Big data analytics in the health sector: challenges and potentials

DOI: 10.7595/management.fon.2019.0001

Abstract:

Research Question: The introduction of the Big Data concept in the healthcare sector points to a major challenge and potential. Motivation: Our goal is to indicate the importance of analyzing and processing large amounts of data that go beyond the typical ways of storing and processing information. The data have their own characteristics: volume, velocity and variety. There are different structures. Analysis of these data is possible with the Big Data concept. Its importance is most evident in the health sector, because the preservation of the health status of the population depends on adequate data analysis. Idea: The idea of the paper is that big health data analytics contributes to a better quality provision of health services. The process is more efficient and effective. Data: Health analytics suggests that more and more resources are being utilized globally. In order to achieve improvements, health analytics and Big data concepts play a vital role in overcoming the obstacles, working more efficiently and aiming at providing adequate medical care. Tools: The Big data concept will help identify patients with developed chronic diseases. Big data can identify outbreaks of flu or other epidemics in real time. In this way, they are managed by the healthcare system, reducing overall healthcare costs over time, and increasing revenues. Findings: A key policy challenge is to improve the outcomes of the healthcare system, data collection and analysis, security, storage and transfers. Big data are the potential to improve quality of care, improve predictions of diseases, improve the treatment methods, reduce costs. Contribution: This paper points to the challenges and potentials of Big Health Data analytics and formulates good reasons to apply the Big Data concept in healthcare.

Keywords: healthcare, health analytics, big health data analytics, challenges and potentials, quality and efficiency of care.

JEL Classification: I15, M15, P47

1. Introduction

In modern business, organizations are increasingly introducing the Big Data concept globally. Several definitions of Big Data have been put forth to date. One of them is that Big Data involves data storage, management, analysis, and visualization of very large and complex datasets (Russom, 2011).

The term Big Data cannot be identified only with a large amount of diverse data, but rather with the techniques used in the processing of these data and ways of making important business decisions (Chen, Chiang & Storey, 2012; Martin-Sanchez & Verspoor, 2014). Due to these characteristics the more often used term is Big Data analytics or “analytics of large amounts of data” because its purpose is the application of advanced technology and statistical techniques for finding hidden patterns in data and generate information for making important business decisions (Mikalef, Pappas, Krogstie, & Giannakos, 2017; Chawla & Davis, 2013).

The Big health data analytics often has a predictive purpose: in a way, the use of relevant predictive analytical tools makes it possible to find the cure before people get affected by a specific disease (Khennou, Khamlichi, & Chaoui, 2018). The health system involves large amounts of data - records of patients, management of data, analysis... (Choong & Hyung, 2017). By definition, Big Data in the healthcare system relates to electronic health data, which are so large and complex that it is difficult (or impossible) to manage data on the basis of traditional software and/or hardware (Raghupathi & Raghupathi, 2014; Groves, Kayyali, Knott & Van Kuiken, 2013).
Big Data in the health sector bring challenges and potentials. Many authors have written on this topic, so this paper gives an overview. The study section 2 will discuss health analytics. Then in section 3 Big Health Data analytics will be presented. After that, section 4 will present challenges and potentials for Big Data in health care. At the end of the paper, the conclusion is given along with the recommendations and suggestions.

2. Aspects of Health Analytics

Health analytics is a tool that is used to make decisions about health care based on health data. While such aspects of health analysts, such as the use of statistical models, data mining and support for clinical decisions, existed for decades, the availability of large amounts of data contributed to an integrated process of decision-making. Aspects of healthcare analytics are:

- Provide a combination of financial and administrative data,
- Identify areas to simplify operations and thus reduce costs,
- Research and development - monitor, evaluate and analyze new solutions,
- Clinical data provide information on the effectiveness of treatment,
- Doctor's understanding of what patients feel and how they react to treatment contributes to the expansion of services,
- Optimize efforts in managing hospital and foundation donations and grants,
- Allows hospitals to track physician records, patient histories, and needs to ensure the right doctor is deployed to the patients most in need,
- Provide an easier approach to track existing claims, clients, and premiums for insurance companies.

