

# Advanced Analytics for Managerial Decision

by Emanuele Borgonovo, Luca Molteni\*

## MANAGEMENT & APPLICATION

**T**he data-driven economy has emerged in recent years. The digital revolution, the increase of connectivity, and the instant spread of information are revolutionizing the way we operate and work, at the individual, company, and institutional level. There is increasing recognition of the importance of data. The period of emergency we are living through is also pushing us to use quantitative information more and more: the degree of freedom of millions of citizens is decided based on data on the spread of the pandemic, that after being studied through mathematical-statistical models, determines the color of the region we live in.

**H**umanity is creating growing quantities of data, and leaves enormous digital trails of its tastes, needs, consumption, activities, and interests. Properly trained machines are able to perform increasingly sophisticated tasks that in many cases help human beings and amplify their abilities and results. So we speak more and more of artificial intelligence (AI). On the one hand, we have that part of AI that stirs our imagination: autonomous robots that substitute humans in futuristic, completely automated cities (but this is fortunately still far off). On the other we have AI that, based on machine learning, is very close to us and is already having a significant impact on our everyday lives. The digitalization of information has almost completely replaced information on paper, making available large quantities of data that reside in the training systems of businesses. The new systems of data analysis are provoking a large change in the way managerial decisions are made. A growing number of both small and large businesses recognize the potential residing in the data at their disposal. Discovering



\* Emanuele Borgonovo is Full Professor at the Department of Decision Sciences of the Bocconi University, Professor of Decision Sciences & Business Analytics and Director of the Management Science Laboratory at the SDA Bocconi School of Management.

Luca Molteni is a Researcher at the Department of Decision Sciences of the Bocconi University and the Department of Decision Sciences Liaison Officer at the SDA Bocconi School of Management.

trends, patterns, and behavior in advance is a factor of competitive advantage.

### THE “DIGITAL ECONOMY”

From an institutional point of view as well, for some years now there has been recognition of the potential impact of the new methods of analysis. In 2012, the Agency for Digital Italy (Agid) was born; in 2014, a document presented to the European Parliament by a specific commission spoke expressly of the “digital economy.” The commission’s report highlighted how the economy was (and still is today) about to be overwhelmed and upset by an enormous mass of data generated by humanity, and by a just as fast digitalization of decision-making processes. The document stresses that new types of skills are becoming essential: in particular, professionals are needed who know how to transform the information produced by calculators and advanced analytics algorithms into indications relevant for decision-makers, institutions, and businesses. The commission’s report also indicates the need for joint training efforts by universities and industry to succeed in filling the professional gap, which was already significant at the time and expected to grow even more in the subsequent decade. In that sense, we can stress the birth of numerous new programs at the university level created specifically to train the professional figures able to meet the challenges of data science.

### MACHINE LEARNING TO SUPPORT DECISIONS

Some critical reflections on the revolution underway are necessary. The first is positive: the fact that increasingly reliable data and predictions are used to make decisions cannot but lead to an improvement in the decision-making process. As Laplace already pointed out two centuries ago, the support of a theory or a quantitative formulation provides indispensable support for the weakness of the human mind at the moment it must make difficult decisions. On the other hand, though, the danger is that of leaving decisions to be made automatically by machines without human supervision or intervention. In fact, an advanced analytics algorithm blindly follows preset rules dictated by lines of code that in any case represent an abstraction of the problem and cannot take into account all of the possible combinations and all of the peculiarities of specific situations. Think of an algorithm that is able to grant or deny a loan. If we were to use the following as a simplified rule of the algorithm: “if the person has been unable to pay back a loan in past years, they shall not have the right to a new loan,” then all the persons who, even only to a small extent, did not make a payment in previous years will not have access to a new loan for an indefinite period of time, even if their financial situation has clearly improved in the meantime. Therefore, we would have an unjust, but automatic de-

nial of the right to loan access. Therefore, lawmakers set time limits. So let’s try again, with a different rule: “If the requesting person has been unable to repay a loan in the past three years, they may not receive a new loan for the next three years.” In this case, all of the people who, even to a small extent, did not make payment of an installment in the previous three years, will not have access to a new loan for the three years following their late payment. However, a separate examination of the financial situation of the requesting party or a significant improvement of their financial situation could induce a risk manager to propose an override of the rule, and anyway grant the loan; based on an improvement of the surrounding conditions that would not be recognized by the rule by which the algorithm makes the decision.

