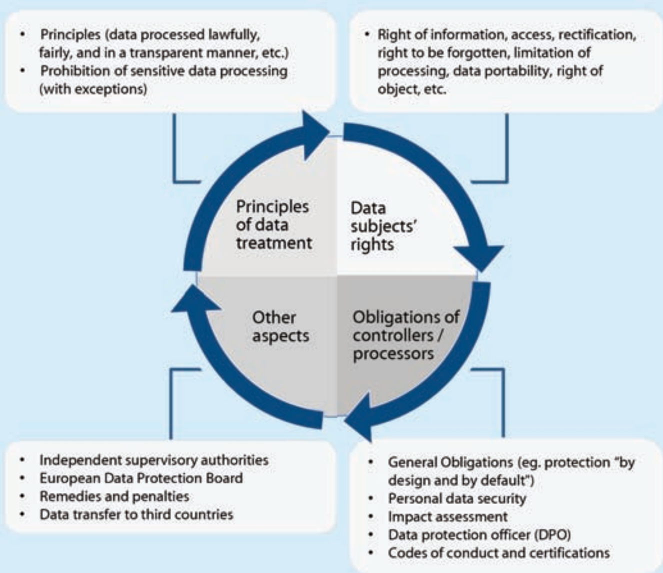
Information sources and regulatory constraints

Globalization and the rapid advance of technology have led to new challenges in the area of personal data protection. The extent of personal data collection and exchange has increased significantly, and today technology allows the processing of personal data on an unprecedented scale.

These advances and the use of information sources that incorporate personal data, require a solid and coherent data protection framework given how important it is to create the confidence required for the digital economy to develop. This gave rise to the need for people to have control over their own personal data, and to reinforce security for individuals, economic operators and public authorities[68] from both a legal and a practical perspective. As a result of this, regulators have developed standards to ensure data are protected. A paradigmatic example is the case of the European Union, where the General Data Protection Regulation (GDPR), which entered into force in May 2018, has established data protection and processing requirements of a general nature that unify data processing across all member countries (Figure 13).

[68] European Parliament (2016).

Figure 13. GDPR main requirements.



Quantitative exercise

Your job in a world of intelligent machines is to keep making sure they do what you want, both at the input (setting the goals) and at the output (checking that you got what you asked for)

Pedro Domingo[69]

Machine Learning techniques introduce several novelties with respect to the classical statistical and econometric techniques, leveraged in the use of greater volume of information and greater algorithm complexity.

This introduces a new set of elements in modeling, which allows for more robust analysis but requires solving issues such as the increase of available variables (which increases the complexity involved in both data processing and the selection of relevant information), the correct identification of spurious relationships, and non-compliance with the assumptions of traditional models (e.g. lack of cointegration in time series, or presence of multicollinearity and autocorrelation).

As a result of the above, the modeling approach can be modified by introducing different tools to take advantage of the Machine Learning algorithms while avoiding the mistakes associated with these new tools.

This section presents a behavioral scoring model that was developed to classify bank loans. In doing so, some of the already discussed new techniques were used to be able to compare results, as well as to assess the challenges posed by their use in the modeling process, the opportunities that arise and the existing risks.

Purpose

The purpose of the exercise is to analyze and illustrate how the use of Machine Learning techniques impacts on model development. Specifically, the aim is to determine how the estimation process and results obtained vary depending on the modeling approach using Machine Learning techniques.

For this, a scoring model has been trained using different Machine Learning approaches, and these approaches have been compared against traditional techniques. Two modeling phases have been developed based on a loan sample (data processing tasks involving the preparation, cleaning and selection of variables, and use of different model estimation approaches). Then, the sample is described and the different phases of the analysis are detailed, as were the findings and main conclusions.

Data used

The analysis was conducted using a set of over 500,000 loans spanning over 10 years of history, with a default rate of 6%. The sample included loan-related variables (amount, term, interest rate, financing purpose, etc.), customer-related variables (annual and monthly income, equity, indebtedness, external credit scores, information about defaults such as bankruptcies in public records, accounts with payment delays, etc., or variables measuring seniority as a customer and seniority in other products), and other variables that may be useful in modeling, such as commissions, number of transactions carried out, etc. During the exercise, all variables were maintained in the initial sample for knowledge discovery purposes.

Study development

The analysishas been developed in two phases, which are detailed below:

Phase 1: Knowledge Discovery. An initial knowledge discovery phase deals with missing values, outliers, and the grouping of variables.

Phase 2: Implementation of machine learning techniques. During a subsequent phase, different models have been developed using modeling techniques on the training samples generated.

[69]Domingos, P. (2015). Professor at University of Washington. He is a researcher in machine learning and known for markov logic networkenabling uncertain inference.

Phase 1: Knowledge Discovery

The first phase of the exercise involved the use of knowledge discovery techniques. Some of the processes, described below, may be more appropriate than others, depending on the final algorithm type used in the modeling. This initial phase allows to obtain a set of training samples that can be used as data to train different algorithms.