Healthcare industry worldwide has recently been faced with various challenges such as a high cost of providing health services, a high percentage of aging population, patients with chronic diseases, a significant shortage of medical experts... (Krumholz, 2014; Bates, Saria, Ohno-Machado, Shah & Escobar, 2014).

For example, the cost of health care spending amounted to $2.6 trillion in 2012, in the US, an estimated 17.6% of the gross domestic product (GDP), while in 2016 it amounted to about $4.1 trillion. In the report, IMS Research shows that the costs for pharmacy drugs increased to $1.2 trillion in 2016 (IMS Institute, 2012; Chen, Mao, & Liu, 2014). Figure 1 shows a summary of the total costs by region and by year. Total operating expenses in 2006 amounted to $658 billion, the sum in 2011 amounted to $956 billion and in 2016 there was an increase of 1.175-1.205 billion dollars. The US has participated in costs from 41% in 2006, followed by 34%, and then 31% in 2016. The states of the European Union participated in the costs with 19% in 2006, followed by 17%, while in 2016 they covered 13% of the total costs (IMS Institute, 2012).

Figure 1: Cost of drugs by years and regions (adapted from IMS Institute, 2012)
Further analysis shows that the health industry is large, but also critical. Constant investment in the health sector leads to inefficiency. According to the Institute of Medicine, one-third of health spending (about 750 billion dollars) today is lost and does not contribute to improving the health care of patients (IMS Institute, 2012; Meeker & Hong, 2014). This refers to the various inadequate services, administrative work, unproductive work processes, inefficient service delivery, the high price of health care services, deception and missed opportunities for providing prevention (Alyass, Turcotte & Meyre, 2015; Kim & Groeneveld, 2017).

Costs relating to the inefficient delivery of health services are (Groves, Kayyali, Knott & Van Kuiken, 2013) as follows:

- 130 billion dollars for an inefficient provision of health services,
- 105 billion dollars for prices that are too high,
- 190 billion dollars for excess administrative costs,
- 210 billion dollars for the unnecessary service,
- 75 billion dollars for fraud,
- 55 billion dollars for missed prevention opportunities.

The Big Data concept is as potential way to go in the quest for better efficiency of the healthcare sector (Choong & Hyung-Jin, 2017; Sanskruti & Atul, 2016). However, the concept also contributes to better decision-making in relation to the patient’s health status, accurate diagnosis, effective treatment and better management of health data.

3. Big Health Data Analytics

Big data related to healthcare have rapidly expanded with the development of the Internet. Management and analysis of voluminous and varying Big data is possible using the Information-communication technology (Ryu & Song, 2014). Such data include basic patient data, various analyses, laboratory reports, medical images, a register of doctors and nurses.

Many healthcare institutions from different countries have successfully applied solutions of Big Data in traditional healthcare. The processes that occur are improving the quality of care, increasing the effectiveness and efficiency of healthcare (Ryu & Song, 2014; Moon-Koo & Jong-Hyun, 2016; Witjas-Paalberends, Laarhoven, Burgwal, Feilzer, Swart, Claassen & Jansen 2017). The effectiveness and efficiency of healthcare is the application of big data analytics to improve healthcare: 1) predictive modelling for risk and resource use; 2) population management; 3) drug and medical device safety surveillance; 4) disease and treatment heterogeneity; 5) precision medicine and clinical decision support; 6) quality of care and performance measurement; 7) public health; and 8) research applications (Rumsfeld, Joynt & Maddox, 2016).

The concept of Big data in the health care system is characterized by several features: the volume, the variety and the velocity (Ganjir, Sarkar & Kumar, 2016; & , 2013). According to data from McKinsey Global Institute, the amount of data generated globally is expected to grow at an average annual rate of 41%. In the period from 2008 to 2020, this amount will increase by 44 times, from 0.8 zettabytes in 2008 to 35 zettabytes in 2020 (Nasscom and Crisil Global Research & Analytics, 2016). This is supported by the fact that every day on Facebook 10 terabytes of data and on Twitter 7 terabytes of data are generated (Groves, Kayyali, Knott & Van Kuiken, 2013). The highest volume of health data is collected in North America (Nasscom and Crisil Global Research & Analytics, 2016).

