Why data analytics is an art

Reference to this article should be made as follows: Charles, V., Emrouznejad, A., Gherman, T., & Cochran, J. (2022). Why data analytics is an art. *Significance*, *19*(6), 42-45. DOI: https://doi.org/10.1111/1740-9713.01707

Why do so many data analytics projects fail? In 2015, Gartner Research[1] estimated that 60% of big data projects would fail within the next two years; in 2017, it was revealed that the actual percentage was closer to 85%. Harvard Business Review[2] reported that a four-year study of major companies that had implemented major analytics initiatives revealed that only slightly more than one-third of the 36 companies studied met the long-term objectives of their analytics initiatives. Most recently, Gartner[3] predicted that 80% of analytic insights will fail to deliver the promised business outcomes by the end of 2022. So, why are so many data analytics projects unsuccessful? Before answering this question, it is necessary to comprehend what data analytics entails.

Data analytics encompasses multiple dimensions. In the literature, the term has been used interchangeably with *business analytics*, *data science*, *data mining*, and *machine learning*, among others. In essence, the underlying goal is to extract insightful information and, consequently, actionable knowledge from a preprocessed set of data for better business outcomes (whichever these may be). In this article, we use data analytics to refer to all of these possible applications.

Data analytics is to be understood as a broad field involving data, tools, and processes that include both computational and management steps to make data-driven informed decisions. In other words, data analytics "is about extracting meaning from raw data using specialized computer systems and software that organize, transform, and model the data to draw conclusions and identify patterns"[4]. The practice of data analytics encompasses a number of processes that comprise a data science pipeline or workflow. Although closely related, *data analytics* should not be confused with *data analysis*. Data analysis is a subset of data analytics that refers to very specific actions performed on data. Data analysis "is the process of systematically applying statistical techniques to collect, describe, condense, illustrate, analyze, interpret and evaluate data"[4].

Talking about data analytics without mentioning 'big data' feels wrong. This is the era of (big) data. An era where data sets and problems are so large and cumbersome that traditional data processing and/or analytic methods can be inadequate. Big data pose a series of challenges which stem from their characteristics: they have many variables and/or many observations (volume), they are created and updated frequently (velocity), and they come in all different types (variety) and quality (veracity). As a result, performing data analytics on massive datasets becomes more challenging.

There are many types of analytics that organisations can perform. Analytics can be studied along three dimensions:[5] domain (*i.e.*, the subject field is which analytics are being applied, such as HR analytics, customer analytics, financial analytics, supply chain analytics, etc.), technique (*i.e.*, the way in which analytics are being performed; techniques can be technology-based or practice-based, qualitative or quantitative, etc.), and orientation (*i.e.*, the 'direction of thought', such as descriptive, predictive, or prescriptive analytics).

The orientation dimension is most used to derive analytics frameworks. Figure 1 presents a comprehensive orientation-driven framework, which involves five types of analytics: descriptive, diagnostic, predictive, prescriptive, and cognitive. There is no consensus over what this framework should look like. As a matter of fact, most orientation taxonomies consider only three or four types of analytics: descriptive, predictive, and prescriptive, and sometimes diagnostic, although the differentiation between descriptive and diagnostic analytics is not always made. Cognitive analytics, on the other hand, represents an emerging trend that promises to deliver additional value.

Unlike current representations that position the analytics alongside two dimensions of *complexity* and *value*, Figure 1 positions this framework in a three-dimensional view along three axes: *complexity*, *difficulty*, and *value*. It is relevant to note that the distinction between *complexity* and *difficulty* has not been properly clarified in the analytics literature, which has led to the two being interchangeably used. They are, however, quite different.

