INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X
Impact factor: 4.295
(Volume 5, Issue 3)
Available online at: www.ijariit.com

Predictive Analytics in Healthcare

Ashish Asthana
ashish10.work@gmail.com
Dr. M. C. Saxena College of Engineering and Technology, Lucknow, Uttar Pradesh

ABSTRACT

The growing healthcare industry is generating a large volume of useful data on patient demographics, treatment plans, payment, and insurance coverage—attracting the attention of clinicians and scientists alike. In recent years, a number of peer-reviewed articles have addressed different dimensions of data mining application in healthcare. However, the lack of a comprehensive and systematic narrative motivated me to construct a literature review on this topic. In this proposal, I would like to seek more information on healthcare analytics using data mining and big data. Following Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines, I conducted a database search between 2005 and 2015 Critical elements of the selected studies—healthcare sub-areas, data mining techniques, types of analytics, data, and data sources—were extracted to provide a systematic view of development in this field and possible future directions. I found that the existing literature mostly examines analytics in clinical and administrative decision-making. Use of human-generated data is predominant considering the wide adoption of Electronic Medical Record in clinical care. However, analytics based on website and social media data have been increasing in recent years. Lack of prescriptive analytics in practice and integration of domain expert knowledge in the decision-making process emphasizes the necessity of future research.

Keywords— *Big Data, Analytics, Data Mining, Healthcare*

1. INTRODUCTION

The US healthcare system faces significant changes, including high cost, poor quality and variable performance. The digitized information is omnipresent everywhere because data is growing and moving faster than healthcare can consume it. This is due to mainly because of the efforts of research in the medical field and their discoveries. Widespread use of electronics record has totally transformed medical care, the latest innovation concerning genetics, smartphone and smart homes. This enables patient self-monitoring and treatment by simpler devices. The appearance of sensing technology like M-health, have increasingly contributed to the making data voluminous and appear like a digital flood. With an abundance of numbers, words, voices, images, video there is this increase in the parallel rapid growth of data. This humongous repository of data is Big Data. The concept of big data is not new, however, the way it is being defined is constantly changing. Various attempts have been made at defining big data, which essentially characterized as a collection of data elements. Their size, speed, type and complexity require one to seek and adopt and invent new hardware and software mechanism. In order to successfully store, analyze, and visualize data. Healthcare can be primarily said as a prime example of the three V's of data- velocity (the speed at which data is generated), volume and variety, which are innate aspects of data as it's produced. (Schroeck, Shockley, Smart, Romero-Morales, and Tufano, 2012). The data is part of multiple healthcare systems; research government entities and many more. The data repositories are soiled and inherently capable of providing a platform for global transparency. To add to the three V's, the veracity of the healthcare data is also critical for its meaningful use towards developing transitional research. The data resides in multiple places like individual EMR's, lab and imaging systems, physical notes, CRM systems and financial transactions. For that, a data-intensive research effort that pushes the limits of available data processing is required. Despite the inherent complexities of healthcare data, there is a potential and benefit in developing and implementing big data solution within this realm.

In this paper, I document the basic concept relating to big data. It will attempt to consolidate the hitherto fragmented discourses on what constitutes big data and what tools and technologies exist to harness the potential of big data. A key factor of this paper will be the use of analytics including data mining, text mining and big data analytics, assisting healthcare professionals in disease prediction, diagnosis treatment, resulting in improved service quality and reduction in cost. According to some estimates application of data mining can save up to 450 billion each year from US healthcare. This paper will also review literature related to the applied and theoretical aspects of data mining. We will clarify the types of analytics (descriptive, predictive, and prescriptive) also healthcare applications area (clinical decision support, mental health) and data mining techniques.

2. LITERATURE REVIEW

There has been a plethora of literature and conceptual studies on healthcare and data analytics. We will outline the characteristic of these studies for example scope, healthcare subarea, timeframe and number of papers reviewed in the table.