New data are generated at high speed, which is far greater than the speed of data processing. Health data by various sources can be collected by monitoring the health status of individuals or of the entire population (Kyoungyoung & Gang, 2013). However, different stakeholders have different goals and hopes for Big Data analytics (Jovanovic Milenkovic, Milenkovic, Vukmirovic & Radojicic, 2016):

- Physicians – to improve better medical care;
- Patients - want daily use of new technologies in order to get better care with the proper diagnosis;
- Providers - want access to data of the clinic, the patients, in order to improve decision making, achieving efficiency in the work without error;
- Researchers - want a new tool to improve the quality and quantity of the workflow. These include the various statistical tools and algorithms;
- Pharmaceutical companies - they want to better understand the causes of disease and thus get a safer drug market;
Mobile devices producers – data provided by the various mobile health applications integrated into the mobile devices (heart beep, physical activity, sports and recreation, sleeping habits etc.)

- Government - trying to reduce costs, enforce regulations;
- Company’s software development - see the opportunities to serve the wholesale market, to develop software that will find application in the health system.

The US health care system faces the problems of the health insurance. Resolving these problems has been described by the former US President Barack Obama, who states that The Affordable Care Act has made a significant progress towards solving long-standing challenges related to access, affordability and quality of care (Barack, 2016). In March 2012, the former US President Barack Obama and his administration launched an initiative to approve $200 million for the project “Big Data Research and Development Initiative” whose main objective is to transform the use of Big Data for scientific discovery and biomedical research, with the participation of several federal departments (Kyoungyoung & Kim, 2013; Weiss & Zgorski, 2012; Chen, Mao & Liu, 2014). The Obama administration proposed a program “Health 2.0” for efficient data management: patients, medical institutions, health insurance. One of the proposed models in the “Health 2.0” is a Pillbox. Pillbox involves providing accurate information about specific drugs that the patient is interested in. The aim of Pillbox is to reduce costs in the sale of drugs to keep medical records in order to avoid inadequate selling of drugs. It is believed that the service is necessary for the elderly (Sagiroglu & Sinanc, 2013; Kyoungyoung & Kim, 2013). A report by McKinsey Global Institute suggests that if the US healthcare were to use Big Data creatively and effectively, the sector could create more than $300 billion in value every year. Two-thirds of the value would be in the form of reducing the US healthcare expenditure (Manyika Chui, Brown et al., 2011; Belle, Thiagarajan, Soroushmehr, Navidi, Beard & Najarian, 2015)

By applying the Big Data analytics, physicians can make timely decisions. The Big Data concept will help identify patients with developed chronic diseases, such as diabetes, heart failure or chronic obstructive pulmonary disease (Fong, Ng & Yuen, 2017). In this way, the data are connected, identifying patients who need additional care (Nasser & Tariq, 2015). An early prevention of diseases start with a change of the lifestyle. Management of these diseases is significant because they are the most costly to treat. Big data can be tracked by software and thus identify outbreaks of flu or other epidemics in real time. In this way, they are managed by the health system, reducing overall healthcare costs over time, and increasing revenues (Murdoch & Detsky, 2013).

4. Challenges and Potentials

A key policy challenge in both developing and developed countries is to improve outcomes of the healthcare system while containing cost pressures. Furthermore, health care costs are rising rapidly, driven by population ageing and rising relative prices. To improve public healthcare spending projections and meet a rapidly growing health care demand, efficiency gains in health care will be crucial (Mitrovic, Vujosevic & Savic, 2016).

The potential of the Big Data concept can help healthcare organizations improve quality and efficiency in the following (Luna, Mayan, Garcia, Almerares & Househ, 2014):

- It leads to new knowledge, based on the analysis of data, that can provide solutions to some of the problems that cannot otherwise be detected in clinical conditions;
- disseminating knowledge, helping physicians through clinical decision support systems, which can provide suggestions and predictions in real time based on the personal data of patients;
- translating personalized medicine initiatives into clinical practice by providing the opportunity to use analytical capabilities;
- Empowering patients - patients can be informed of their health and healthcare;
- Improving epidemiological surveillance: some tools are developed and keep track of the highly prevalent or deadly diseases in the population.

