Application of Data Analytics Principles in Healthcare

R Arokia Paul Rajan

Abstract— Information technology has transformed the healthcare field worldwide. In many areas of the healthcare industry, implementations of data analytics tools are commonly used recently. Applying data analytics principles in medical sciences appropriately transforms the mere storage of medical records in to discovery of drugs. Data science and analytics are essential tools because they can help make better decisions when it comes to spending and reducing inefficiencies in healthcare. The proposed model of healthcare data analytics provides a framework to accelerate the adoption and implementation of predictive analytics in healthcare. Healthcare data analytics can be applied to prove formulated hypotheses, test those using standard analytics models and predict patient health conditions. It can be used to classify patients at risk of developing diseases such as diabetes, asthma, and other life-long illnesses. In spite of the challenges faced while applying data science predictive analytics in the healthcare environment, there is an enormous opportunity for its usage in providing quality healthcare for patients.

Keywords: Healthcare, data mining, data analytics, predictive analytics, healthcare informatics.

I. INTRODUCTION

Healthcare and finance have nothing in common except that both the fields are being disrupted by technology. $3.5 billion was invested in 188 digital health firms i n the first half of 2017 [1]. For the past decennium, millions have been invested in building data warehouses. The historical problem is that the data reported informs on the actual data as what's happening and neither on the reason for its happening nor the steps to be taken to have the right outcomes. The secret to the successful transformation of the industry is therefore the use of data science in healthcare, as there is a high demand for value-based care [2].

The main objective for data scientists and machine learning experts is to gain a deeper understanding of the high volume of clinical data and attempt to reduce the expenses in health care. It is important to advance along the rapidly developing data science field and medicine.

Data science provides different tools, extraction methods for unstructured patient information and contributes to making healthcare more efficient, accessible and personalized. One of the areas in which data science predictive analytics can be applied is in the prognosis and diagnostic accuracy.

The methods of data science predictive analytics learn from historical data and forecast the outcome correctly. The data of the patients are collected, processed, analyzed and predictions are made. The algorithm choice will rely on the four factors: reliability, volume, data type and timeline of the project. Different algorithms may be tested to provide the most accurate prediction.

Data science in health care as shown in Fig. 1 uses scientific methods, processes, algorithms to extract knowledge and insights from data in various forms in order to test a hypothesis and predict possible outcomes to provide quality patient care.

Fig.1. Data science in healthcare

II. RELATED WORKS

In the medical healthcare sector, the role of predictive analytics is to achieve a high level of effective overall care and preventive care, as the results of predictive systems allow treatments and actions to be taken when risk recognition is identified in early stages, which in turn helps to minimize costs [3]. In addition, [4] it has been established that patients can also work and support medical care by tracking and updating their medical status so that they can obtain the necessary treatment at the right time. The innovation revolution has brought a significant benefit to the healthcare decision-making support system as decision-making processes in the healthcare sector can be improved by concentrating on patient identification, behavior and prevention to achieve a high level of care and improve the healthcare economy [5]. The prediction of health care is another method of data analysis that focuses on reducing future medical costs. Predictive analysis uses medical history of the patient to determine all potential health risks and foresee a future medical diagnosis beforehand. It is stated that

Revised Version Manuscript Received on 10, September 2019.
Dr R Arokia Paul Rajan, Associate Professor, Department of Computer Science, CHRIST (Deemed to be University), Bangalore, Karnataka, India. (Email: arokia.rajan@christunivewrsity.in)

predictive methods can help by predicting, minimizing time and cost by extracting and analyzing patient history records, information and diagnoses from the databases.

A predictive program has been developed by Parkland hospital in Dallas, Texas, which scans the details and information of all patients to identify potential hazards and outcomes. As a result, the hospital saved more than half a million dollars, particularly in cardiac failure and predictions of disease in terms of monitoring patients and avoiding future complications. Data Mining is defined as a process in which data is collected, analyzed and stored to generate information and knowledge of useful and high quality. This definition also involves how the data is collected, analyzed, ready for use, and finally the data processed in support of data analytics and predictive modeling [6]. Through information technology, the healthcare industry is largely regarded as one of the most important industries [7].