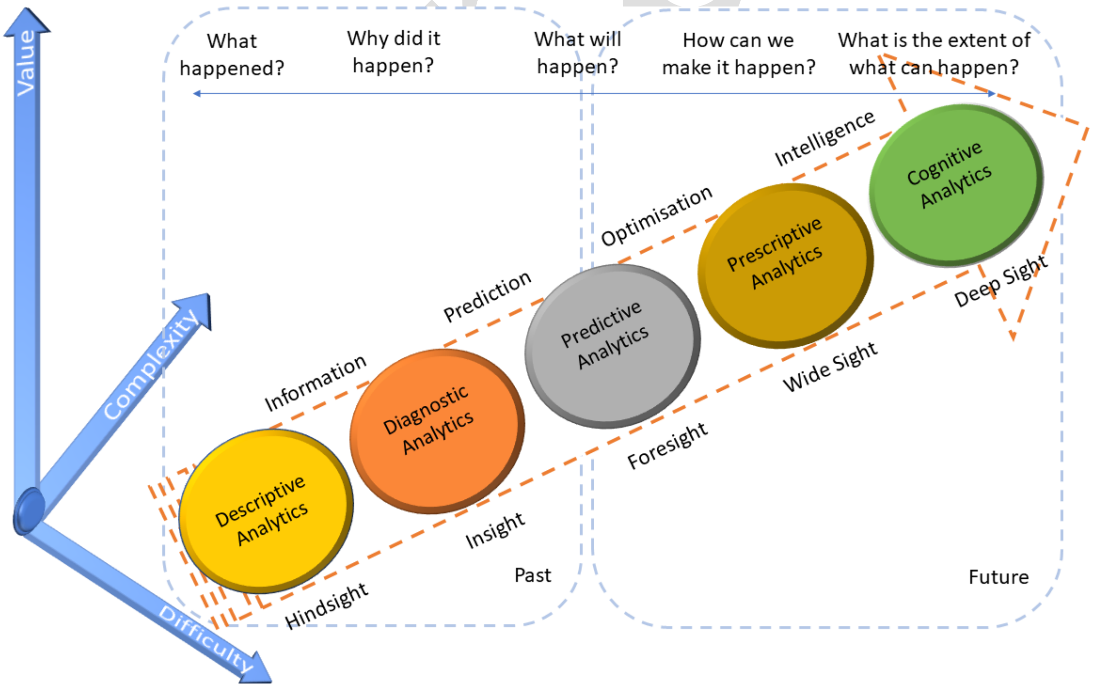

Figure 1. Data analytics framework[6].

Complexity is equal to computational complexity and is a measure of the amount of computing resources (with particular focus on time and memory requirements) that algorithms consume when they run. To be noted that complexity does not necessarily stem from large or big data. Complexity can also arise when dealing with data for which neither the number of variables nor the number of observations is necessarily large, but for which computational need increases rapidly (i.e., the procedure to solve the problem involves a computationally intensive algorithm that requires a relatively large number of steps to complete, thus taking more time and more memory).

Difficulty, on the other hand, stems from the difficulty of the problem being studied and the challenges associated with extracting valuable knowledge from the data in a meaningful manner. Difficult problems have many unknowns and are characterised by connectivity and mutual dependencies between involved variables, which means that, overall, these problems are very difficult to formulate. This also means that there is no established procedure for solving difficult problems, as it is very hard to adopt or develop a rule-based approach. The more difficult a problem is, the more likely it is that organisations will require both advanced technological capabilities and enhanced problem-formulation and problem-solving skills from their workforce.

Lastly, *value* means various superior outcomes, such as better decision-making, improved performance, realisation of organisational objectives, and enhanced competitiveness, among others.

Some additional observations should be made about Figure 1. First, although depicted sequentially, the types of analytics are not mutually exclusive; they are interconnected and can also work in parallel, depending on the problem to be addressed. Second, visualising analytics in this way reflects the belief that as analytics become more advanced, more complexity and difficulty should be introduced, necessitating more resources, both computer and human, and, in most cases, that this should be complemented by an increase in the value proposition offered. However, reality can be quite different. For example, there are times when there is insufficient evidence to justify implementing more sophisticated levels of analytics (like predictive, prescriptive, or cognitive).

Each type of analytics depicted in Figure 1 offers a unique value proposition (hindsight, insight, foresight, wide sight, deep sight) and serves a unique purpose (information, prediction, optimisation, or intelligence) (see Box 1). The techniques that can be used are numerous and span disciplines, including statistics (e.g., various types of regression, inferential statistics); machine learning (e.g., ensemble decision trees such as random forest), in particular kernel methods (e.g., support vector machines); operations research (e.g., linear, stochastic, and dynamic programming); and biology-inspired (e.g., genetic algorithms, neural networks, and nature-inspired algorithms such as swan particle algorithm and ant colony algorithm, and so on).

Organisations have been working on developing capabilities and adopting data analytics methods for extracting actionable knowledge from big data for what they have seen to be as better and more timely decision-making and enhanced competitive advantage. Indeed, the most common benefit associated with big data is data-driven decision-making. Or, as W. Edwards Deming said, “the ultimate purpose of collecting the data is to provide a basis for action or a recommendation”. A recent study[7], for example, showed that the adoption of a retail descriptive analytics dashboard by more than 1,500 e-commerce websites led to an increase of 4%–10% in average weekly revenues postadoption.