During this phase it was observed that 40% of variables had over 10% missing observations. Different treatments were carried out on these variables in order to use the available information[70]:

a. The mean value (or the mode in the case of qualitative variables) was assigned to the missing observations.

b. A clustering analysis was run, and the mean (or the mode in the case of qualitative variables) for the cluster was assigned to each missing value. The optimal number of clusters resulting from the best classification was determined using quantitative methods[71].

c. A regression was carried out on other variables in order to estimate a value for the missing observations in each variable based on the remaining information.

d. For the specific case of the Random Forest, a maximum value (e.g. 9999) was assigned in order to identify the missing observation as an additional value for the tree to automatically create a specific branch for this information.

An outlier treatment was also carried out, which for simplicity reasons consisted of using a floor in the 1st percentile and a cap in the 99th percentile for variables containing outliers.

Finally, a process was followed to simplify the variables. This required using again a clustering algorithm to replace those variables that make it possible to identify a good classification into groups with the assigned cluster (Figure 14). Also after a first exploratory analysis, those variables that were found not to contain relevant information were removed.

After those treatments, different training samples were obtained that would be used to estimate the algorithms[72].

Together, these techniques made it possible to reduce the number of variables in order to make the subsequent processes more efficient, prepare the existing information so that it would adapt to the specific requirements of the different models and potential limitations of the algorithms, as well as to anticipate and validate assumptions in order to evaluate the results obtained from the different modeling techniques.

Phase 2: Use of machine learning techniques

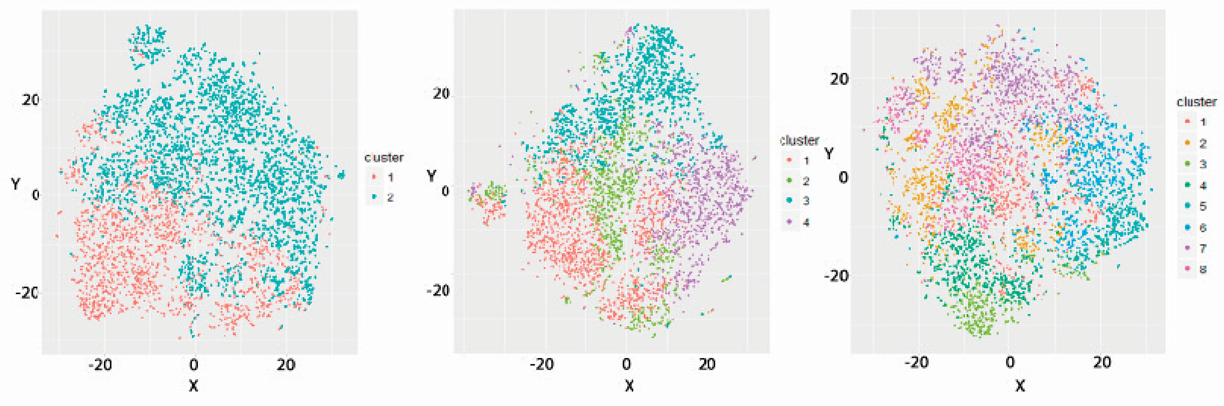
During the second phase, different techniques were used on the different samples already treated, and different measures of the discriminatory power of the model were calculated on a validation sample. The following techniques were used:

a. First, a logistic model was built in order to be able to compare against a traditional modeling method.

b. An elastic net was used as a regularization technique to observe the impact of this technique against traditional modeling.

[70]Treatments other than the classical ones consisting of replacing a missing value with a zero, and creating a dummy variable to recognize the missing values.
[71]Several quantitative methods can be used to perform this selection, such as the elbow method, or the partition around medoids (PAM) algorithm. Overall, the aim is to analyze the number of clusters from which the marginal gain from information in the classification by expanding the number of clusters becomes residual. The PAM algorithm was used in this case.
[72]Although different training samples can be used for a single algorithm, therefore resulting in different outcomes, this study shows only optimal results.

Figure 14. Clustering examples for K=2, K=4 and K=8 based on two variables. Variables X and Y would be replaced with the variable that defines the cluster.



c. A random forest was developed as one of the ensemble methods.

d. An adaboost was estimated so that a second ensemble method would be available.

e. A model was estimated through an SVM with a linear function.

f. Finally, a second SVM model with a radial function was estimated.

The following models were obtained after using the above techniques:

a. Logistic model (base model for comparison). A preselection of variables was carried out using bivariate and multivariate techniques, taking into account the individual significance of each variable, the discriminatory power of the global model, and different information criteria methods. This resulted in a model that has 11 variables the most important of which are behavioral (e.g. number of accounts that have never been in default, whether there are pending debts, and the number of accounts that have ever been in an irregular situation for more than 120 days). The importance of each variable was measured as the marginal contribution of that variable to the goodness of model fit. This made it possible to analyze the pattern of incorporation of each variable according to its relevance (Figure 15).

b. Model with regularization techniques. After using the elastic net, a model with a total of 45 variables was developed, with the emphasis being on behavioral variables (e.g. accounts in irregular situation, outstanding debts in the last 12 months, positive information in public records, etc.). Although loan and customer related variables were included, the behavioral variables were much more important (Figure 16).

c. Random Forest. A 50-tree algorithm was obtained where the maximum branch depth was 3 and the number of variables randomly included as candidates in each node was 7. A total of 80 variables were used to form the set of trees. The variables with greater relevance in the algorithm were behavioral (e.g. any outstanding debts, other accounts with unfulfilled due dates, number of active accounts, etc.), and some other variables that characterized the customer or the loan (e.g. level of indebtedness). It was observed that, in terms of relevance, the former variables had a higher weight, while the relevance of the latter was relative low (Figure 17).