Progressively, IT has been viewed as a practice that facilitates healthcare performance through the efficient use of data and information within the healthcare sectors. Thus, [8] stated that, in order to understand the relationship between IT and healthcare, it is essential to understand the technologies used in healthcare. Over the past few years, IT functions have evolved not only as a provider of technology services, but also as a strategic provider that develops and integrates the infrastructures of industries to promote and ensure service quality [9].

The information technology of the mid-80s changed the healthcare industry and brought numerous benefits when using microcomputers, which for that time were small in size, simple and very strong. This also enabled hospitals to develop clinical applications for a variety of medical care settings. As a result, hospitals started to buy and implement healthcare information systems, and then problems began to emerge as practitioners tried to integrate data into these systems [10]. However, [11] acknowledged that information technology has enhanced healthcare industries, but also highlighted numerous difficulties related to the use of information technology in healthcare sectors, as they noted that the implementation of information technology in small clinics and organizations is exhausting, with high costs due to reduced scale efficiencies. IT implementation furthermore requires ongoing training and retention of experienced professionals.

[12] claimed that real-time analytics yielded more accurate results and data as it analyzed the history and symptoms of current patients, thereby correctly analyzing the diagnosis of patients and offering the best care. Real-time control strategies ensure that data is kept updated and that information reliability is improved, as suggested by [13]. Real-time issues in health care analytics are believed to be very important because it produces accurate results, such as where patients with diabetes can recover if their syndrome is correctly discovered and managed in the earlier stages. [14] acknowledged that they also highlighted some of the drawbacks, such as high costs, a high level of training needed and a lengthy period of time to complete.

III. GENERIC MODEL OF HEALTHCARE DATA ANALYTICS

The generic model of healthcare data analytics is a proposed framework to measure the adoption of processed patient data and the use of data science algorithms for predictive analytics in healthcare. The Generic model of Healthcare Data Analytics is organized into four major phases as shown in Fig. 2 where the understanding of each phase in stepwise fashion is necessary before moving on the next stage. The model is not essentially linear in its progression though in a perfect state that will be the case.

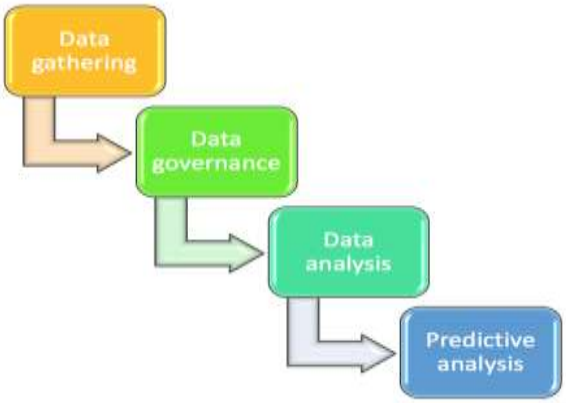

Fig. 2. Generic model of healthcare data analytics

A. Level 1- Data gathering

Data gathering is the building block for the model. The primary goal of any data gathering is to collect quality data that translates to rich data analysis leading to credible and conclusive answers to the framed hypothesis. The clinical data could be obtained from different sources represented in Fig. 3. By including data from patient cohort studies, the precision and quality of data could be maintained. Accurate data collection is extremely important to maintain the integrity of research.

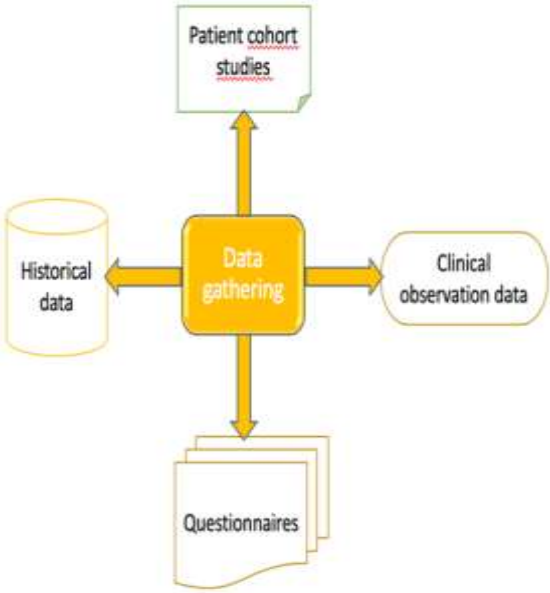

Fig. 1. Different data gathering methods

B. Level 2 - Data governance

At this level, standardized vocabulary and patient registries are established. The master reference table is defined to include the anonymized patient identity, clinical data, evolution parameters over time and others. The core data are related and organized according to the local master reference data. This level ensures that data is managed as an asset and transformed into meaningful information.

