

# DATA AND MACHINE LEARNING IN DECISION-MAKING

All models are incomplete and approximate. Modelling is extracting from a reality that is always chaotic and complex, the essential characteristics for the decision-making process in question, organizing, simplifying, and creating meaning and purpose. This is only possible at the cost of a high dose of abstraction and simplification.

**PEDRO AMORIM** <sup>(1,2)</sup>  
**JOSÉ FERNANDO OLIVEIRA** <sup>(1,2)</sup>

<sup>(1)</sup> INESC TEC

<sup>(2)</sup> FACULTY OF ENGINEERING OF THE UNIVERSITY OF PORTO

pedro.amorim@inesctec.pt

jose.f.oliveira@inesctec.pt

In the summer of 2019 in Seattle, we attended a talk by Jeff Wilke, global CEO of Amazon's Consumer Business. The first slide of that presentation was split into two. On the left, the slide was subtitled Decision Support and showed an analyst analysing the output of a mathematical model, and on the right side, the subtitle was Hands-Off Wheel and showed an analyst programming a model that made decisions autonomously. With the following click, Jeff Wilke put a cross on the left side and simultaneously commented that at Amazon the way to make decisions should be like the one described on the right side of the slide - investing the time necessary for the model to best portray the decision to be made, but without interfering with the result.

Decision-making supported by analytical models is usually described on a scale of three main categories (Figure 1). The first category - descriptive analytics - concerns models that support the understanding of past events. For example, when analysing a retailer's promotional campaign, these models can identify how effective and efficient that activity is. In the second category - predictive analytics - the goal becomes to anticipate the impact of a certain business activity. Using the same example, with these models, the retailer could predict the sales of a particular promotional campaign. Finally, in the prescriptive analytics category, mathematical models have the responsibility of suggesting actions that are then analysed and refined by decision makers. Going back to the retailer's case, these models would suggest the best parameterization of the promotional campaign given a business objective and constraints.

This last category of analytical models - prescriptive analytics - currently brings many challenges in terms of adoption by organizations. These challenges are rooted

in the fact that decision-makers in these organizations do not believe it is possible to codify and improve the current decision-making process. This baseline position causes the requirements for modelling not to be fully mapped out and the mathematical description of the problem to fall too short of reality. Even beyond these initial challenges, changing the decision-making process is always challenging and transformational in nature. This reality makes it necessary to accompany the technical rigor of model development with a practical sense of changing minds and habits.

Going back to Jeff Wilke's presentation, it is clear that Amazon has extended the scale of analytic models and brought autonomous analytics to the forefront. This category has at its foundation a distinct stance on the development and application of decision-making models. Being a natively digital company, Amazon's employees have never made decisions in any other way and this makes it easier to overcome the challenges elicited for prescriptive analytics. With the intensive use of this category of analytical models, Amazon puts substantial effort into the development stage, using successive iterations. Thus, considering, again, the case of defining promotional campaigns, Amazon will attempt to determine, after multiple experiments, the price elasticity profiles of different customer segments and model comprehensively the corresponding business dynamics. In use, these models, as they have no human intervention downstream, will produce systematic deviations that can be continuously analysed and refined.

All models are incomplete and approximate. Modelling is extracting from a reality that is always chaotic and complex, the essential characteristics for the decision-making process in question, organizing, simplifying, and creating meaning and purpose. This is only possible at the cost of a high dose of abstraction and simplification. In a simple way, and quoting George Box<sup>[1]</sup>, all models are

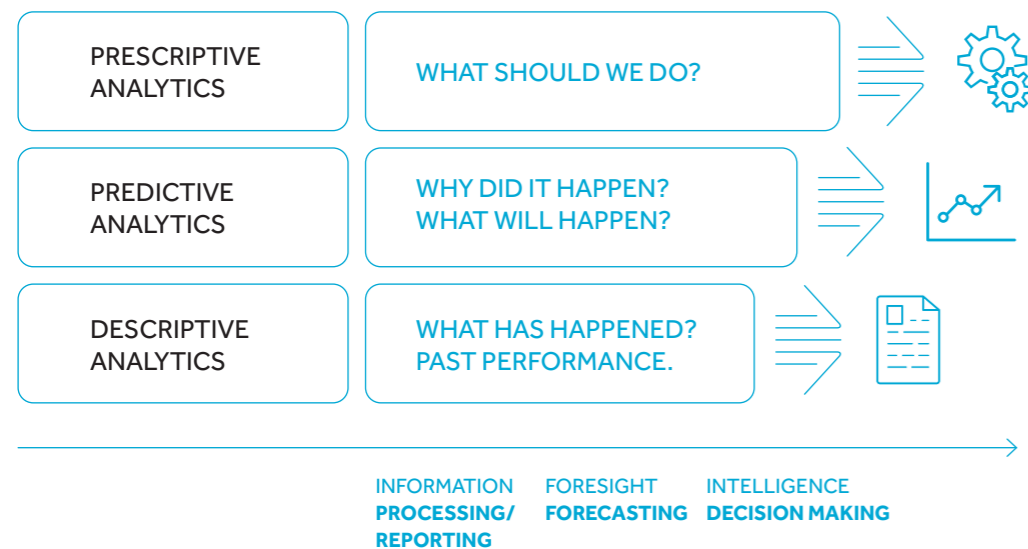


iiLab - Industry and Innovation Lab and Innovation Lab

wrong, some are useful. George Box derives from this assumption two important conclusions. The first is that as all models are wrong, it is not possible to obtain the "correct" model by overlaboration. The second, which follows from the first, is that if we must live with the error, we have to be particularly attentive to those aspects where error is important and relevant. What distinguishes the models of prescriptive analytics from the models of autonomous analytics is the focus on human intervention and, consequently, the sources of subjectivity and error. Thus, prescriptive analytic models carry within themselves the subjectivity of the analyst who decided what was relevant or not to the quality of the decisions to be obtained, incorporating in greater or lesser detail such features into the model. If the model automatically generates decision proposals, it does so based on the rules and objectives mathematically modelled by the analyst. The validation of these models is done by the controlled feeding of data, which allows the verification and validation of the results. There is thus a huge ethical responsibility on the part of the analyst in the construction of the prescriptive analytics model. Autonomous analytic models seek to be immune to the analyst's subjectivity themselves, building the decision rules based on huge volumes of historical data, which allow correlations between actions and consequences to be established. But the "Achilles heel" of autonomous analytics is exactly that correlations are not cause-and-effect relationships. On the other hand, these models

also simplify the data used by selecting the features that have the most impact on the correlations, they are also wrong and can produce significant systematic biases, as we stated earlier. Thus, these models require human intervention in search of these deviations, which will also be fraught with subjectivity. If the error is inherent to the use of analytical models, autonomous or otherwise, then human subjectivity will also always be present, and, consequently, ethical considerations. The discussion around which analytical methodology is more permeable to human (lack of) ethics has advocates on both sides of the barricade, but it will only be serious if we keep in mind that decisions will always have to be made by concrete women and men, who, informed by science and technology, cannot alienate the ultimate responsibility of the decision. By doing so, we will dehumanize our society.

Figure 1 - Scale of the of the three key analytics models



iiLab - Industry and Innovation Lab

[1] Box, G. E. P. (1976), "Science and statistics", Journal of the American Statistical Association, 71 (356): 791-799, doi:10.1080/01621459.1976.10480949