ORIGINAL RESEARCH PAPER

Volume-8 | Issue-4 | April-2019 | PRINT ISSN No 2277 - 8179

INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH

USING BIG DATA ANALYSIS IN HEALTH CARE INDUSTRY

Neurology

Dr. Tania Mahapatra MBA, MPT, Neurology, Medico Research Analyst, Samapriya Health Data Analytics, Delhi

ABSTRACT

Big Data is not new but its utility in health care field is still at beginning stage. Health data range from medical histories, lab reports, treatment outcomes and patient's experiences. This vast and voluminous data when scientifically analysed unearth hidden relations and conclusions that can aid in improving public health at economical costs. It can be used to describe, prescribe or predict occurrences. A clinician can provide the best health care service to a patient based on past observations and present information. It can also contribute to efficient administration. If challenges pertaining to privacy concerns and data accuracy are managed well, Big Data analysis can turn around the picture of health care industry for good.

KEYWORDS

Big Data, health data analysis, predictive analysis, clinical DSS

INTRODUCTION

With the digitization of health care industry, terms like Big Data have taken a place beside the usual medico jargons. For a medical organization, it refers to the large, complex, distributed and ever-growing information being produced continuously (Kuo *et al*., 2011) and that need an infrastructure beyond the traditional system to store and analyze it. The era of Big Data has not dawned upon us suddenly. Sun and Reddy (2013) had already foretold a manifold increase in worldwide digital healthcare data making it evident from the beginning that no one tool or application will be capable enough to manage such compound information. Further, health data is complex (Kuo *et al*., 2014) because it includes X-rays, MRI, unstructured physician notes, clinical terminologies or codes that when analysed properly can unlock hidden trends, providing insights into new knowledge, understanding public health or combating epidemics (Agrawal *et al*, 2012). The correlations among various parameters can help describe or predict certain behaviours or events that can solve many of daily dilemmas. Therefore, the vast knowledge can be positively exploited for remarkable insights to provide better judgment and better resource allocation to the public (Figure 1).

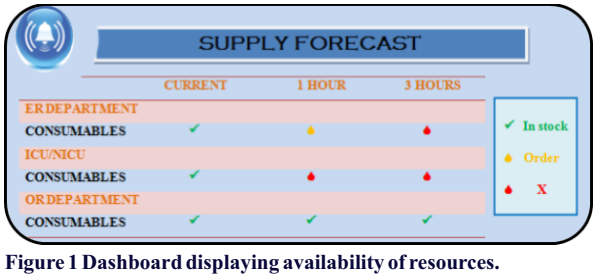

Figure 1 Dashboard displaying availability of resources.

Literature supports the use of Big Data analysis for healthcare professionals. Bates *et al* (2014) demonstrated identifying and managing high-risk patients in terms of cost, triage and patient readmissions using analytics. Kim (2013) had also described various ways in which Big Data could be exploited to improve many sectors like clinical decision support, administrative services, and user behaviors. Manogaran, *et al* (2017) described important characteristics of Big Data as following,

1. Volume: Diverse information in the form of text, audio, video etc. are generated on a mass and huge scale on daily basis generating Big Data.
2. Velocity: Within minutes, any information on social media becomes old and there are the latest feeds. Millions of posts or tweets are circulated and news spread like raging fire across the globe.
3. Variety: Data can be in any format like XML, CSV, TXT, MP3, FLAC, DOCX, XLS, AIFF, PDF and JPEG.
4. Veracity: The true data that is relevant for the organization must be sorted out from the irrelevant so as to increase the reliability of the collected information.
5. Validity: The data accrued must be precise and relevant for any future application.
6. Variability: The data stream needs to be consistent and regular even in case of unpredictable circumstances.

ARCHITECTURAL STRUCTURE

The process of analytics (Figure 2) begins with the extraction of health data from multiple sources – internal and external. Information from patient records, hospital data, and clinical sheets constitute internal data sources. On the other hand, publicly available information, like WHO, or data from government sources constitute external sources. Either the raw data is aggregated and processed further for use or the data remains as such to be retrieved for processing as and when required. Features to regularly upload new data, like patient discharge or emergency procedure, are a must for real-time update. When the data is ready for analysis, the system may employ a number of tools depending upon the requirements. Statistical software like "SPSS" or "SAS" or advanced BI (business intelligence) or data mining tools like "Dundas" or "IBM Cognos" are well-known (Raghupathi and Raghupathi, 2013). The outcome depends upon the kind of analysis performed on the data. If a query is performed, answers to questions like "identify patients who would most likely miss the next appointment" can be gathered. For viewing reports related to compliance, performance indicators or outcome measures, dashboards and scorecards can be generated. In case of Online Analytic Processing (OLAP), unlimited report viewing and predictions for "what if" situation for diagnosis and treatment is done. Finally, using data mining, clinical data can be extracted, analyzed and interpreted to help in the clinical decision-making process.