Data governance processes represented in Fig. 4 ensures that data meet precise standards and rules as it is organized in the system. It is necessary to develop a set of standards and procedures that determine how licensed staff will use the information. In addition, audit procedures and controls should be put in place to ensure continued compliance with internal policies and external regulations that ensure consistent use of data across multiple business applications.

Fig. 4. Data governance processes

C. Level 3 - Data analysis

Data analysis is also used as a synonym for data modeling. This level is shown in Fig. 5 is characterized by the analytic motive, focusing on consistent, efficient agile production of reports required for analyzing clinical data.

Fig. 5. Phases in data analysis

The emphasis is on inspecting, cleaning and modeling with the goal of finding useful information and drawing conclusions. Data mining is a data analysis technique focusing on modeling and knowledge discovery for predictive analysis. The analysis process may result in added data cleaning or additional data requests. These activities may hence be an iterative process. Data analysis plays a key role in making scientific decisions and helping to achieve effective operation.

D. Level 4 - Predictive analysis

The focus of this level represented in Fig. 6 includes predictive modeling, forecasting and risk stratification. A branch of advanced analytics, predictive analytics, is used to forecast unknown future events. It is important to use historical and transactional information trends to determine potential threats and opportunities.

Fig. 6. Predictive analytics process

IV. DATA VISUALIZATION & RESULTS

Data visualization is a common term used to describe any attempt by putting it in a graphical context to help people understand the importance of the data. Patterns, trends and correlations may not be detected in text-based data, but with data visualization software they can be exposed and recognized more easily. Visualization of data no longer includes flat bar charts or pie charts. The tools have been developed to enable complex, interactive visualizations with dynamic information.

The analyzed patient data may be reported in many formats. The data visualization techniques could be used to clearly and efficiently communicate a message. The chosen methods are those best suited to the processing of visual information by our brains. It is important to adjust and customize graphs to the audience when deciding on a visualization approach. For example, an internal usage exploratory plot will be different from a graph intended to communicate a result to a general audience. Efficient data processing is a good solution to knowledge that is unreliable.

V. CASE STUDY: THE IMPACT OF VACCINES ON INFECTIOUS DISEASES BATTLING

Vaccines have saved millions of lives. The controversial paper [4] published by Andrew Wakefield and his colleagues in 1988 alleged a connection between the vaccine and autism for measles, mumps, and rubella (MMR). Although the sample size was meagre (n=12), the chaotic format, and the speculative nature of the results, the paper gained broad attention, and vaccination rates for MMR began to drop as parents became worried about the risk of autism following vaccination. Nevertheless, misunderstandings exist in part because of self-proclaimed protesters who continue to spread myths about vaccinations. The Tycho Project collected, compiled and distributed the information used for the study, including weekly recorded count data from fifty states in the U.S. for seven diseases from 1928 to 2011.

The chart shown in Fig. 7 strikes an argument of the value of vaccines are very interesting. The outbreaks of infectious diseases over time and the effects of vaccination demonstrate that it is not possible to take for granted the effectiveness of immunization programs.

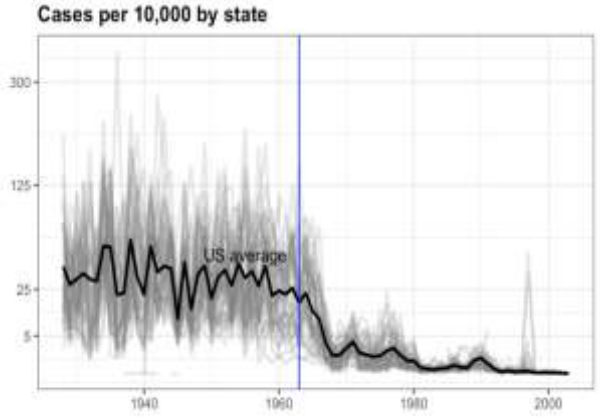

Fig. 7. Impact of vaccines on battling infectious diseases

Most immunization effects are recognized over decades that may not be immediately apparent. From this it is apparent that while it is possible to accurately predict the direct benefits of immunization, these form only a reference. For example, prevention of childhood disease is linked to better educational success and higher later-life earnings [5], [6]. Assessing the overall impact of immunization should be regarded as an ongoing process that requires pre-implementation modeling and subsequent long-term monitoring.