Big data have the potential to improve patient care, save lives and reduce healthcare costs (Srinivasan &, 2013; Koumpourosm, 2014). The potential health lies in combining traditional details with new forms of data at the individual level and at the population level. In Table 1, the differences between traditional analytics and Big Data analytics are shown in relation to the data sources, processing techniques, intensity, time and purpose of data processing.
Table 1: Traditional vs. Big Data analytics (Lazarevic, 2015)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Traditional analytics</th>
<th>Big Data analytics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data capture</td>
<td>Analysis of the data part (Partial Data)</td>
<td>Analysis of all the data (All Data)</td>
</tr>
<tr>
<td>Sources</td>
<td>A small number of homogeneous data sources</td>
<td>A large number of heterogeneous data sources</td>
</tr>
<tr>
<td>Data processing techniques</td>
<td>Standard tools and techniques (classic software, classical statistical techniques)</td>
<td>Advanced tools and techniques (advanced software, advanced statistical techniques)</td>
</tr>
<tr>
<td>Time data processing</td>
<td>After generating the data (Ex-Post)</td>
<td>During the generating of the data (Real-Time)</td>
</tr>
<tr>
<td>The intensity of data processing</td>
<td>Occasional data processing (Ad-Hoc)</td>
<td>Continuous data processing</td>
</tr>
<tr>
<td>The goals of data processing</td>
<td>Diagnosis and description of the event</td>
<td>Behavioural analysis and forecasting of events</td>
</tr>
<tr>
<td>Types of reports</td>
<td>Standard uniform reports</td>
<td>Complex reports</td>
</tr>
</tbody>
</table>

Table 1 shows that Big Data analytics reveals many advantages. Big data in the health sector mainly consist of huge amounts of digital health data such as electronic health records (EHR), computerized physician order entry (CPOE), picture archiving communications system (PACS), clinical decision support systems (CDSS), and laboratory information systems.

The following factors lead to the improvement of quality of healthcare (Archenaa & Anita, 2015):

- **Providing patient centric services:**
  This involves detecting diseases at the earlier stages based on the clinical data available, minimizing drug doses to avoid side effects, and reducing cost for the patients.

- **Detecting spreading diseases earlier:**
  Predicting the viral diseases earlier before their spreading, based on the live analysis. Preventive measures shall be taken in a particular geo-location.

- **Monitoring the hospital’s quality:**
  All hospitals must apply the norms issued by the medical council. This periodical monitoring indicates necessary measures against disqualifying hospitals.

- **Improving the treatment methods:**
  Customized patient treatment - Doctors monitor patients suffering from certain diseases. Based on the analysis of data, doctors can give the persons who may suffer from the same symptoms, the same drugs that lead to healing.

In the literature, many authors gave their views on the challenges and potentials provided by Big Data. Electronic health records are designed and implemented to move data between disparate organizations, provide a follow up of patients, and develop strategies to improve overall outcomes. Cloud storage is increasingly used, because costs drop and reliability grows. A large number of healthcare organizations use some sort of cloud-based health IT infrastructure, including storage and applications. The Security Rule includes a long list of technical safeguards for organizations storing protected health information, including transmission security, authentication protocols, and controls over access, integrity, and auditing (Alekxandru, Radu & Bizon, 2018). Table 2 lists the main challenges and potentials for big data in health care.

