Available online at http://www.journalijdr.com

ISSN: 2230-9926

International Journal of Development Research
Vol. 08, Issue, 12, pp.24445-24448, December, 2018

ORIGINAL RESEARCH ARTICLE OPEN ACCESS

BIG DATA ANALYTICS IN HEALTH INFORMATICS

***Dr. Kahkashan Tabassum**

Department of Computer Sciences, CCIS, Princess Nourah Bint Abdulrahman University, Riyadh, KSA

ARTICLE INFO

Article History:
Received 17th September, 2018
Received in revised form
28th October, 2018
Accepted 22nd November, 2018
Published online 26th December, 2018

Key Words:
Health Informatics, Bigdata Analytics, Healthcare, Data Cleaning, Electronic Health Record.

ABSTRACT

Big Data analytics in healthcare is emerging as a promising field and is most trusted technology today that is used to determine significant insights of huge data sets with minimized costs and improved results. Big data is defined as very large volume of high velocity, complex and variable data that needs innovative techniques to enable the data or information capture, storing, dissemination, management and analysis. The data volume of healthcare systems in enormous and is gaining fast momentum with respect to almost every area of research as well as industry. Nonetheless it provide appropriate storage and access platform for the healthcare systems also. Big data encapsulates the data characteristics such as variety, velocity, veracity with respect to healthcare. The conventional analytical techniques focuses on the large amount of (but currently unanalyzed) patient-related health and medical data available to influence more deeper understanding of the results and these could be further extended to the point of healthcare. The strategic methodology is based on the patient informing the data and the physician using it to perform decision making and giving most appropriate treatment to the latter. This paper discusses the Big data relevance, rationale, utilization and benefits with the healthcare perspective. It provides a summary of recent improvements in big data in health informatics and discusses how these will benefit from the integration of different personalized information from a varied range of data gathered by variable sources (structured or unstructured). In future it is predicted that recent advances in big data will help in increasing our knowledge for testing new theories of disease management from diagnosis to prevention.

Citation: Dr. Kahkashan Tabassum, 2018. "Big data Analytics in Health Informatics", *International Journal of Development Research*, 8, (12), 24445-24448.

INTRODUCTION

The healthcare industry is expanding drastically and thus generating a large volume of valuable data on patient demographics, treatment plans, payment, and insurance coverage. These upcoming aspects attract the attention of large base of clinicians and scientists to study and apply for future research. The current technology addresses different dimensions of data mining application in healthcare and has gained substantial attention of the researchers evident from the study of number of peer-reviewed articles published recently. A review of the literature on healthcare analytics using data mining and big data between 2005 and 2017 has revealed that the existing literature (Raghupathi and Raghupathi, 2014 Ahlemeyer-Stubbe and Coleman, 2014) mostly examines analytics in clinical and administrative decision-making and is based on human-generated data considering the wide adoption of Electronic Health Record (EHR) in clinical care (Vijay Sikka; Newton *et al.,* 2004). In recent years, the analytics rely on information of website or social media data but this lacks in prescriptive analytics in practice and integration of domain expert knowledge in the decision-making process emphasizes the necessity of future research. Analytics is an effective method of generating evolving insights through the efficient use of data and application of quantitative and qualitative analysis. It can generate elemental decisions for "planning, management, measurement, and learning" purposes. For instance, the Medicare Centers and Services rely on analytics to minimize the rate of readmission in hospital. And saved about $115 million in fake payments (Coleman *et al.,* 2016). Use of analytics involving data mining and text mining had solved many issues, new strategies based on big data analytics offer promising services to provide assistance to healthcare professionals in various ailment (pertaining to disease) prediction with effective diagnosis and

Corresponding author:* **Dr. Kahkashan Tabassum,
Department of Computer Sciences, CCIS, Princess Nourah Bint Abdulrahman University, Riyadh, KSA.

improved treatment facilities, resulting in progress in quality of service with reduced cost (Williams *et al.,* 2004). The study shows that the estimated reduction in cost saving is 450$ in US healthcare system on applying data mining every year (Bain , 2002). Dental practices are persistently collecting patient data but it is held as a patient dental record (PDR) often and remains as an underutilized resource. But due to the advancement in current technological strategies this data can be used to perform analysis and visualization. The analysis or the statistics once created using the PDR will help the doctors to diagnose the dental problems quickly and further guide the treatment also on a fast track. The emerging trends towards use of data analytics within various fields for instance in business will be used to guide business decisions. Similarly they could be applied in dental practices. The dental practices use computers (digital systems) to store a large reserve of patient information and it could be readily and effectively made available if the data tools of data analytics are used. Simple data analytic techniques are presented which can be used to extract substantial insight into patient demographics, DNA (did not attend) rates and many other areas of practical relevance to clinical service delivery and business management.