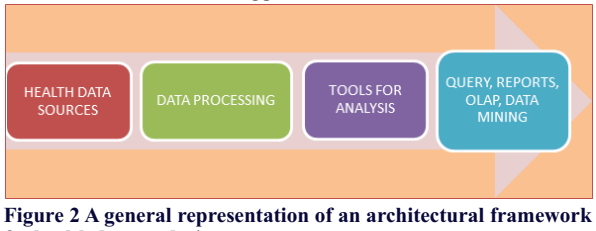

Figure 2 A general representation of an architectural framework for health data analysis

PRACTICAL ILLUSTRATIONS

In practical life, a clinician has to manage hundreds of cases having co-morbid complications with poly drug therapy. It is unreasonable to expect the treating doctor to be updated with the latest findings of every drug's risk factors, probable interactions, and complications at real-time. Descriptive software that acts an electronic drug database may come to physician's aid by providing the latest drug information and assist in prescribing them (Figure 3).

Figure 3 Illustration of a Medical Reference App.

Further, as depicted in Figure 4, patients' medical profile can be matched against current medications as well as recommended therapies for a comprehensive approach. All relevant details at one place assist in effective clinical decision making.

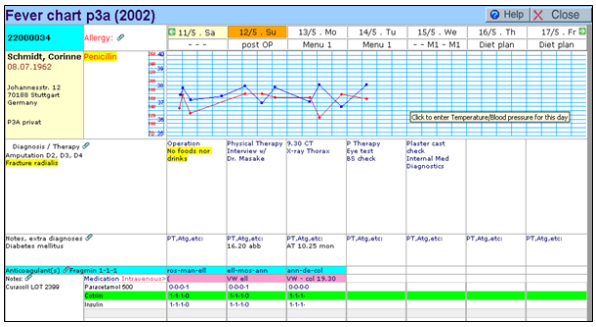

Figure 4 Descriptive analysis of a patient. ***Image Source: Wikimedia Commons***

More often than not, a doctor has to go through multiple treatment plans before finalizing which drug regime is suitable to a patient. Still, the patient might feel dissatisfied causing drift. To avoid such scenarios, predictive analysis and machine learning provide an arsenal to assist in clinical decision making (Elhanan *et al,* 1996) through quick determination of *best* treatment plan based on the patient's history and medical records (Figure 5).

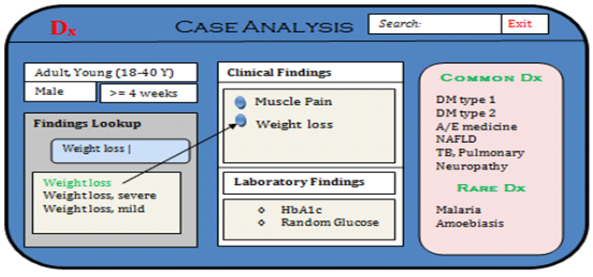

Figure 5 Illustration of a Clinical Decision Support System.

The technique uses the large data from the patient population for quick analysis of the present condition for an accurate diagnosis and a treatment plan with clear milestones and expected outcomes (Figure 6).

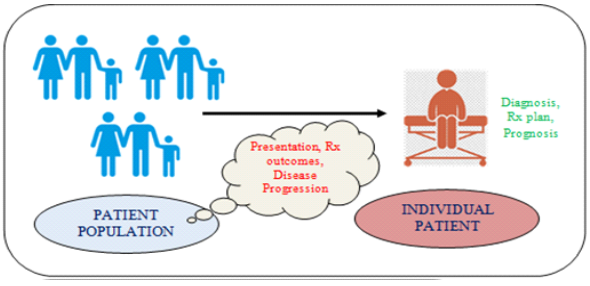

Figure 6 The basis of Predictive Analysis.

Healthcare analytics also helps in triaging, identifying co-morbidities, and suggesting the *best* patient-doctor match. It can predict readmissions, length of stay (Lisk *et al*, 2018), bed availability (Figure 7) or hospital-acquired infections (Qin and Ward, 2018).