Table 1: Characteristic of studies

Data mining techniques and guidelines for clinical medicine	Timeframe 2005-2015
EMR and Visual Analytics	NA
Big data, level of usage	
Big data Analytics and its opportunities	
Text Mining, Adverse drug reaction	
The methodology of Big data analytics in Healthcare	2012-2015
Application and theoretical perspective of data mining and study of data analytics in healthcare	

It is drawn from above research that papers related to healthcare big data emerged in 2002 in the genomics area. The paper in this focus on several major themes, basic safety and protection of healthcare data in this era. The structured and unstructured data in the EMR and EHR include familiar data or redundancy of data is detected. The need to field code data at this point of care for electronic handling is a major barrier. On the other hand, the most provider agrees that an easy way to reduce errors is to use digital entries. The potential of big data in healthcare lies in continuing traditional data with new forms of data. Other characteristics that have added to big data are "veracity" [Dirette, D. P. (2016)] or data assurance. Big data analytics and outcomes should be error-free and credible. The veracity is a goal, not yet the reality. Data quality issues are of acute concerns in healthcare industry mainly for two reasons: The life and death are uncertain, their decision cannot be predicted and quality of healthcare data, especially unstructured data is highly variable and is often incorrect.

Veracity assumes the scaling up in granularity. The analytics architecture; tools for structured and unstructured big data are very different from traditional Business Intelligence tools. Veracity in healthcare data faces many of the same issues as financial data, especially on the payer side. Improving coordination of care, avoiding errors and reducing cost depends on high-quality data.

3. RESEARCH QUESTIONS

In guiding the research, the paper will seek to ask a number of questions, as listed below

(a) The source of data from organizations hospital, pharmacies, companies, and medical centres is in different formats. These organizations have data in different systems and settings. How to use this huge amount of data and be able to manage it?

(b) Quality of data is a serious limitation. Data collected are, in some cases, unstructured, improper, and no standardized. How this data can be standardized to have a meaningful outcome.

(c) 3. A big investment is required for companies to acquire staff (data scientists), resources and also to buy data analytics technologies. How and why Medical organizations will be convinced about using big data analytics?

(d) Using data mining and big data analytics requires a high level of expertise and knowledge. Will the companies be motivated enough to hire such persons as will be a costly affair for companies?

(e) Due to serious constraints regarding the quality of collected data, variations and errors in the results are not excluded.

4. METHODOLOGY

The research methodology will be both qualitative and quantitative. Qualitative methods will be used to collect primary data, in the form of observations, interviews and questionnaires as applicable. In other instances, the data will be collected through the use of meta-analysis, specifically secondary data. Applicable online search methodologies will be applied as necessary. The data collected will then be analyzed using statistical inferencing, and in line with the theoretical model proposed under the theoretical framework section.

5. CONCLUSION

Big data analytics has the potential to transform the way healthcare providers use sophisticated technologies to gain insight from their clinical and other data repositories and make informed decisions. In the future, we'll see the rapid, widespread implementation and use of big data analytics across the healthcare organization and the healthcare industry. Big data, including predictive analytics tools, have the potential to change the healthcare system from reporting to predicting results at earlier stages. In this paper, it has been aimed to present a global vision of big data for collecting, analyzing and managing healthcare data. 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.

6. REFERENCES

[1] Yang, J.-J.; Li, J.; Mulder, J.; Wang, Y.; Chen, S.; Wu, H.; Wang, Q.; Pan, H. Emerging information technologies for enhanced healthcare. Compute. Ind. 2015, 69, 3–11. [CrossRef]

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

[3] Raghupathi, Wullianallur and Viju Raghupathi. "Big data analytics in healthcare: promise and potential" Health information science and systems vol. 2 7 Feb. 2014, do: 10.1186/2047-2501-2-3

[4] Schroeck, M and Shockley, R and Smart, J and Romero Morales, Dolores and Tufano, P. (2012). Analytics: the real-world use of big data: How innovative enterprises extract value from uncertain data, Executive Report.

[5] Jeble, Shirish and Kumari, Sneha and Patil, Yogesh. (2016). Role of big data and predictive analytics. International Journal of

Automation and Logistics. 2. 307-331. 10.1504/IJAL.2016.10001272.

[6] Dirette, D. P. (2016) "Why the Veracity of Data Matters in Health Care Research," The Open Journal of Occupational Therapy: Vol. 4: Iss. 4, Article 1. Available at: https://doi.org/10.15453/2168-6408.1324