Figure 15. Relevance of variables in a traditional algorithm.

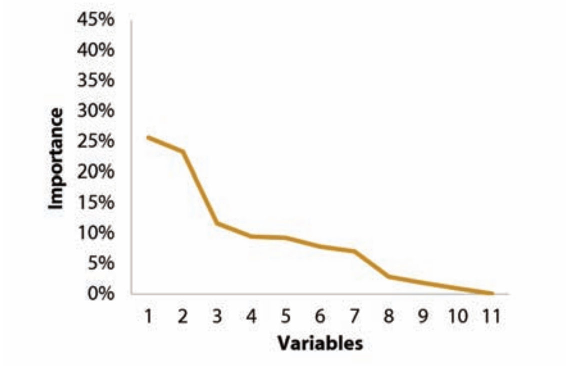


Figure 16. Relevance of variables after using the elastic net.

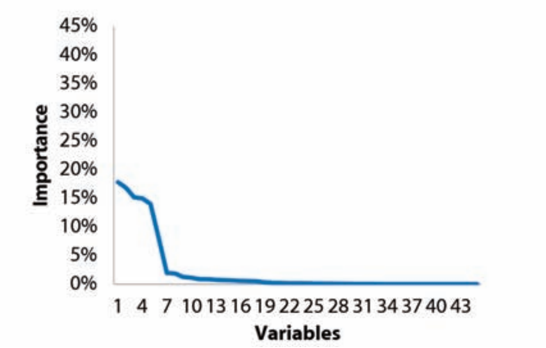
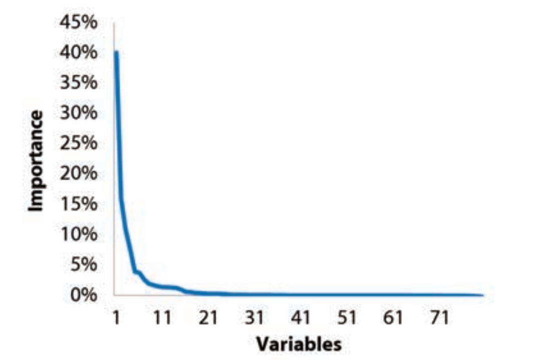


Figure 17. Relevance of variables in the Random Forest algorithm.



d. Adaboost. A 20-iteration algorithm was selected and its reduction parameter was set by default to 1. As a result, 68 significant variables were obtained including behavioral variables that were again particularly relevant (e.g. average revolving balance for all accounts, number of credit card accounts without past due payments, principal repaid by the client to date, etc.), as well as other variables relating to the customer. In this case, the relevance of variables was observed to have a slight and gradual decrease (Figure 18).

e. SVM with a linear function. Different models were obtained based on the variables used, and the best model was selected using cross-validation techniques. In the chosen model, 68 relevant variables were obtained, of which behavioral ones were the most relevant (e.g. outstanding debts in the last 12 months or percentage of accounts never in an irregular situation, Figure 19).

f. SVM with a radial function. As with the linear function case, a selection was made from among different models. This model also had 68 variables of which the most relevant ones were also behavioral. Although both SVM models are very similar, the importance of the variables in each of them differs markedly (Figure 20).

Figure 18. Relevance of variables in the Adaboost algorithm.

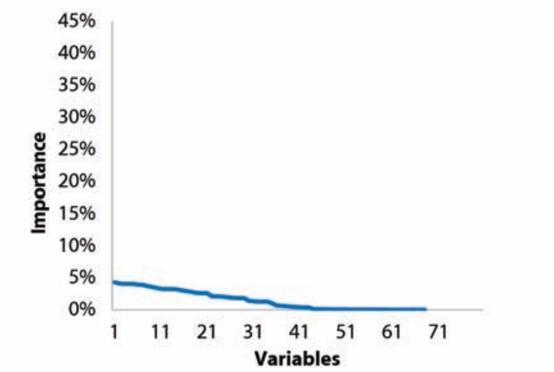


Figure 19: Relevance of variables in the "SVM a linear function" algorithm.