The question is, in the presence of an algorithm that analyzes thousands of data points and produces a signal of “no, we will not grant the loan,” how can a human being, based on specific sensations, reverse that analysis? Is a human being able to assume this responsibility?

### THE IMPACT OF COVID-19 ON DATA SCIENCE

One of the most critical aspects we are dealing with in this period is certainly the impact of Covid-19 on our lives, activities, and institutions. We are more isolated than before, with much more digital contact, forced to revise our modes of consumption. It is clear that the pandemic will have a non-negligible impact on the world of data science. Prediction models – and in particular those based on historical statistics – have been and in the coming months will be strongly affected by the discontinuity in those statistics, suggesting the need to incorporate more qualitative research tools that consider the scenario of demand behavior.<sup>1</sup> Moreover, if on the one hand the importance of the analysis of data in general and big data in particular emerges even more clearly in a critical period such as this one, on the other, the investments by companies facing difficulties could decline due to the drop in their margins; investments could also decline in relation to the hiring of dedicated personnel, and for the acquisition of external analytics services judged not to be essential, and lastly, also for the acquisition of data management and analysis tools. The increase in smart working is also creating greater security threats for companies, since it is accompanied by a forced acceleration towards digital and the exchange of documents and information outside of the company’s premises; these working conditions could thus favor additional investments in the area of cybersecurity.

More in general, in these months of great difficulty companies are asked to be more efficient and have a greater ability to exploit the potential of machine learning applications, with effective solutions able to concretely assist business decision-making processes.