Background

Healthcare sector in many countries is growing rapidly and faces many challenges including rising costs, inefficiencies, poor quality, and increasing complexity (Yang *et al.,* 2015). Better decision-making based on available information could mitigate these challenges and facilitate the transition to a value-based healthcare industry (Berwick and Hackbarth, 2012). Healthcare institutions are adopting information technology in their management system (Prokosch and Ganslandt, 2009). A large volume of data is collected through this system on a regular basis. Analytics provides tools and techniques to extract information from this complex and voluminous data (Cortada *et al.,* 2012) and translate it into information to assist decision-making in healthcare. Data analytics is an emerging technology that is finding its application in almost every industry and has wide potential for providing doctors, patients and analysts to keep trach of their data and understand their patient base and changes over span of time. The dental practitioners could then be capable of practicing these information and could gain insight into patient demographics to guide dentists, make useful business decisions and also improve patient care. Its historical for the healthcare industry to produce large amounts of data (Connolly and Wooledge, 2013; Courtney, 2013), use the record keeping techniques, apply compliance and also accomplish the regulatory requirements and observe patient care. The data storage is conventional made and uses hard copy form. Due to the advent of new techniques based on the advancements of the current technology, a rigorous digitization of these data is made to happen and is reality today. To improve the healthcare delivery quality these huge masses of data (known as 'big data') while reducing the costs, supports an extensive range of medical and healthcare clinical decision support functions, disease surveillance, and population health management (Burghard, 2012; Dembosky, 2012; Feldman *et al.,* 2012; Bonnie, 2012). It is evident from the reports of U.S. healthcare system, that the data about the patients' health records (or Electronic Health records- EHR) is drastically increasing and soon it may become as huge in capacity as several yottabyte (one yottabyte = 1024 giga bytes) (Connolly and Wooledge, 2013). Big data needs to be accessed via complex tools and techniques since the conventional hardware and software management methods cannot be used. Therefore big data in healthcare also is very huge and utilizes sophisticated tools to access and manipulate healthcare data sets. The speed and diversity of data are other factors influencing big data management in health informatics industry. The diverse data storage in healthcare systems arises by the clinical PHR's corresponding to prescriptions of medical doctors, written notes on cases, medical imaging (such as x-ray, mammographic, panoramic, etc), data containing laboratory tests details, partial patient information with medical insurance, pharmacy, sensor data if sensors are deployed, details of emergency care, medical magazines, periodicals and journals, news, social media data from facebook or twitter and other managerial data. Big data analytics could be used to identify the relationship and patterns of association within the data stored and undoubtedly could be used to enhance healthcare facilities, saving lives at reasonable price.

Techniques for Data Analysis in Health Informatics

Recently published review/ research studies on healthcare and data mining outline the characteristics such as scope of healthcare sub-area and timeframe. For the medical practitioners it is required that they can effectively provide cure for their patients. Due to the advancement in technology it has become obligatory to many practitioners to be aware of the latest technological information to study the PHR of existing patients to give effective treatment solutions to the new patients by using the statistics of the PHR in the history. So one has to collect data for the research of the PHR and this data is either collected using data gathering techniques or existing online data available in the form of datasets. If the data collected is correct before data analysis can begin, the accuracy of the data collected has to be verified. Following data collection, the data needs to be critically analyzed. For any research, data analysis is very important as it provides an explanation of various concepts, theories, frameworks and methods used. It eventually helps in arriving at conclusions and proving the hypothesis. Data analysis is a process used to inspect, clean, transform and remodel data with a view to reach to a certain conclusion for a given situation. Data analysis is typically of two kinds: qualitative or quantitative. The type of data dictates the method of analysis. In qualitative research, any non-numerical data like text or individual words are analyzed. Quantitative analysis, on the other hand, focuses on measurement of the data and can use statistics to help reveal results and conclusions. The results are numerical. In some cases, both forms of analysis are used hand in hand. For example, quantitative analysis can help prove qualitative conclusions. Among the many benefits of data analysis, the more important ones are: Data analysis helps in structuring the findings from different sources of data. Data analysis is very helpful in breaking a macro problem into micro parts. Data analysis acts like a filter when it comes to acquiring meaningful insights out of huge data set Data analysis helps in keeping human bias away from the research conclusion with the help of proper statistical treatment. When discussing data analysis it is important to mention that a methodology to analyze data needs to be picked. If a specific methodology is not selected data can neither be collected nor analyzed. The methodology should be present in the dissertation as it enables the reader to understand which methods have been used during the research and what type of data has been collected and

