Chapter 6
Big Data Is Decision Science:
The Case of COVID-19 Vaccination

Jacques R. J. Bughin
Université libre de Bruxelles, Belgium

Michele Cincera
Université libre de Bruxelles, Belgium

Dorota Reykowska
NEUROHM, Poland

Rafał Ohme
WSB University in Torun, Poland

ABSTRACT

Data science has been proven to be an important asset to support better decision making in a variety of settings, whether it is for a scientist to better predict climate change for a company to better predict sales or for a government to anticipate voting preferences. In this research, the authors leverage random forest (RF) as one of the most effective machine learning techniques using big data to predict vaccine intent in five European countries. The findings support the idea that outside of vaccine features, building adequate perception of the risk of contamination, and securing institutional and peer trust are key nudges to convert skeptics to get vaccinated against COVID-19. What machine learning techniques further add beyond traditional regression techniques is some extra granularity in factors affecting vaccine preferences (twice more factors than logistic regression). Other factors that emerge as predictors of vaccine intent are compliance appetite with non-pharmaceutical protective measures as well as perception of the crisis duration.

INTRODUCTION

Big data is the handling of vast amounts of data through a flexible, mostly cloud-based IT architecture. Big data is here to stay for many reasons. The first is that digitization is data extension, with new digital

DOI: 10.4018/978-1-7998-6985-6.ch006

content creation growing at a rate between 40 and 60% a year. This rate means that Google, which was indexing a million pages for a few million searches in 1998, now indexing more than a trillion pages ten years later, for1.2 trillion searches a year. Second, big data can fuel new powerful machine learning techniques to uncover otherwise hidden relationships between data, that can support new powerful insights. As an example, Netflix shifted its content recommendation engine, based on customer rating, to a machine-learning algorithm fed by a large set of big data ten years ago. The tool offered a revolution to support personalized recommendations, in such a way, that now, four out of five Netflix movies and TV series viewed by Netflix subscribers originate from machine-based suggestions (Amatriain, 2013).

The seminal work by Brynjolfsson, et al. (2011) reveal that companies leveraging big data for more fine-tuned decision making, could increase their revenue productivity by more than 5%. A few years later and using a more global sample, Bughin (2016) had reached essentially the same conclusion- that is big data is usually responsible for an uplift of more than 5% in labor productivity for firms globally. Besides being used for business, big data is also important when it comes to helping major social issues, in particular when it comes to complex issues such as predictions of climate change (Bauer et al., 2021), vote intent (Mavragani and Tsagarakis, 2019), or traffic congestion (Teseng and colleagues 2018).

One another current case in point is the vaccination issue linked to the covid-19 pandemic. While the covid-19 pandemic continues unabated across the world, affecting hundreds of million people worldwide, the only serious way to eradicate the pandemic is mass- vaccination. However, large vaccination acceptance is not warranted. Social movements against vaccination have increased during the covid-19 pandemic, fueled by a series of conspiracy theories (Nguyen and Catalan, 2020). There are also possibly multiple, and interacting elements that affect vaccine intent, which make predictions of vaccine used complex—A recent poll performed by the high profile organization Pew Research in the US also emphasizes that the shape of vaccine intent is driven by « rather complex and interrelated factors » that might require more deeper data analytics to sort out a clear ranking of actions to support mass-vaccination[1].

We thus leverage machine learning to profile vaccination intent for covid-19, concentrating outside of known product factors (such as the effectiveness and safeness of the vaccine) to determine additional factors that could be boosted or alleviated to enhance fast vaccination against the Covid-19. Specifically, after aligning with recent theories of vaccine determinants to build an extensive database of behaviors and vaccine intent across 5 Continental European countries—Germany, France, Spain, Italy and Sweden, we use Random Forest (RF) techniques to provide a granular view of vaccine intent. RF is one of many machine learning techniques, but is found to generate rather accurate predictions regarding protective uptake (Bughin and Cincera, 2021)

Reaching for advanced analysis techniques has been complemented by a dedicated approach to data collection. Predicating behavior in a sensitive context such as vaccination intent in the face of the covid-19 pandemic required the usage of tools that would help decrease the natural tendency to distort declarations.

Understanding the real motives of behavior has always been a challenge for researchers. Numerous studies have shown that there is a weak correlation between declarations and behavior (Krauss, 1995) making it hard to predict behavior based solely on explicit answers. Especially when testing attitudes around sensitive topics, full of emotional load. Political correctness, post rationalizations or auto presentation needs are all important aspects that frequently influence and distort explicit, declarative answers. Additionally in the last 50 years researchers have consistently shown that most of our cognitive processes take place outside of conscious awareness and control and still they influence our perceptions, judgments and actions (Zajonc, 1968, 1980; Uleman & Bargh 1989; Bornstein & Pittman 1992; Greenwald 1992; Murphy & Zajonc 1993; Bargh 1997; Ohme, 2001; Nosek, Hawkins, & Frazier, 2011). For this reason