## THE POTENTIAL OF MACHINE LEARNING

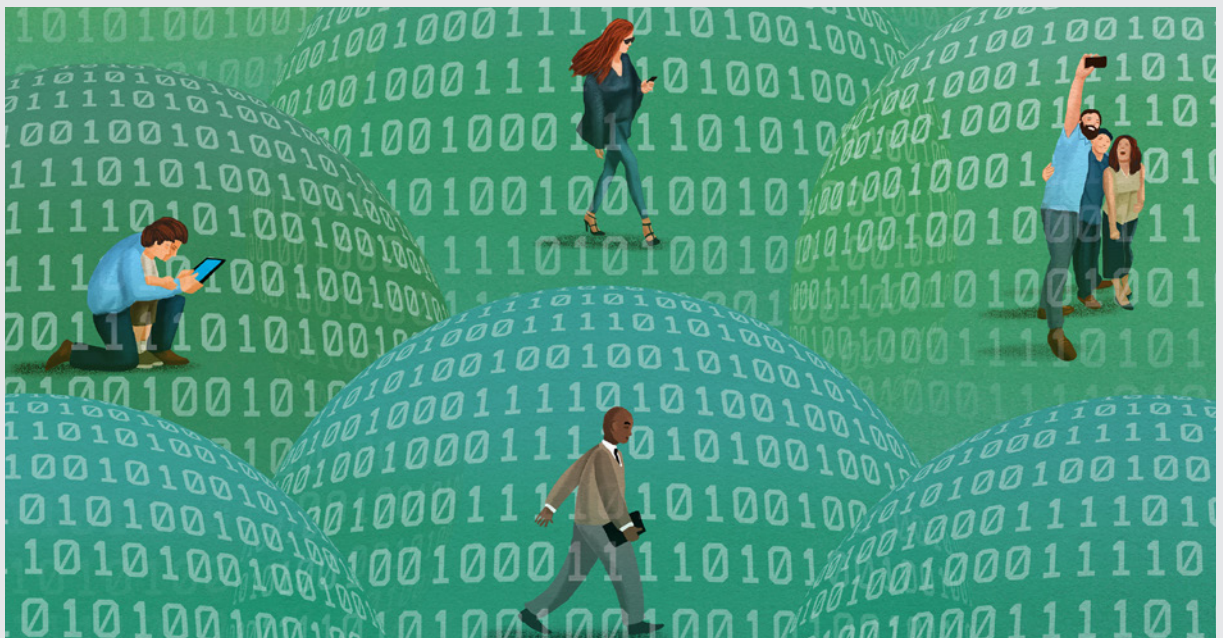
Regardless of the difficulties of the moment linked to the pandemic, some advanced machine learning algorithms, such as the most recent evolutions of artificial neural networks (deep learning), have by now become the most widespread method for a vast range of tasks. In applications such as the classification of images, the recognition of speech and the processing of natural language, deep learning networks show surprising performance, with predictive precision that often exceeds that of human beings. In 2016, Google DeepMind's AlphaGo beat Lee Sedol, one of the best go players of all time. Go is a strategic board game that is very popular especially in China and Korea. With respect to chess, it is based heavily on visual intuition, and until 2016 it was thought a computer could not beat a human due to the type of tasks the game requires. For chess, on the other hand, we know that humans have been regularly beaten by the computer since the famous match between Deep Blue and Kasparov in 1997. In the duel between Alpha Go and Lee Sedol, the twenty-sixth move was the key to the victory, even though it wasn't clear during the game why this move was advantageous. After that specific game, Fan Hui, an expert go player, stated: "It's not a human move. I've never seen a human play this move." So it should be stressed that by now, the most advanced machine learning algorithms scan enormous amounts of data and could detect relationships and paths that human beings are unable to perceive. This learning by the system becomes of interest in various areas, for example for physicists and chemists, finding hidden laws of nature, or as mentioned previously, showing new possible strategies of action: Go players have incorporated the AlphaGo move in their repertoire, while professional chess players regularly train against computers.

## THE INTERPRETABILITY OF RESULTS

The increase of machine performance is due to the improvement of the models and methods of analysis, the presence of better and better data, and greater computing power, especially thanks to the use of Graphics Processing Units (GPU) and parallel computing. But this performance increase is not without costs; these models with high predictive performance are characterized by a lack of interpretability and transparency in the relationship between inputs and outputs that lead to specific predictions.

Breiman, among others, has emphasized the trade-off between performance and interpretability.<sup>2</sup> For a long time, it has thus been common opinion that the simplest models are easier to interpret, but at the cost of inferior predictive performance.<sup>3</sup> In recent years, one of the efforts of researchers who work on the development of advanced models has been to weaken this assertion, developing methods that maintain high predictive performance but provide greater interpretability.

It should be stressed that the importance of the interpretability aspects of models is not uniform in its different applications. If we think of an algorithm that recognizes images or speech or also in some tasks such as playing go, the key point is that performance and interpretability is a minor goal. But there are other applications in which it is inevitable to seek a balance between interpretability and predictive capacity, and the majority of managerial applications certainly fall into this category: for example, think of the problems linked to the propensity to purchase a specific product by a customer, the issue of customer loyalty (churn/attrition), or that of the behavioral segmentation of customers. All of these applications require the analyst to be able to provide an interpretation of results both



to communicate them to management, and to be able to implement the most appropriate actions to reach the goal of improving the business conditions, that are the premise of the very idea of applying a machine learning model.

Moreover, the interpretability of a model can also reveal its weaknesses, for example providing an explanation of why it fails or even showing that the decision suggested is mathematically correct, but based on imprecise or misleading data. Being fully aware of the weaknesses of an algorithm is the first step towards improving it.

Interpretability also goes hand in hand with the issue of trust. A decision-maker cannot blindly trust predictions when they do not know how the results have been obtained. This is particularly true in delicate fields such as medicine, where it is essential to know why a certain diagnosis has been made by the machine. A doctor must verify the output of the black box, in order to fully understand the reasons and be able to trust the machine's suggestions. The conditions are very different in a managerial context, since the sharing of the choices made and their motivations is almost always a fundamental element of the business decision-making process.