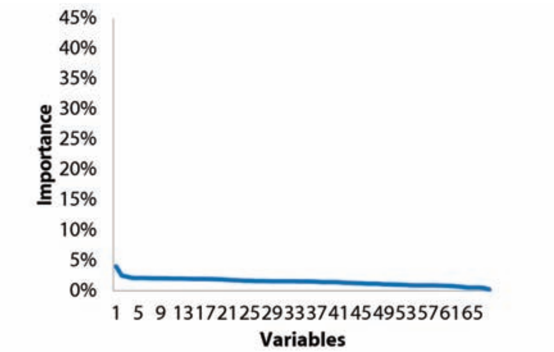
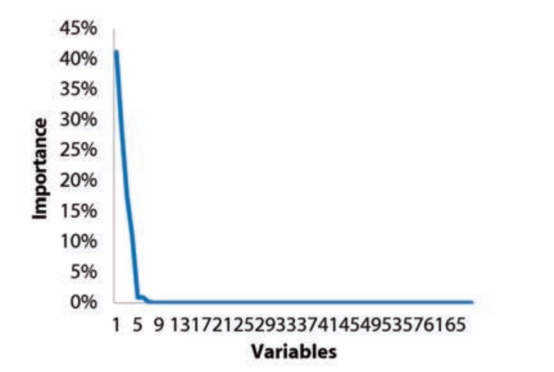


Figure 20: Relevance of variables in the "SVM with a radial function" algorithm.

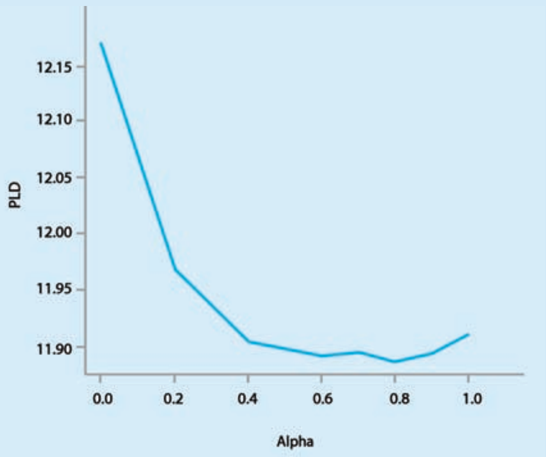


Detail of techniques used in the analysis[73]

Elastic net. As described in the previous section, the elastic net is applied by incorporating the term $\lambda_2\|\boldsymbol{\beta}\|_2^2 + \lambda_1\|\boldsymbol{\beta}\|_1$ into the target function that is to be optimized when estimating the parameters $\boldsymbol{\beta}^t = (\beta_1, \dots, \beta_n)$. To do this, the parameter is defined $\alpha = \frac{\lambda_2}{\lambda_1+\lambda_2}$ and the term $(1-\alpha)\cdot\|\boldsymbol{\beta}\|_2^2 + \alpha\cdot\|\boldsymbol{\beta}\|_1$ is estimated.

Then the partial likelihood deviance function, or PLD[74], is built as a function of the parameter α, and the function's minimum value is obtained. In this case, the minimum value is found at α=0.8 (Figure 21). The target function is trained using this value of α, resulting in variables whose associated estimator is close to zero and are therefore not used for building the model.

Figure 21. Optimization function for the elastic net method. The value of that minimizes the partial likelihood deviance function (PLD) is shown.



Random Forest. This technique is based on combining independent models (bagging), so that a large number of trees are generated and an average of their results is calculated. Growing each tree requires randomly assigning a random selection of the sample with replacement and partitions of the variables in order to reduce overfitting and control multicollinearity (Figure 22). To give a prediction for a new observation, the tree is fully applied. In the case of the classification trees, the mode of the response variable is often used as the prediction value, i.e. the node's most frequent class. Besides, the proportion of trees that take the same response is interpreted as the probability, which provides information on the prediction confidence level.

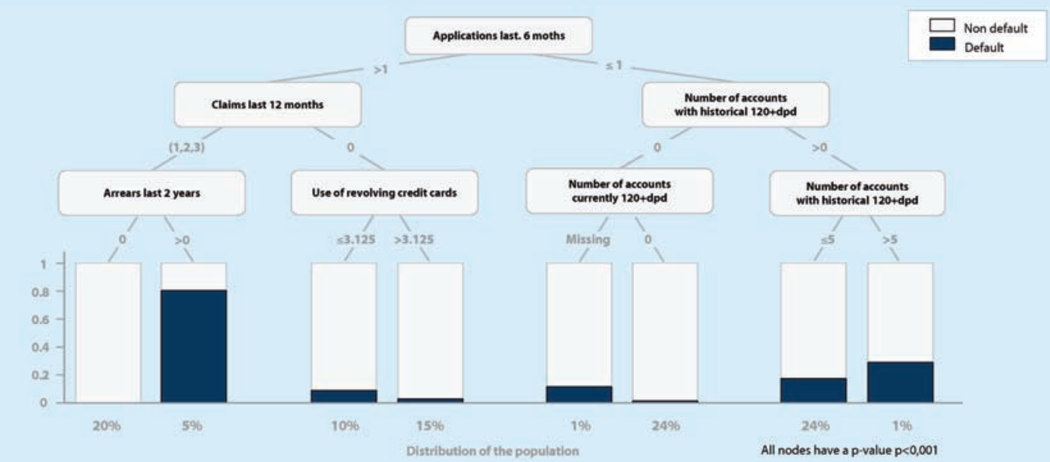
Adaboost. The use of this technique involves calculating different weak estimators on the sample that are later aggregated through boosting techniques using different weights, which in an iterative context allows the algorithm to learn from the errors of the previous iterations. The process is as follows:

- With the training sample, a submodel is created assigning equivalent weights for all observations.
- Relative weight is increased for all incorrectly classified observations, and decreased for correctly classified observations, and a new submodel is built taking the new weights into account. This is done until converging at a low error level.
- All submodels are sequentially combined, obtaining a complex final model based on simple intermediate submodels.