analyzed throughout the process. The dissertation also presents a critical analysis of various methods and techniques that were considered but ultimately not used for the data analysis. An effective research methodology leads to better data collection and analysis and leads the researcher to arrive at valid and logical conclusions in the research. Without a specific methodology, observations and findings in a research cannot be made which means methodology is an essential part of a research or dissertation. The digital world of computers have several programs that will analyze data. These can be divided into those designed for use with quantitative data and those designed for analyzing qualitative data. These programs are designed to cover a very wide range of statistical techniques and rich in many features. There is a need therefore to work with technical experts who are familiar with the program. Some examples of data analysis programs for data are discussed in the following section.

Tools for analyzing quantitative data

1. **Epi-info:** Very good, simple, program which covers most of the statistical analyses which researchers would want to use. User-friendly screens, and guides to selecting and interpreting the statistics available.
2. **Minitab:** Simple program which covers all the basic statistical analyses. Not as easy to use as Epi-info.
3. **SPSS:** Commonly used comprehensive statistical package. Popular amongst social scientists. Complex to use.
4. **STATA:** It's a complex tool but is based on a powerful program with many sophisticated features. Is used by many statisticians and epidemiologists.

Data cleaning

Prior to data analysis, it is required to check data and perform data cleaning (Rahm, 2000). One should check that subjects that do not meet the original inclusion criteria are excluded at this stage but provided a report is created for how many are excluded and why. Assume that the data is on a spreadsheet or in a database it makes it easy to check out-of-range values. For instance, the possible values for gender may be (1 = Male, 2 = Female, 9 = Missing) (Bonnie, 2012). On examining the data if we found value as 3 then it is an error that needs correction. The data-cleaning process is used to recognizes absurd responses. For instance if data contains a record as being both male and pregnant which clearly marks it as an error.

METHODS

The preliminary step in the analysis is to define the characteristics of the members (participants) followed by comparison of the sample data selected with the population data from where it was taken. In describing the characteristics of the sample, there is typically a mix of data types. Some of the expressive statistics are given below: 1)Frequency counts 2) Proportions 3) Measures of mean, median, mode 4) Measures of standard deviation, inter-quartile range and so on

Ordinal data: This could be referred to as nominal data or categorical data. To determine how many times a specific value can appear in data-set the frequency count method could be used. Therefore when a sample data is to be considered and it has two elements one is 100 in number and the other is 70 then the frequencies can be expressed as proportions. With So, for example, the sample may consist of 100 males and 50 females. These can be represented wither by a pie or a bar chart. The average values (mean values) are however are not appropriate with respect to ordinal data. For unordered data (variables) generally *mode* can be determined which happens to be the most repeatedly occurring value in a data set. For ordered data (categorized) mode or median (middle value among ordered dataset) can be used.

Continuous data: The data here is based on three aspects: data distribution pattern by plotting frequency distribution graph, measuring average value within data, exploring the boundary of data. Most distributions for the data representation are found to be normal distributions. If not normal, they may have skewed distribution, with the curve mostly bent to left or right direction.

Benefits of Big Data Analytics: Big data analytics is the most reliable platform to analyze patient demographics to understand the various categories of patient types and predict the possible cure at early stage with respect to patients with similar health record. It provides a range of Personalized Healthcare systems to monitor the common health problems of patients. In addition it can be used to improve patient care and offer timely treatment to the patients and relief them from the both physical and psychological agony by reducing the time for suffering by the statistics provided after analyzing the PHR. Big data and analytics is being used to escalate the patient engagement offers wearable's including other health tracking devices by integrating with other technologies in the healthcare industry. It provides useful applications to monitor emergency health care systems. Due to shortage or in availability of data in proper form, it makes patient care programs difficult but if the PHRs are digitalized it becomes easy to recognize patient patterns effectively thus supporting the predictive analytics to reduce emergency situations by detecting chronic health issues, providing more flexible and customized curing plans to patients.

Conclusion

Big data analytics could be effectively applied in the health informatics in the area of clinical research studies since it is liable to gather varied and voluminous information from the real-world situations containing heterogeneous stakeholders and provide appropriate services. Patient satisfaction is the most significant criteria and a key task for the success of health care systems and Big data analytics with its set of techniques of analysis and visualization helps to achieve highest level of it. In spite of the growing challenges and issues of the patient needs it can provide sophisticated and well managed services and influences the future impact in the health industry. It facilitates rapid and seamless procurement of health data from huge and varied databases. Big data analytics with sensor technology could be used to avoid a narrow aspect of diagnosis of a disease by supporting a periodic monitoring of the disease which will help early diagnosis and may lead to prevention of the disease.