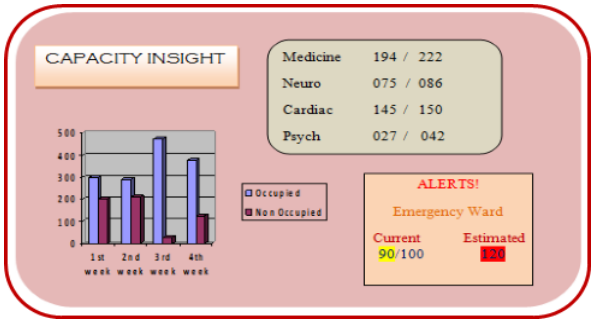

Figure 7 Tracking bed occupancy.

It has also shown advantages in revenue management. New reimbursement plans, consumer-driven infrastructures, and competitive market has made it imperative to deliver higher quality and more accessible healthcare services under minimum administrative and clinical expenditure (Holmes *et al*, 2017). Predictive analysis consider factors like treatment results, length of stay, patient readmissions, insurance claims or panel reimbursements to predict future events in regards to revenue and payments (Smith *et al*, 2015).

Patient satisfaction is another crucial parameter that is affected by waiting time, service delay, query resolution and conduct by staff etc., and can be analysed for improving it (Batbaatar *et al*). Some organisations also adopt voice recognition software in their business models to interpret the level of patient satisfaction through keywords, inflection, and tone.

Another pertinent issue in clinical zone is the 'No-Show' by the patients on the appointed date (Curtis *et al*, 2018) thereby compromising their health and placing a greater burden on the resources. Using the data of absence (Figure 10) for previous patients, analysis can assist in predicting which patient profile has the probability of not showing up in their next appointment (Boyle *et al*, 2012). Reminders or counseling the significance of keeping up with follow-ups can improve compliance.

	A	B	C	D	E	F	G	H	I	J
1	Patient ID	Age	Gender	Appointment Date	Neighbourhood	Hypertension	Diabetes	Alcoholism	Handicap	SMS Reminders	No Show
2	1323	43	M	12-03-2019	West District	0	0	0	0	yes	0
3	1324	56	M	12-03-2019	West District	1	0	0	0	yes	0
4	1325	57	M	12-03-2019	East District	1	0	1	0	yes	1
5	1326	23	F	12-03-2019	South District	0	0	0	1	yes	1
6	1327	34	M	13-03-2019	West District	0	0	0	0	yes	0
7	1328	45	F	13-03-2019	East District	0	0	0	1	yes	1
8	1329	43	F	15-03-2019	East District	1	1	0	1	yes	0
9	1330	42	F	15-03-2019	North District	1	0	0	1	yes	0
10	1331	67	F	15-03-2019	North District	1	1	0	0	yes	0
11	1332	39	M	15-03-2019	East District	0	0	1	0	yes	0
12	1333	32	F	15-03-2019	South District	0	0	0	0	yes	0
13	1334	76	M	15-03-2019	South District	1	0	0	0	yes	0
14	1335	46	F	16-03-2019	South District	1	1	0	0	yes	0
15	1336	45	F	16-03-2019	East District	0	1	0	0	yes	0
16	1337	53	F	16-03-2019	North District	1	1	0	0	yes	0
17	1338	60	M	16-03-2019	West District	1	1	0	1	yes	1
18	1339	89	M	16-03-2019	North District	1	1	0	0	yes	1
19	1340	35	F	16-03-2019	North District	0	0	0	0	yes	1
20	1341	57	M	16-03-2019	South District	0	1	0	0	yes	1

Figure 10 Dataset for 'No-Show'.

Likewise, the data can help in making arrangements to schedule appropriate staff and other resources to minimise waiting time (Miro *et al*, 2003), enable smooth operations and support efficient practices (Gopakumar *et al*, 2016).

Issues And Prospects For An Health Organisation

Like every technology, health data analytics comes with its share of challenges. The first one is accuracy of results. History regarding lifestyle practices (e.g. smoking) or genetic diseases (e.g. Thalassemia) is usually underplayed or ignored by patients that affect the data accuracy. Certain cultures discourage open discussions on health issues raising concerns for privacy. Techniques that utilise anonymous submission of data and standardised formats of data collection can achieve consistency and reliability to a large extent.