Furthermore, again linked to the question of trust, we must seek to adequately quantify uncertainty (and to date we are still far from this goal). This aspect is also very important in sectors such as medicine.<sup>4</sup> Modern

black boxes based on data analysis, and in particular medical AI, need to develop a strict principle of quantification and regulation of uncertainty (uncertainty quantification).

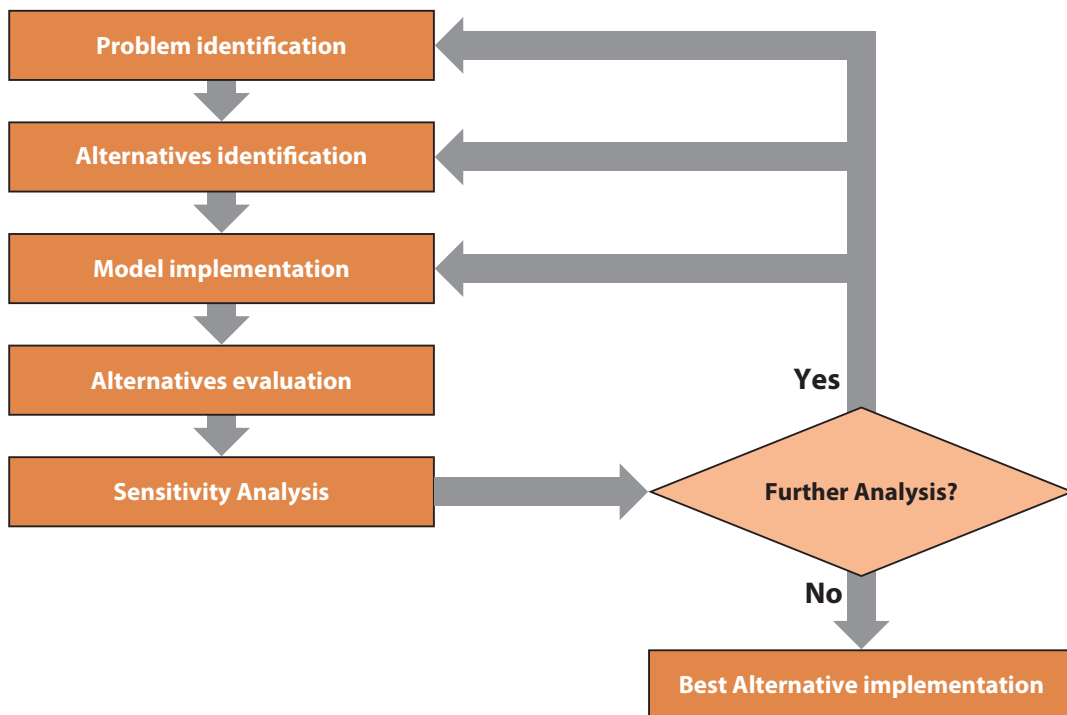
In the context of this discussion, legal issues also cannot be ignored. With the growing role of AI algorithms, the request for regulation increases. Decision-makers are often obliged to justify the decisions they make, especially in delicate sectors such as banking and criminal law. It is known that algorithms can learn rules and use data that, due to the way they are collected, conceal some forms of distortion; so returning to the initial example, people with certain specific characteristics could have a lower likelihood of obtaining a loan.<sup>5</sup> On this point, the European Union has adopted a "right to explanation," a new regulation that allows users to request explanations on decisions made with the support of artificial intelligence. This is an area in which the interpretability of models plays an extremely important role.

#### THE MODERN DECISION-MAKING PROCESS<sup>6</sup>

We thus come to the modern decision-making process. The process has been studied for a long time in the literature.<sup>7</sup> In figure 1, the main steps are illustrated.

