Chapter 26
Data Science in Economics and Business:
Roots and Applications

Mara Madaleno
https://orcid.org/0000-0002-4905-2771
GOVCOPP, DEGEIT, Portugal & University of Aveiro, Portugal

João Lourenço Marques
https://orcid.org/0000-0003-0472-2767
GOVCOPP, DCSPT, Portugal & University of Aveiro, Portugal

Muhammad Tufail
https://orcid.org/0000-0002-6013-9157
National University of Sciences and Technology (NUST), Pakistan

ABSTRACT

Economics and business are a great background for data science provided econometricians and data scientists are sets with an intersection, although remaining unknown. In econometrics, data mining is somewhat a monstrous word, a field that traditionally seeks causal inference and results in interpretability. When we go deeper into what data science usually is, the boundaries between more traditional econometrics and even statistics and the hip and cool machine learning become shorter. In economics and business, we find examples and applications of simple and advanced data science techniques. This chapter intends to provide state-of-the-art data science applications in economics and business. The review and bibliometric analysis are limited to the research articles published through Elsevier Scopus. Results allowed the authors to conclude that despite the number of already existent research, a lot more remains to be explored joining both fields of knowledge, data since, and economics and business. This analysis allowed the authors to identify further possible avenues of research critically.

DOI: 10.4018/978-1-7998-6985-6.ch026

INTRODUCTION AND FRAMEWORK

Undoubtedly, mathematical modeling and statistics are central to quantitative economics, highlighted by the fact that some of the biggest data repositories are maintained by economic research organizations. Economists try to understand human behavior through Homo economics or the "economic man", to model human behavior. For that, they rely on economic theories and use some analytical tools and techniques, many of which are no more than standard statistical and mathematical models. We find in the literature economists walking around in all domains; from unemployment and inequality to the economics of climate change, to advertisement and revenue collection, from consumer and production behavior, up to profit, sales, and purchases, and finance data explorations (Provost and Fawcett, 2013; Brooks et al., 2019; Nosratabadi et al., 2020a, 2020b; Taddy, 2019; Basdas and Esen, 2020; Consoli et al., 2021).

We may think of some reasons why we should not dissociate economics, business, and data scientists. At least economists know exactly what the term machine learning means but with more econometric and statistical concepts. Machine learning ends up being a fancy word to describe statistical or predictive modeling that is used by programmers (Cao, 2017). Surprisingly, when we look for the most popular machine learning courses, the first two modules of their syllabus are still linear regression and logistic regression, basic concepts learned by economists in introductory econometrics. Econometrics is the application of statistical modeling in understanding complex social and environmental issues, being a big area of applied economics. We may even argue that in this case, economists have a deeper knowledge of linear regression than the average data scientists abroad. However, it is still hard for an economist to understand concepts such as neural networks, machine learning, and support vector machines, among many others, and still, heteroscedasticity is something away from the syllabus of machine learning, putting forward the pros and cons for economists or managers desiring to be involved in data science (Basdas and Esen, 2020; Consoli et al., 2021). Neural networks combine layers of logistic-like regressions to model non-linear relationships among variables, usually more complex, that simple regression analysis is not able to capture. Therefore, nothing as difficult as it may initially seem. Still, they build powerful algorithms, in what we believe to be complementary areas.

Economists have higher standards given their obsession in finding causal relationships, a goal unable to be fulfilled unless randomized controlled trials are pursued and sensitivity analysis to models basic assumptions scrutinized (we may think here of many emerging biases like attenuation, survivorship, selection bias, measurement error, reverse causality, truncation, censoring, among others). For machine learning, things are easier, given that models are not solved explicitly, and they do not need to rely on stricter assumptions (Donoho, 2017). Instead, models are estimated interactively with gradient descent and its derivatives, usually ignoring the theory behind the relationship we are trying to study (Taddy, 2019; Nosratabadi et al., 2020b). Even so, they use cross-validation and testing, and instead of t-statistics bootstrapping. In reality, this is summed up to the use of the right tools in the correct applications. Moreover, presentation and writing are important parts of data science provided non-technical audiences will be the interesting parts in research (managers, policymakers, marketers, copywriters, customers, clients). After all, the final result expected to be driven from the research is that we can demonstrate why our results matter and how stakeholders can use them and act over them. Economists, have generally a broader picture of the overall panorama and deepen these results, not just reading numbers as most data scientists, but presenting and clearly explaining results, with fluid writing. Whereas mathematicians and a computer scientist are not so comfortable when it comes to presenting and explaining the work