VI. CONCLUSION

A lot of challenges are faced with data science and predictive analysis in the healthcare sector. The primary challenge is that adopting the latest advancements in technology and tools that are widely used for analytics is far behind other sectors and facing many problems. It leads to plenty of opportunities and scope for improvement of each technique in the healthcare sector. Another important challenge is the sensitivity of patient information. Healthcare companies are fearful of data privacy and cannot rely on sensitive information for technology. The ultimate challenge is the end-use of every predictive model. There are situations where false positives and false negatives produce varied results. However, a thorough knowledge of prediction models and the limitations that can come across end-users is essential.

In the healthcare sector, data science predictive analytics have a large scope for the future and much can be offered to the industry. These models once adopted in the system could provide better healthcare for the patient.

VII. ACKNOWLEDGMENT

The author would like to sincerely appreciate the efforts made by Ms. Aroul Prissila who is currently working as Senior Project Manager, Modul-Bio, Marseille, France for making this article.

REFERENCES

1. https://www.altexsoft.com/blog/datascience/7-ways-data-science-is-reshaping-healthcare
2. Islam, Md Saiful et al., A Systematic Review on Healthcare Analytics: Application and Theoretical Perspective of Data Mining". Healthcare
3. Conley, E., Owens, D., Luzio, S., Subramanian, M., Ali, A., Hardisty, A., & Rana, O., Simultaneous trend analysis for evaluating outcomes in patient-centred health monitoring services, Health Care Management Science, vol. 11, no. 2, pp. 152-66, 2008.
4. Obenshain, M. , Application of data mining techniques to healthcare data, Infection Control and Hospital Epidemiology, vol. 25, no. 8, pp. 690-695, 2004.
5. Cannon, M., & Tanner, M., Healthy competition: What's holding back healthcare and how to free it, Cato Institute, Washington, D.C., 2007.
6. Russom, P., Executive summary big data analytics, TDWI best practice report, Renton, Washington, 2011.
7. Wager, K.A., Lee, F.W. & Glaser, J.P., Managing Health Care Information Systems: A Practical Approach for Health Care Executives, Wiley, United states of America, 2005.
8. LeRouge, C., Mantzana, V. & Vance Wilson, E., Healthcare information systems research, revelations and visions, European Journal of Information Systems, vol. 16, pp. 669 – 671, 2007.
9. Bhattacherjee, A & Hikmet, N., Physicians' resistance toward healthcare information technology: a theoretical model and empirical test, European Journal of Information Systems: Including a Special Section on Healthcare Information, vol. 16, pp. 725–37, 2007.
10. Murphy, K., What about healthcare must change for analytics to flourish?', https://ehrintelligence.com/2013/08/29/what-about-healthcare-must-change-foranalytics-to-flourish, 2013.
11. Taylor, J., Transforming healthcare delivery with analytics: Improving outcomes and point-of-care decisioning, Decision management solutions, ftp://public.dhe.ibm.com/software/data/swlibrary/infosphere/analyst-reports/Transforming_Healthcare_Delivery.pdf, 2010.
12. Walker, M., Fonda, SJ., Salkind, S. & Vigersky, R., Advantages and disadvantages of real-time continues glucose monitoring in people with type 2 diabetes, Blood glucose monitoring, vol. 8, no.1, p. 25, 2012.

13. http://www.searo.who.int/entity/health_situation_trends/topics/health _data_standrads/en/
14. C. Roed, L.H. Omland, P. Skinhoj, K.J.Rothman, H.T. Sorensen, N., Obel Educational achievement and economic self-sufficiency in adults after childhood bacterial meningitis, JAMA, 309, pp. 1714-1721, 2013.

AUTHOR PROFILE

Dr R Arokia Paul Rajan is currently working as Associate Professor, Department of Computer Science, CHRIST (Deemed to be University), Bengaluru, Karnatka, India. He holds Ph.D. in Computer Science & Engineering from Pondicherry University, India. His research area is data management in Cloud architectures. He published 15 research papers in various international journals and conferences.