However, as previously stated, big data analytics initiatives fail quite often in practice[8]. It makes one wonder: why is having more data or sophisticated data analytics tools not a guarantor of success? There are many reasons why data analytics projects may fail. This is partially due to the fact that actions are not necessarily taken based on the gained insights, as a result of inertia/management style, resistance, a lack of leadership support, internal politics, and an unwillingness to change. But the answer may also lie in the lack of a problem-centric thinking approach.

Having mountains of data and powerful analytics tools simply is not enough. “If you do not know how to ask the right question”, W. Edwards Deming said, “you discover nothing”. It is important and relevant to first understand the variety of questions that can be answered with the different types of data analytics (see Box 1). Not only that, but it also becomes important to understand the context within which these types of analytics operate.

Independent of the application area, certain requirements should be considered while developing analytics models; these requirements include business relevance, as well as statistical performance (predictive capacity in terms of relevant performance measures), interpretability, explainability, justifiability, operational efficiency, and economic cost.

Notably, interpretability is frequently weighed against statistical performance, which is a critical trade-off to consider. For example, while neural networks may be more efficient in some cases and may provide better accuracy, they provide no information about the patterns in the data or how they work. A linear regression model, on the other hand, has limited modelling capacity but is quite intelligible and interpretable. In this sense, sophisticated models, colloquially referred to as black-box models, may be less reliable than simpler models. When much is unknown about a subject (like the COVID-19 epidemic), having models that can be explained, i.e., how they operate, how they predict the future, and so on, is more advantageous than having models that are difficult to understand. So, being able to specify the question will help in selecting the right analytics tools.

Another trade-off is between value and computational complexity. Dealing with higher computational complexity will not automatically result in a more important solution. There is a need to consider if the impact of the knowledge that is derived is substantial enough to make a difference or if it is just marginal. Large amounts of time, effort, and money can be spent on addressing computational complexity to get results, but if the enhancement in the value aspect turns out to be very low, it may not be worth the cost. Organisations need to ask themselves the exact value proposition attained from addressing higher computational complexity. Once again, understanding the problem that needs solving will help in understanding the level of added value desired.

Furthermore, it is not just about solving a difficult problem, but also about solving a difficult problem in time to allow for practical changes to happen. The time factor is important in generating added value. No analytics-driven solution will make any sense in real-life (no matter how difficult the problem) if a decision cannot be made at the right time.

Analytics work for a variety of different problems. So, analytics solutions need to be bespoke solutions that are designed to tackle exactly the problem they are supposed to tackle and get the correct insights. Otherwise, it is like throwing darts in the dark hoping to hit a specific target. It may or may not happen, but if it does, it is simply by chance. There is a need to implement analytics with problem-solving in mind, especially considering the high cost of customisation of bespoke analytics.

Turning data into meaningful knowledge is an art, and not a matter of adopting and applying analytics models blindly. All types of analytics are useful in driving decision-making for organisations, but one should carefully consider what is the level of computational complexity required and the level of problem difficulty involved against the value that can be generated. In the end, data analytics tools do not have to be too complicated or expensive. It all depends on what problem you want to solve and what the end goal is.

Box 1. The types of data analytics

The most popular and straightforward type of analytics is *descriptive analytics*. It tries to figure out what happened over a given length of time in order to answer the question "What happened?". As a result, its goal is to develop a comprehension of historical facts, as well as the meaning and character of past events; in other words, it offers *hindsight*. It focuses on presenting past data in a graphical format, commonly in the form of dashboards. Visual analytics is a common method for performing descriptive analytics.

Diagnostic analytics aims to identify the underlying causes of past events, answering the question "Why did it happen?". Data anomalies or outliers can be spotted using this tool, which then investigates the data sources that can explain the anomalies. If an event occurs again, this type of analytics can help the organisation to recognise patterns in data so that it is ready to respond quickly and avoid a negative outcome. Simply said, it provides *insight*, helping to achieve a better understanding of the world around us. It is also relatively easy to perform. Clustering algorithms, outlier detection, Naïve Bayes, and time-series data analytics are all common diagnostic analytics approaches.