The first phase is that of identifying the problem. For example, a company that wants to better understand the perceptual positioning of its brand on the

**Figure 1 - The modern decision-making process**



market could be interested in performing a customer churn analysis.

Among the available alternatives we can consider that of performing it in house through the company's own personnel, or entrusting the analysis to an outside supplier. At equal time and financial resources, the first option depends on the decisive condition of having the necessary internal skills to do it.

The next step is to formulate a model. This is the sum of the predictive algorithm and the available data on which to carry out those predictions. We are entering a complex world here, that of the preparation of the data and the selection of the predictive algorithm. In a certain sense, the available alternatives, in principle, contain infinite mathematical models. For example, let us consider a churn analysis. In this case, the analyst could choose to opt for a logistic regression model, a decision tree classification model, a random forest, or a gradient boosting model. But it is clear that not all of the variables make sense for the problem in question. Once the ones pertinent for the problem are identified, it is necessary to then collect quantitative information on them, orienting the data collection.

Once the model is implemented, the results obtained will be examined. From the analysis of the model's

sensitivity and strength, and a careful examination of the process that has taken place to that time, it will be possible to determine if the results are satisfactory. In this case, we can move to the implementation phase. Otherwise, we return to repeat and refine some of the previous activities so as to arrive at the most robust decision possible.

The development of the discipline is such that many companies are creating internal data science teams. This choice has the advantage of allowing the company to internalize the necessary skills, enriching the company's human resources in a critical function for providing support to management. It also has the advantage of managing data internally, that on the one hand contributes to minimizing the risks linked to the question of privacy, and on the other allows for constant and direct monitoring.

Obviously, the creation of a data science team also entails costs associated with human and IT resources (computation and storage), that need to be entirely covered by the company and are constant in time. On the other hand, if the size of the company or the investment possibilities are relatively limited, the most flexible and practicable solution is that of outsourcing the data management and/or analysis process through the use of qualified suppliers.<sup>8</sup>

---

<sup>1</sup> A. Dearmer, "How Has Covid-19 Impacted Data Science?", *Xplenty*, December 16, 2020.

<sup>2</sup> L. Breiman, "Statistical Modeling: The Two Cultures", *Statistical Science*, 16(3), 2001, pp. 199-231.

<sup>3</sup> C. Rudin, "Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead", *Nature Machine Intelligence*, 1(5), 2019, pp. 206-215.

<sup>4</sup> On this point, see E. Begoli, T. Bhattacharya, D. Kusnezov, "The Need for Uncertainty Quantification in Machine-Assisted Medical Decision Making", *Nature Machine Intelligence*, 1(1), 2019, pp. 20-23.

<sup>5</sup> Rudin, *op. cit.*

<sup>6</sup> This subsection is in part borrowed from E. Borgonovo, L. Molteni, *Quando ai manager danno i numeri*, Milan, Egea, 2020.

<sup>7</sup> See, among others, R.T. Clemen, *Making Hard Decisions: An Introduction to Decision Analysis*, Pacific Grove, Duxbury Press, 1997.

<sup>8</sup> In the book *Quando ai manager danno i numeri*, *op. cit.*, we set out to introduce the reader to this new and complex discipline, presenting an overview of the main data ingestion, data management, and data analysis methods, with particular attention also to the issues of data preparation and the evaluation of predictive models. The same book also illustrates some applications selected in the business world, based on small and big data.

---

## SYNOPSIS

- The balance between the predictive ability of machines and the interpretability of data by management is a fundamental requirement today to reach a company's goals and improve its business.
- In many managerial contexts, the interpretability of data goes hand in hand with the issue of trust: a decision-maker will clearly have some hesitation to blindly trust predictions developed by artificial intelligence systems if they have no idea how they have been obtained.
- The definition of an adequate decision-making model based on the information produced by advanced analytics algorithms is increasingly entrusted by companies to an internal data science team. This choice has the benefit of allowing the company to internalize the necessary skills and manage its own data independently, minimizing cybersecurity risks.