Chapter 4
Big Data and Analytics: Application to Healthcare Industry

Misbahul Haque
Aligarh Muslim University, India

Mohd Imran
Aligarh Muslim University, India

Mohd Vasim Ahamad
Aligarh Muslim University, India

Mohd Shoaib
Aligarh Muslim University, India

ABSTRACT

In today’s world, humungous and heterogeneous data are being generated from every action of researchers, health organizations, etc. This fast, voluminous, and heterogeneous generation leads to the evolution of the term big data. Big data can be computationally analyzed to uncover hidden trends and patterns that help in finding solutions to the problems arising in various fields. Analysis of big data for manufacturing operational acquaintance at an unparalleled specificity and scale is called big data analytics. Proper utilization of analytics can assist in making effective decisions, improved care delivery, and achieving cost savings. Recognizing hidden trends and useful patterns can lead us to have a clear understanding of the valuable information that these data holds. This chapter presents a quality overview of big data and analytics with its application in the field of healthcare industries as these industries require their stream of data to be stored and analyzed efficiently in order to improve their future perspective and customer satisfaction.

DOI: 10.4018/978-1-5225-3870-7.ch004
INTRODUCTION

Data has a historic significance of being the most vital asset for organizations and governments. In the contemporary era of wearable devices and smartphones, very hefty amount of data and information are being produced about and by the peoples, things and their relations, which are being kept as a record. These data are being stockpiled in the databases and are of humungous and heterogeneous kind and are often termed as Big Data. Big data is going to be a bigger reservoir of huge, complex, structured or unstructured data that are being generated and collected from many digital sources like network devices, wireless sensors, medical equipment, legacy systems and many other such sources. Big Data can always serve as being a good source for several industries in different sectors that aim to automatically mine strategic information in reasonable time. While we insert effort on big data, it is critical to decide whether the benefits prevail over the overheads of storage and maintenance. Several analysis tools (Gupta & Saxena, 2014) are being designed and developed to investigate large data set and to have a better understanding of the impact of massive amounts of data in business improvements. To extract more benefits researchers and experts are trying to find a way to look into the future of big data.

Big data can generally be characterized by a set of V’s: volume, variety, velocity, value and veracity. Volume refers to the amount in order of petabytes or even zettabytes of data being generated in various healthcare industries and it is expected to get doubled every year. Healthcare systems are generating data at very fast pace. It includes information at the individual as well as disease/population specific levels such as medical record of a person, health information of a patient, radiology images, biometric sensor readings, 3D imaging and genomics. These data are very complex and large in volume and needs to be stored, managed and analyzed in order to extract useful information. The KMPG report (Galloro, 2008) indicates that in 2013, the healthcare data volume was about in excess of 150 Exabyte and it is increasing constantly. Variety refers to the diversity or variability of data being generated from a number of sources in the form of structured, semi-structured or unstructured data (Groves, 2013). Earlier, the organizations and enterprises were handling data that were restricted to a limited data sources and not so randomized in nature. In contrast, today’s scenario has changed a lot and organizations have to deal with a more complex data that may come from unalike sources and possesses more challenge while being stored, mined and analyzed. Healthcare data like clinical data, medical records, doctor notes, paper prescriptions, MRI images and radiograph films are so complex and difficult to augment with traditional data in order to get accurate precautions for patients. Velocity may be referred as the rate with which the data has been generated by the various sources like human interaction, business processes and healthcare systems. The regular updating of healthcare data doesn’t guarantee its correctness. The continuous generation of data is also accountable for being processed and analyzed. Also, the delivery of extracted and beneficial information has to be done in real time. The data moves through a number of systems of an organization varying from batch integration, then being loaded at a certain fixed interval of time to the streaming of data in real time (Sabharwal et al., 2016). Value provides the opportunity to answer questions and make decisions in emergency on the basis of insights gathered from analysis of data storage through some quality governance strategies and mechanisms that were previously considered out of reach. It also answers to the question that if the data analysis can discover a critical causative effect which may result in a remedy to a syndrome. Veracity means trustworthiness of data of different quality with meaning and applicability. To obtain proficient and effective results, high-quality data are required for analysis. Every form of information has different quality. In healthcare system, the quality of data is one of the most important aspects as the correctness of the data directly affects the life or death of patients.