Table 2: The challenges and potentials of health Big data analytics

<table>
<thead>
<tr>
<th>CHALLENGES</th>
<th>POTENTIALS</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Data Structure</td>
<td>• Improve Quality of Care</td>
</tr>
<tr>
<td>Data are unstructured and in large volumes.</td>
<td>By applying the Big Data Concept, the quality and efficiency of care is improved. Analytics can indicate the number of chronic diseases and epidemics.</td>
</tr>
<tr>
<td>• Storage and Transfers</td>
<td>• Prediction</td>
</tr>
<tr>
<td>High data costs occur when data are transferred and stored.</td>
<td>Big Data allow for the prediction of diseases. Big Data analytics can also help in the prevention of deadly illnesses and personalized monitoring and disease management.</td>
</tr>
<tr>
<td>• Security</td>
<td>• Inadequate knowledge</td>
</tr>
<tr>
<td>There is concern for data confidentiality.</td>
<td>Individuals need to be further educated with the development of technologies.</td>
</tr>
</tbody>
</table>
Management challenges related to Big data are a group of challenges encountered, for example while accessing, managing and governing the data. Data warehouses store massive amounts of sensitive data such as financial transactions, medical procedures, insurance claims, diagnosis codes, personal data, etc (Sivarajah, Kamal, Irani & Weerakkody, 2017). The health information scientists and computer scientists discussed big data in medicine. They focused on how to collect, store, integrate, and manage data. However, the future of big data in health sector is in using new analytic techniques such as machine learning to answer clinical questions, educating doctors and policy makers to understand big data, and promoting the use of tools generated by big data and big data technologies that support clinical decision making. In the future, it is important to discover the challenges and potentials that health Big Data analytics provides (Zhang, Wang, Zhan & Zhan, 2018).

Big Data technologies have the potential to transform the way healthcare providers use sophisticated technologies to gain insight from their clinical and other data repositories and make decisions based on collected information. In the future, we will see rapid, widespread implementation and use of Big Data analytics across the healthcare organization and the healthcare industry. Big Data analytics and applications in healthcare are at a nascent stage of development, but rapid advances in platforms and tools can accelerate their maturing process (Vukmirovic, Rajnai, Radojcic, Vukmirovic & Jovanovic Milenkovic, 2018).

The goal of the health system is a healthy population (Jovanovic Milenkovic, Radojcic, Milenkovic & Vukmirovic, 2009; Jovanovic Milenkovic, Jeremic & Martic, 2014). Due to the expansion of information and communication technologies in all spheres of life, especially in the field of health care, the importance of theories that predict and explain the factors that influence the adoption and use of information and communication technologies in health care is growing (Rodic Trmcic, Labus, Bogdanovic, Babic & Dacic-Pilcevic, 2016).

The concept of “Big data” is not new; however, the way it is defined is constantly changing. The introduction of Big Data concepts in the health system contributes to better health care characterized by speed, accuracy and timeliness. Health Analytics and Big Data concepts play an important role in overcoming obstacles in order to a more efficient and faster medical care. The application of new technologies is essential. Medical and health technology is used by everyone: patients, physicians, health workers, end-users, engineers, clinics and hospitals.

In this paper, our systematic literature review revealed both challenges and opportunities that big data offers to the health care industry. The literature listed the challenges of data structure, storage and transfers security and standardization. The literature also listed the opportunities of increased quality, better management of population health, prediction of diseases, cost reduction and globalization.
Despite the complexities of healthcare data, there is a potential and benefit in developing and implementing Big Data solutions within this realm. These findings point to the direction for future research. Ultimately, the key to success will be to remain focused on ultimate goal: gaining an effective insight into the best ways to treat patients in the healthcare system (Schneeweiss, 2014).

REFERENCES

About the Authors

Marina Jovanović Milenković
University of Belgrade, Faculty of Organizational Sciences, Serbia
e-mail: marinaj@fon.bg.ac.rs

Marina Jovanović Milenković is an Associate for Ph.D. studies within the Department of Ph.D. studies at the Faculty of Organizational Sciences, University of Belgrade. She completed her PhD thesis in the field of Information and communication technology in health care management in 2011. Her major interests are e-health, ICT implementations in health systems and business decision making.

Aleksandra Vukmirović
Higher Education School for Applied Studies Belgrade Business School, Serbia
e-mail: aleksandra.vukmirovic@bbs.edu.rs

Aleksandra Vukmirović is a lecturer at the Belgrade Business School, Higher Education Institution for Applied Studies. She is the author of several scientific studies of national and international importance. She participated in and managed a number of projects in the fields of e-business and marketing research.

Dejan Milenković
Serbian Armed Forces General Staff, Serbia
e-mail: dejan.milenkovic@vs.rs

Dejan Milenković is a professional officer in the Serbian Armed Forces - General Staff. He completed his PhD thesis in 2013 in the fields of statistical management, electronic documents management and repositories of electronic documents. His major interests are e-business, documents management system, business process analysis and business decision making.