Use of Support Vector Machines (SVM). Two SVM models were used, a linear model and a radial model. The difference lies in how the kernel function that allows the groups to be separated is defined. The linear SVM function used is $K(\boldsymbol{x_i}, \boldsymbol{x_j}) = \boldsymbol{x_i}^t\boldsymbol{x_j}$ (the scalar product), whereas the radial SVM function is $K(\boldsymbol{x_i}, \boldsymbol{x_j}) = \exp\left(-\gamma\cdot\|\boldsymbol{x_i} - \boldsymbol{x_j}\|_2^2\right)$, where γ>0. The values used in the models built were γ=0.014 for the linear model and γ=1 for the radial model.

[73]Since these techniques were described in the previous chapter, some technical details associated with their specific use have been added.
[74]Partial Likelihood Deviance (PLD) is an indicator that compares the predictive power of a model under analysis with that of a reference model (which can be a constant model). It is calculated as the difference between the log-partial likelihood functions of the model under analysis and a reference model.

Figure 22. Example of segments of a tree for the Random Forest model.



Results and conclusions

In order to compare the results obtained in this analysis after using different techniques, a confusion matrix and the area under the ROC curve (AUC) were used as measures of discriminatory power:

- The confusion matrix was used to measure the model's success rate. For this, the cut-off point was set at the value that minimizes the prediction error.

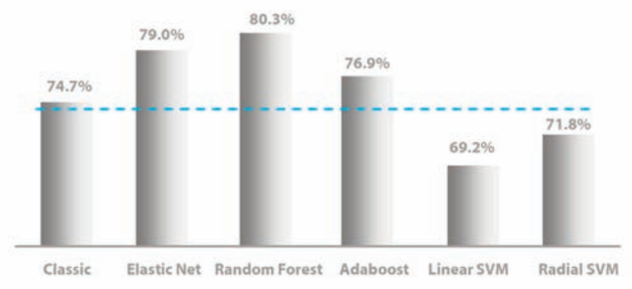
- As for the AUC indicator, the higher the value of the indicator, the greater the discriminatory power. Practical experience[75] shows that this indicator usually takes values within the range 75% to 90%. These indicators were estimated on a validation sample[76].

The impact on the business of the change in predictive power was also compared. To do this, the traditional model was chosen as the basis for comparison (with the cut-off point set at the optimization of type I and type II errors). Based on that model, the resulting default rate was estimated for each of the models analyzed assuming the same approval rate as with the traditional model (i.e. same level of business), and the approval rate was also estimated for each model assuming the default rate was the same as in the traditional model.

The following results were obtained[77] (Figures 23, 24 and 25):

- After a comparison against these statistics (Figure 24), a better performance was observed in the ensemble models compared to the remaining methods. The random forest in particular yielded the best results: the success rate improved from 74.7% using the traditional model to 80.3%, representing a percentage increase of 7.5%. As for discriminatory power, the area under the ROC curve improved from 81.5% using the traditional model to 88.2% a percentage increase of 8.2%. However, this improvement was achieved by increasing the estimation complexity, since many variables were used between all 50 trees.

Figure 24. Success rate in the confusion matrix (sum of the diagonal) for the different approaches.



- The second best method was the use of the elastic net, which achieved success-rate and area-under-the-ROC-curve values of 79% and 86.4%, respectively, representing a percentage increase of 6% for both indicators. As observed, the use of regularization techniques in samples with many variables resulted in a significant improvement with respect to traditional methods for selecting variables (stepwise).

- Likewise, in this particular case it was observed that the extra difficulty added by the use of SVM (both linear and radial) did not bring an improvement in the model's predictive power. This was due to the fact that the traditional logistic model already had a high discriminatory power (82%), which usually occurs when the variables allow the two classes to be linearly separated, and this means the use of SVM does not add information on the separability of classes (in fact, in this case predictive power was lost and the success rate went down).

[75] BCBS (2005).
[76] According to modeling practices, the total population was split into different training and validation samples, so that each model trained with the training sample was used to score the validation sample, and the validation results were estimated on said sample.
[77] These results were obtained on specific validation samples, although in a Machine Learning environment it is appropriate to assess the adequacy of results using different sub-samples (via bootstrapping or cross-validation techniques).

Figure 23. Results under the different approaches.

