Editorial

Can Data Science Help Fight Corruption in Public Procurement?

General overview

In recent years, there has been a growing interest in the adoption and implementation of new technologies associated with the fourth industrial revolution, or Industry 4.0. Concepts such as Big Data (BD) and Artificial Intelligence (AI), with their advantages and countless practical uses, have become drivers of transformation in any organization. We often hear about medical diagnoses based on artificial neural networks; the selection of innovation factors through machine learning models; AI-powered chatbots for marketing channels; and prediction, control, and speed patterns in urban public transport, among others. The list of these novel technologies and future scenarios for their application is seemingly endless.

In this regard, computer science has provided integrative concepts that have allowed systems to learn behaviors through specific algorithms and by processing large volumes of data. The primary basis of AI is, for instance, data generation and sorting, also known as BD—its backbone. The aim of data management is to receive and process new pieces of information, in addition to quickly and

efficiently forecasting and developing repetitive human tasks. This is precisely one of the main paradigms of the fourth industrial revolution because technological development is what is driving the new industrial era in which humankind is already immersed.

However, the way non-experts use the concepts and potentials of these new technologies often leads to errors and confusion. They, for instance, are employed interchangeably, even as if they were synonyms, irrespective of the fact that there may be substantial differences between them to serve their own purpose and a specific application. It is, therefore, not surprising that, given the wide variety of concepts, many can hardly achieve a very limited conceptual understanding of this entire field. For example, data, a concept whose flat and simple definition would indicate that it is something of important value to people, organizations, or systems, has triggered an avalanche of ethical and legal discussions on how it should be handled.

Yet, even more complex concepts, such as data science, are considered an entry point towards more complex tools. In data science, hypotheses about the real world are developed and then validated by analyzing the collected data. Then, there is also machine learning, a concept that employs algorithms to help a computer system to acquire knowledge through the use of data. This concept is, in turn, complementary to that of data science, which is why it is fair to say that machine learning can be, and is often, part of an adequate data science. This latter is fundamentally a process, whereas machine learning is a tool that can be quite useful to carry out such process. Finally, across the entire network of concepts, we find AI, whose application has focused on making machines emulate human intelligence.

Big problems require big solutions

The use of Industry 4.0 technologies has evolved to find new and potential niches, some of which, unfortunately, have seen recurrent difficulties and controversies and become a source of concern for our society. Some of those problems include corruption and fraud, which, although not new, have found a way to persist due to people's misuse and lack of knowledge of technologies.

Public corruption is defined as the abuse of public power or position for material gain. The different forms of corruption, which are driven by several factors, have caused multiple negative effects. In particular, corruption hinders economic growth, thus affecting businesses, jobs, and investments, which results in a poorer quality of life and a declining trust in the government, as it deprives us all of opportunities. Broadly speaking, corruption is a systemic, complex, and adaptive phenomenon that requires to be addressed from various disciplines in order to be properly detected and fought.

One of the most recent cases of corruption occurred during the crisis caused by the COVID-19 outbreak. Such a crisis has not only affected people's well-being but also health systems, as unscrupulous individuals have taken advantage of the situation to steal public money through the governments' digital procurement platforms. As is well known, an inefficient public procurement management often makes it vulnerable to corruption. This is mainly due to factors such as rigid structures, lack of coordination between the government and institutions, poor time and money management, and limited or non-existent monitoring.

Fortunately, this whole situation has also led to considering creating mechanisms to react to, detect, raise red flags, and prevent those actions that are presumably fraudulent or unethical and stop them from being repeated. Now, the question that naturally arises is: How can new computer technologies be effectively and efficiently used to categorically prevent fraud and corruption in public procurement processes?

Science, technology, and innovation as the solution

Albert Einstein wisely coined the phrase "Insanity is doing the same thing over and over again and expecting to get different results." This phrase should encourage us to choose the path of change, of divergent thinking, which leads us to unexplored pragmatic approaches, that is, to innovation ideas applied to problems that have traditionally been tackled by means of rigid or traditionalist methodologies.

In this context, using data science and analysis as drivers of innovation to transform public procurement processes and devise mechanisms to combat public corruption has turned out to be not only a feasible but also profitable solution from

multiple perspectives. Some of the solutions that have been addressed in different studies include exploratory analyses of data in public procurement contracts for the provision of health services, the use of natural language processing to analyze vaccination contracts during pandemics, the detection of fraud patterns in public contracts by means of unsupervised learning and clustering algorithms.

In general, the growing attention the public procurement sector is receiving and the various innovative efforts have allowed scholars, researchers, professionals in various areas (such as technology managers), and the private sector to converge towards new forms of innovation and work together for a common goal using open data tools such as the Public Procurement Electronic System (abbreviated SECOP in Spanish), a platform employed by Colombia's government to report on all the contracts executed with public funds. Other methodologies include project management, Agile, Scrum, and CRISP-DM.

As a result, novel and improved applications, many of which are still in the proposal, development, or improvement stage, have been increasingly developed to provide a solution to this serious social problem.

Future challenges

The future calls for proposals to meet the coming challenges. Detecting corruption in public procurement is not only an academics' and businesses' issue but a society's commitment that involves all sectors and citizens. Therefore, to tackle such problem, innovative approaches must be adopted, and governments must support society in their implementation so that the improvement strategies are in line with the needs of citizens. In addition, efforts should be made to ensure that control frameworks turn individuals and social organizations into allies to develop data science proposals for the identification of risks or possible acts of corruption, which, ultimately, would strengthen co-responsibility between the public and private sectors.

Julián Alberto Uribe-Gómez
julianuribe@itm.edu.co
MSc in technological management-data scientist,
Professor at the Faculty of Economic and Administrative Sciences,
Instituto Tecnológico Metropolitano - ITM, Medellín, Colombia