To truly benefit from analysis, it is important to have required infrastructure to store every bit of information, like medical history, diagnosis, lab findings, treatment or billing, to corroborate, control and integrate them into meaningful data source for the BI. Researchers can be granted restricted access based on valid credentials for studying the public data.

If leveraged ethically, health data analytics is a significant skill that can convert massive amount of everyday data into relevant information for better healthcare services, improved patient satisfaction and efficient practices at reduced costs.

REFERENCES:

1. Agrawal, D. et al. (2012) Challenges and Opportunities with Big Data, Big Data White Paper- Computing Research Association [online] http://cra.org/ ccc/docs/init/ big data white paper. pdf
2. Batbaatar, E., Dorjdagva, J., Luvsannyam, A., Savino, M. M., & Amenta, P. (2017). Determinants of patient satisfaction: a systematic review. Perspectives in public health, 137(2), 89-101.
3. Bates, D., Saria, S., Ohno-Machado, L., Shah, A., Escobar, G.: Big Data in health care: using analytics to identify and manage high-risk and high-cost patients. Health Aff. 33, 1123–1131 (2014)
4. Boyle, Justin, et al. "Predicting emergency department admissions." Emerg Med J 29.5 (2012): 358-365.
5. Curtis, C., Liu, C., Bollerman, T. J., & Pianykh, O. S. (2018). Machine learning for predicting patient wait times and appointment delays. Journal of the American College of Radiology, 15(9), 1310-1316.
6. Elhanan, G., Socratous, S. A., & Cimino, J. J. (1996). Integrating DXplain into a clinical information system using the World Wide Web. In Proceedings of the AMIA Annual Fall

Symposium (p. 348). American Medical Informatics Association.
7. Gopakumar S, Tran T, Luo W, Phung D, Venkatesh S Forecasting Daily Patient Outflow From a Ward Having No Real-Time Clinical Data JMIR Med Inform 2016;4(3):e25
8. Holmes, G. M., Kaufman, B. G., & Pink, G. H. (2017). Predicting financial distress and closure in rural hospitals. The Journal of Rural Health, 33(3), 239-249.
9. Jee Kim, G.: Potentiality of Big Data in the medical sector: focus on how to reshape the healthcare system. Healthc. Inform. Res. 19, 79 (2013)
10. Kuo, M. H., Sahama, T., Kushniruk, A. W., Borycki, E. M., & Grunwell, D. K. (2014). Health big data analytics: current perspectives, challenges and potential solutions. International Journal of Big Data Intelligence, 1(1-2), 114-126.
11. Kuo, M.H., Kushniruk, A. and Borycki, E. (2011) 'A comparison of national health data interoperability approaches in Taiwan, Denmark and Canada', Electronic Healthcare, Vol. 10, No. 2, pp.14–25.
12. Lisk, R., Uddin, M., Parbhoo, A., Yeong, K., Fluck, D., Sharma, P., & Han, T. S. (2018). Predictive model of length of stay in hospital among older patients. Aging clinical and experimental research, 1-7.
13. Manogaran, G., Thota, C., Lopez, D., Vijayakumar, V., Abbas, K. M., & Sundarsekar, R. (2017). Big data knowledge system in healthcare. In Internet of things and big data technologies for next generation healthcare (pp. 133-157). Springer, Cham.
14. Miro, O., Sanchez, M., Espinosa, G., Coll-Vinent, B., Bragulat, E., & Milla, J. (2003). Analysis of patient flow in the emergency department and the effect of an extensive reorganisation. Emergency Medicine Journal, 20(2), 143-148.
15. Qin, S., & Ward, D. (2018, November). Forecasting Hospital Daily Occupancy Using Patient Journey Data-A Heuristic Approach. In International Conference on Advanced Data Mining and Applications (pp. 150-159). Springer, Cham.
16. Raghupathi, W., & Raghupathi, V. (2013). An overview of health analytics. J Health Med Informat, 4(132), 2.
17. Smith, M. W., Friedman, B., Karaca, Z., & Wong, H. S. (2015). Predicting inpatient hospital payments in the United States: a retrospective analysis. BMC health services research, 15(1), 372.
18. Sun, J. and Reddy, C.K. (2013) 'Big data analytics for healthcare', Tutorial Presentation at the SIAM International Conference on Data Mining, Austin, TX.
19. Wikimedia Commons (n.d.). File:Sshot fever.png. https://commons.wikipedia.org/ wiki/ File: Sshot_fever.png