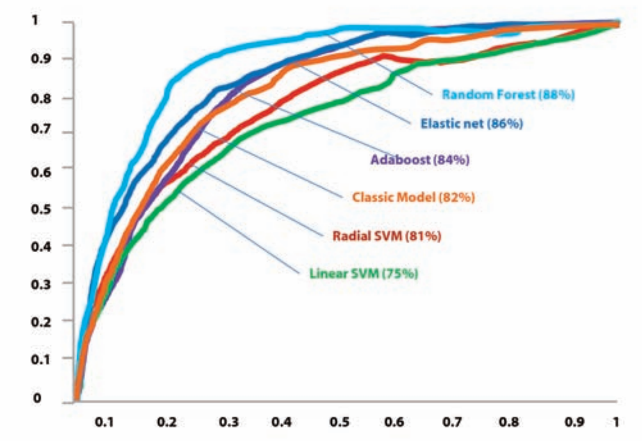
	Modelo tradicional		Red elástica		Random Forest		Adaboost		SVM Lineal		SVM Radial	
Matriz de confusión	70,3%	1,5%	74,4%	1,2%	75,6%	1,2%	72,4%	1,4%	65,2%	1,8%	67,6%	1,6%
	23,9%	4,4%	19,3%	4,6%	18,5%	4,7%	21,8%	4,5%	29,0%	4,0%	26,5%	4,2%
AUC	81,5%		86,4%		88,2%		83,7%		74,6%		81,1%	
Tasa de mora (igualdad de negocio)	2,1%		1,4%		1,1%		1,7%		2,7%		2,4%	
Tasa de aceptación (igualdad de mora)	71,8%		81,4%		83,2%		76,8%		41,8%		61,4%	

- As for the potential impact that the increase in each model's predictive power would have on the business, it was seen that, the volume of business being equal, the default rate would be reduced by 48% using the random forest (a default rate of 2.1% would result from an optimal cut-off point using the classical model, which would be reduced to 1.1% using the random forest provided the number of loans approved was maintained), and by 30% using the elastic net (with the default rate down to 1.4%). Likewise, the default rate being equal (2.1%), business volume would increase by 16% using the random forest, and 13% using the elastic net.

To be able to benefit from the improvements associated with these techniques, it is essential for companies to have a model risk management framework and function aligned with the highest quality standards (already addressed in the Management Solutions publication entitled "Model Risk Management: Quantitative and qualitative aspects of model risk management[78]"), which should include having the Internal Validation function effectively challenge the assumptions used and the results obtained from the models through the use of techniques such as model replication and, for more advanced entities, the development of "challenger" models using traditional techniques.

[78] Management Solutions (2014)

Figure 25. ROC curves obtained for each technique.



The confusion matrix at the cut-off point that minimizes errors

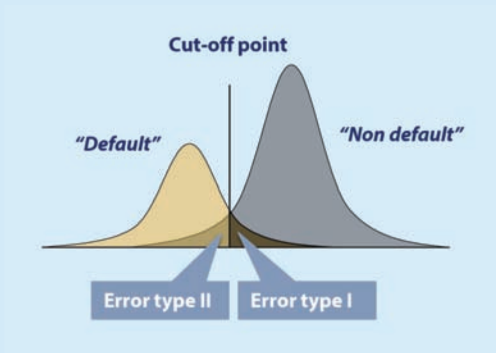
The confusion matrix is a technique for measuring the discriminatory power of a model. This matrix compares a model's prediction with the actual result. In the case of a scoring model, it makes it possible to compare the level of default predicted by the model with the level that actually occurred (Figure 26). This requires calculating the success rate (defaults correctly predicted by the model and operations classified as non-defaults that did not eventually default), and the errors in the model (known as type I and type II errors[79]).

Figure 26. Confusion matrix.

		Actual	
		Non Default	*Default*
Predicted	*Non Default*	Correct inference	Error type I
	Default	Error type II	Correct inference

Since models produce a score (which is a ranking of the operations based on their credit quality), a cut-off point needs to be set in order to determine whether a default is being predicted by the model. A well accepted methodology in the industry is to set a cutoff point that simultaneously minimizes both errors (Figure 27).

Figure 27. Cutoff point that simultaneously minimizes type I and type II errors.



[79] Percentage of defaults erroneously classified by the model, and percentage of false defaults erroneously predicted by the model, respectively.

Bibliography

Acuity Market Intelligence (2016). "The Global Biometrics and Mobility Report: The Convergence of Commerce and Privacy". Market Analysis and Forecasts 2016 to 2022.

BCBS (2005). "Studies on the Validation of Internal Rating Systems". Bank for International Settelments.

Bengio, Y. and others (2015). Bengio, Y. LeCun, Y. Hinton, G.: "Deep Learning"(2015) Nature.

Box, G. and Draper, N. (1987). "Empirical Model-Building and Response Surfaces". Wiley.

Breiman, L. (2001). "Statistical Modeling: The two cultures", Vol 16 (3), Statistical Science, 2001. pp. 199-231.

Brynjolfsson, E. and McAfee, A. (2017). "The business of artificial intelligence". Harvard Business Review.

Cios, K.J. and others (2007). Cios, K.J., Pedrycz, W., Swiniarski, R.W., Kurgan, L.: "Data Mining: A Knowledge Discovery Approach". Springer 2007.