Future Impact of Study

1. The Big data analytics can transform the way healthcare providers in various medical fields including dentistry

use sophisticated technologies to obtain insight from their clinical, dental and other data repositories and make informed decisions.

2. In the future there an anticipation to witness rapid, widespread implementation and use of big data analytics across dental healthcare.
3. Big data analytics can fix the issues related to guaranteeing privacy, safeguarding security, establishing standards and governance
4. Big data analytics utilizes advance tools and technologies and subsequently make rapid advances in platforms to accelerate the process of contant development in healthcare industry.
5. In future there is a lot of potential corresponding to reduced costs, savings and still obtaining effective analytics.
6. Data Analysis is a constant process that continues and new solutions would be determined, the treatment by prediction would be implemented and finally most of the cases may be prevented from curing if problems detected in advance.

REFERENCES

Ahlemeyer-Stubbe A, Coleman SY. 2014. A Practical Guide to Data Mining in Business and Industry. London: Wiley.

Bain R. Analysis of patient attendance pattern. Dental Profile magazine 2002, NHSBSA Dental Service publications. Available at: http://www.nhsbsa.nhs.uk/DentalServices/ 2873.aspxGoogle Scholar

Berwick D.M., Hackbarth A.D. 2012. Eliminating waste in US health care. *J. Am. Med. Assoc.,* 307:1513–1516. doi: 10.1001/jama.2012.362.

Bonnie, 2012. http://www.west-info.eu/files/big-data-inhealthcare.pdf. 5. Fernandes L, O'Connor M, Weaver V: Big data, bigger outcomes. J AHIMA 2012:38–42

Burghard C. 2012. Big Data and Analytics Key to Accountable Care Success. IDC Health Insights.

Coleman SY, Gob R, Manco G, Pievatolo A, Tort-Martorell X, Reis M. 2016. How can SMEs benefit from big data? Challenges and a path forward. *J Qual Reliab Engineer Int.,* 32: 2151–2164.

Connolly S, Wooledge S. 2013. Harnessing the Value of Big Data Analytics. Teradata.

Cortada J.W., Gordon D., Lenihan B. 2012. The Value of Analytics in Healthcare. IBM Institute for Business Value; Armonk, NY, USA, Report No.: GBE03476-USEN-00.

Courtney M. 2013. Puzzling out big data. *Engineering & Technology,* 56–60.

Dembosky A. 2012. "Data Prescription for Better Healthcare." Financial Times, December 12, 2012, p. 19. Available from: http://www.ft.com/intl/cms/ s/2/55cbca5a-4333-11e2-aa8f-00144feabdc0.html#axzz2W9cuwajK.

Feldman B, Martin EM, Skotnes T. 2012. "Big Data in Healthcare Hype and Hope".

IHTT: Transforming Health Care through Big Data Strategies for leveraging big data in the health care industry; 2013. http://ihealthtran.com/ wordpress/2013/03/iht%C2%B2-releases-big-data-research-reportdownload-today/.

Intel: Big Data Analytics; 2012. http://www.intel.com/content/ dam/www/ public/us/en/documents/reports/data-insights-peer-research-report.pd

Newton JT, Bower EJ, Williams AC. 2004. Research in primary dental care. Part 1: setting the scene. *Br Dent J.,* 196: 523–526.

Prokosch H.U., Ganslandt T. 2009. Perspectives for medical informatics. *Methods Inf. Med.,* 48:38–44. doi: 10.3414/ME9132.

Raghupathi W, Raghupathi V. 2014. bBig data analytics in healthcare: promise and potential. *Health Inf Sci Syst.,* 2: 1–10.

Rahm E, Do HH. 2000. Data cleaning: problems and current approaches. IEEE Data Engineer Bull, 23: 3–13.

Vijay Sikka http://www.healthtechzone.com/topics/healthcare/ articles/2015/09/21/410192-business-intelligence-transforms-dental-industry.htm Google Scholar

Williams AC, Bower EJ, Newton JT. 2004. Research in primary dental care. Part 3: designing your study. *Br Dent J.,* 196: 669–674. Medline, Google Scholar

Yang J.J., Li J., Mulder J., Wang Y., Chen S., Wu H., Wang Q., Pan H. 2015. Emerging information technologies for enhanced healthcare. *Comput. Ind.,* 69:3–11. doi: 10.1016/j.compind.2015.01.012.

****