Predictive analytics aims to provide a response to the question "What will happen?" or "What is most likely to happen?" As a result, it seeks to make predictions about the future based on the analysis of past events. Self-awareness or *foresight* is the ultimate goal. Despite the relative ease with which it can be implemented, this type of analytics can present considerable concerns for organisations. This is due to the fact that it necessitates additional technology as well as a more diverse set of talents; it is the quality of the data and the skill of the analyst that determines the accuracy of a prediction or forecast, since both are estimates at best. Ensemble decision trees–random forest, support vector machine, Long Short-Term Memory (for time series prediction), and other types of artificial neural networks–are common supervised machine learning techniques used in predictive analytics.

Prescriptive analytics is one of the most complex type of analytics that seeks to answer the question "How can we make it happen?". It assists organisations in achieving their goals by advising on all possible scenarios and activities that are likely to optimise such goals; it takes into account and builds upon the wider context to develop a *wide sight*. Prescriptive analytics often employs mathematical programming (such as linear, integer, and stochastic programming), game theory, Monte Carlo simulation, nature-inspired algorithms, and optimisation procedures. Prescriptive analytics can provide organisations with an extra competitive edge by improving decision-making in an optimised way, which other forms of analytics cannot. Needless to say, not all organisations can afford the time, effort, and resources (both computational and human) required to implement it.

Cognitive analytics, also referred to as "intelligent analytics", is the most sophisticated type of analytics, although it is still in its infancy. It answers the question "What is the extent of what can happen?", helping to develop a *deep sight* of the data to unfold hidden patterns. Its goal is to replicate human thought and mimic the way the human brain works. It works as a self-learning feedback loop, inferring from current data and then storing those inferences in the knowledge base for future inferences. In this sense, cognitive systems can adapt and become "smarter" over time by learning through their interactions with data and humans. Cognitive analytics is frequently performed using artificial intelligence, machine learning algorithms (more precisely, reinforcement learning), semantics, game theory, and deep learning models. Cognitive analytics combines standard analytics techniques with artificial intelligence and machine learning characteristics to enhance analytics results and provide a distinct competitive advantage.

It should be noted that, irrespective of the type of analytics, human input and expertise is always needed to ensure that the analytics are performed in an ethical, responsible, and safe way, and to ultimately decide on the practicality, validity, and deployability of the "knowledge" thus obtained.

Acknowledgements
The authors would like to thank the editor, former editor, and the anonymous reviewers for the helpful comments which substantially improved the clarity of this article.

Disclosure statement
The authors declare no competing interests.

References
1. Asay, M. (2017) 85% of big data projects fail, but your developers can help yours succeed. *TechRepublic*. Available at https://www.techrepublic.com/article/85-of-big-data-projects-fail-but-your-developers-can-help-yours-succeed/ (accessed 23 October 2022)
2. McShea, C., Oakley, D. and Mazzei, C. (2016) The Reason So Many Analytics Efforts Fall Short. *Harvard Business Review*. Available at https://hbr.org/2016/08/the-reason-so-many-analytics-efforts-fall-short (accessed 23 October 2022)
3. White, A. (2019) Our Top Data and Analytics Predicts for 2019. *Gartner*. Available at https://blogs.gartner.com/andrew_white/2019/01/03/our-top-data-and-analytics-predicts-for-2019/ (accessed 11 March 2022)
4. Sahay, A. (2018) Business Intelligence (BI), Business Analytics (BA), and Data Analysis. In A. Sahay, *Business Analytics: A Data-Driven Decision Making Approach for Business, Volume 1* (pp. 28–56). Business Expert Press. Available at https://hbsp.harvard.edu
5. Holsapple, C., Lee-Post, A. and Pakath, R. (2014) A unified foundation for business analytics. *Decision Support Systems*, **64**, 130–141. https://doi.org/10.1016/j.dss.2014.05.013
6. Charles, V., Emrouznejad, A. and Gherman, T. (2022) 'Strategy Formulation and Service Operations in the Big Data Age: The Essentialness of Technology, People, and Ethics', in Emrouznejad, A. and Charles, V. (eds) *Big Data and Blockchain for Service Operations Management*. Cham: Springer. https://doi.org/10.1007/978-3-030-87304-2_2
7. Berman, R. and Israeli, A. (2022) The Value of Descriptive Analytics: Evidence from Online Retailers. *Marketing Science*. https://doi.org/10.1287/mksc.2022.1352
8. Sankaran, A. (2021) Finding The Data: How To Avoid AI And Analytics Project Failures. *Forbes*. Available at https://www.forbes.com/sites/forbesbusinesscouncil/2021/10/15/finding-the-data-how-to-avoid-ai-and-analytics-project-failures/?sh=46591cdf5a66 (accessed 13 March 2022)