Círculo de Empresarios (2018). "Alcance e implicaciones de la transformación digital: principales ámbitos de actuación".

Cisco (2017). "The Zettabyte Era: Trends and Analysis". White paper. June 2017.

Cybersecurity Ventures (2017). Cyber economy research.

Dahr, V. (2013). “Data Science and prediction”. Association for Computer Machinery.

Digital Consumption (2016). World Economic Forum White Paper Digital Transformation of Industries.

Domingos, P. (2012). “A few useful things to know about Machine Learning”, Vol 55 (10), Communications of the ACM, 2012. pp. 78-87.

Domingos, P. (2015). “The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World”. 2015.

EC (2017). Digital Transformation Scoreboard 2017. European Commission.

Fitzgerald et al. (2013). "Embracing Digital Technology: A New Strategic Imperative, Sloan Management Review

Forbes (2015). “Big Data: 20 Mind-Boggling Facts Everyone Must Read”.

Forbes (2016). “Artificial Intelligence Is Helping Doctors Find Breast Cancer Risk 30 Times Faster”.

Forbes (2017). Janakiram MSV: “3 Key Machine Learning Trends To Watch Out For In 2018”. Forbes, diciembre 2017.

Friedman, J. H. (1997). “Data mining and statistics: what's the connection?” Presented at the 29th Symposium on the Interface: Computing Science and Statistics, May 14–17, 1997, Houston, TX, pp. 3–9.

Frost & Sullivan (2017). “Global Forecast of RegTech in Financial Services to 2020”.

International Telecommunication Union (2009). “Distributed Computing: Utilities, Grids & Clouds”. ITU-T Technology Watch Report 9.

International Telecommunication Union (2017). ITU World Telecommunication/ICT indicators database 2015.

Jordan, M. I., Mitchell, T. M. (2015). “Machine learning: Trends, perspectives, and prospects”, Vol 349 (6245), Science, 2015. pp. 255-260.

LEGOs (2018). Still “The Apple of Toys”? Harvard Business School

Management Solutions (2014). “Model Risk Management: Quantitative and qualitative aspects”.

Management Solutions (2015). “Data Science and the transformation of the financial industry”.

McKinney (2018). “Python for Data Analysis”. O’Reilly Media.

Ministry of Industry, Energy and Tourism. Government of Spain (2018). "New trends and challenges in the world of data”, March 2018.

Mitchell, T. (1997). Machine Learning. McGraw Hill.

Murphy, K. (2012). “Machine Learning: A probabilistic perspective”. The MIT Press.

European Parliament (2016). Regulation(UE) 2016/679 of the European Parliament and of the Council of April 27, 2016.

Pearson (2016). Lukin, R. et alter: “Intelligence Unleashed: An argument for AI in Education”. Pearson.

Samuel, A. (1959). “Some studies in machine learning using the game of checkers”. IBM Journal. July 1959.

Schwab, K. (2016). “The Fourth Industrial Revolution”. World Economic Forum.

Shalev-Shwartz, S. y Ben-David, S. (2014). “Understanding Machine Learning: From Theory to Algorithms”. Cambridge University Press.

Statista (2017). “Number of Machine-to-Machine (M2M) connections worldwide from 2014 to 2021”. 2018.

Statista (2018). “Digital Market Outlook”. 2018.

Techemergence (2017). “How America’s 5 Top Hospitals are Using Machine Learning Today”.

The Economist (2017). Axel Springer’s digital transformation.

Turing, A.M. (1950). “Computer Machinery and Intelligence”. Mind 49: p433-460.

Westerman, G. (2017). “Your Company Doesn’t Need a Digital Strategy”. MITSloan Management Review.

Glossary

Auto-encoder: is a neural network used to learn efficient encodings (e.g. dimensionality reduction, compression and encryption).

Big Data: any voluminous amount of structured, semi-structured or unstructured data that has the potential of being extracted in order to obtain information. By extension, it also refers to all of the IT infrastructures and architectures that store such data.

Blockchain: data structure that implements a cryptographic record of all operations that are performed and previously validated by a network of independent nodes through a consensus algorithm.

Bootstrapping: resampling method used to approximate statistical sampling distribution. Frequently used to approximate the bias or variance of a statistical analysis, and to build confidence intervals or test assumptions on parameters of interest.

Business to Business (B2B): commercial transactions between companies, i.e. those that take place between a manufacturer and a distributor, or between a distributor and a retailer.

Computer Vision: artificial intelligence subfield whose purpose is to program a computer to understand a scene or the characteristics of an image.

Cost function: function which, based on a value of a model's estimates, provides a measure of the mistake that is made when using the model. The model training process usually consists of obtaining the estimates that minimize the cost function.

Cybersecurity: set of tools, policies, methods, actions, insurance and technologies that can be used to protect the different IT assets and data of an organization stored in some kind of physical or logical support in the cyber-environment, as well as the communications between such assets.

Data dredging: Data Mining analyzes large data volumes looking for potential relationships between the data. Data dredging occurs when relationships are found that are not supported by hypotheses or causes really explaining those relationships.

Function as a Service (FaaS): Cloud programming services category that provides a platform for users to develop, use and manage an application's functionality without the complexity of building and maintaining the infrastructure typically associated with that application.

Graphics Process Unit (GPU): coprocessor dealing with the processing of floating point graphics or operations in order to reduce the workload of the Central Processing Unit (CPU).

Greedy: algorithm in which each element to be considered is evaluated once, being either discarded or selected, so that if it is selected it becomes part of the solution, and if it is discarded it is neither part of the solution nor will it again be considered for the same solution.

Internet of Things (IoT): digital interconnection of everyday objects (advanced connection of devices, systems and services) with the internet, allowing for data on these objects to be sent and received without the need for human involvement.

k-fold cross validation: cross- sample validation process consisting of dividing the sample into k groups, and iteratively using each group for validation and the rest for training, changing the validation group in each iteration.

Knowledge discovery: process of identifying patterns in the data that are valid, novel, useful, and understandable.

Machine to Machine (M2M): connectivity between machines where, once an action is performed, the data generated by the different digital components involved are connected to servers in order to store and analyze the information.

Markov chain: discrete stochastic process in which the probability of an event occurring depends solely on the previous event.

Orthogonal partition: partition of an n-dimensional space into regions through hyperplanes perpendicular to each of the axes defined by the explanatory variables.

Overfitting: characteristic that occurs when the model has been fit to the training sample too closely, so that it does not achieve satisfactory results on samples other than this one (e.g. on the validation sample).

Partial likelihood deviance (PLD): indicator that is used to compare the predictive power of a model under analysis with that of a reference model (which can be a constant model). It is calculated through the difference in the log-partial likelihood functions of the model under analysis and the reference model.

Peer to Peer (P2P): computer network allowing the direct exchange of information in any format between the interconnected computers.

Principal Component Analysis (PCA): a technique used to describe a set of data in terms of new uncorrelated variables (components). Since components are ranked by the original variance they describe, this technique is useful to reduce the dimensionality of a data set. PCA seeks the projection that best represents the data in terms of least squares.

RegTech: combination of the terms "regulatory technology", referring to a company that uses technology to help other companies comply with regulations efficiently and at a reduced cost.

Regularization: mathematical technique consisting of adding a component to a cost function in order to detect those variables that are not contributing significantly different information to the model. It is used to avoid overfitting problems (as in the case of elastic nets).

Resampling: set of methods aimed at obtaining new data samples from a data set that is representative of the same population (i. e. bootstrapping). It is mainly used to create training and validation subsets from the original sample.

Roboadvisor: automatic algorithm that provides online advice and management with minimal human intervention. This service is typically provided for the web-based management of financial investments, which facilitates the creation of personalized investment portfolios, adapted to each person´s own profile.

Robotics Process Automation (RPA): software that replicates human actions by interacting with the user interface of a computer system.

ROC curve (Receiver Operating Characteristic): curve used to analyze the predictive power of a binary output model. It represents the relationship between type 1 error (incorrectly classifying adverse events) and type 2 error (incorrectly classifying favorable events).

Smart contract: programmable contract intended to enforce business rules, whose code is registered and can be executed in a distributed way by the different nodes on the network.

Stepwise: iterative method for building models based on the automatic selection of variables.

Our aim is to exceed our clients' expectations, and become their trusted partners

Management Solutions is an international consulting services company focused on consulting for business, risks, organization and processes, in both their functional components and in the implementation of their related technologies.

With its multi-disciplinary team (functional, mathematicians, technicians, etc.) of 2,000 professionals, Management Solutions operates through its 24 offices (11 in Europe, 12 in the Americas and 1 in Asia).

To cover its clients' needs, Management Solutions has structured its practices by sectors (Financial Institutions, Energy and Telecommunications) and by lines of activity (FCRC, RBC, NT), covering a broad range of skills -Strategy, Commercial Management and Marketing, Organization and Processes, Risk Management and Control, Management and Financial Information, and Applied Technologies.

The R&D department provides advisory services to Management Solutions' professionals and their clients in quantitative aspects that are necessary to undertake projects with rigor and excellence through the implementation of best practices and the continuous monitoring of the latest trends in data science, machine learning, modeling and big data.

Javier Calvo
Partner at Management Solutions
javier.calvo.martin@msgermany.com.de

Manuel A. Guzmán
R&D Manager at Management Solutions
manuel.guzman@msspain.com

Daniel Ramos
R&D Methodologist at Management Solutions
daniel.ramos.garcia@msspain.com

Management Solutions, Professional Consulting Services

Management Solutions is an international consulting firm whose core mission is to deliver business, risk, financial, organisation, technology and process-related advisory services.

For further information please visit **www.managementsolutions.com**

Follow us at:

www.managementsolutions.com

Madrid Barcelona Bilbao London Frankfurt Paris Warszawa Zürich Milano Roma Lisboa Beijing New York Boston Atlanta Birmingham San Juan de Puerto Rico Ciudad de México Medellín Bogotá São Paulo Lima Santiago de Chile Buenos Aires