

Big Data Analytics For Medical Applications

R.Durairam, AP/CSE, Mahendra Institute of Technology, Mahendrapuri,
Mallasamudram (W), Thiruchengode, Namakkal, Tamil Nadu 637503

Abstract:

With the advent of technology and growing amount of data (Big Data), need is felt towards implementing effective analytics techniques (Big Data Analytics) to analyse this big volume of data for unknown and useful facts, patterns, associations and trends which can give birth to new line of treatment of diseases and provide high quality healthcare at lower cost to all. This paper gives a wide insight and know how about the various Big Data analytics (BDA) initiatives taken to improve healthcare worldwide. It also explains the various phases involved in BDA process and depicts its benefits and challenges with focus on healthcare industry.

Keywords – BDA, Big Data, Health care.

1. INTRODUCTION

Big Data is a buzzword which is reigning the innovation market from quiet sometime. An enormous amount of data often referred to as Big Data is getting generated everyday by diverse segments of industries like business, finance, manufacturing, healthcare, education, research and development etc. The traditional DBMS's and RDBMS's in market are incapable of storing such vast amount of data. We cannot take full advantage of the hidden knowledge and information from this data as the traditional data mining algorithms do not work effectively on this enormous data. So there is need of developing and using effective, innovative tools and technologies offered by Big Data. wearables collect real time physiological data of patients at a rapid pace or velocity. This new data being generated every second poses a big challenge for data analysts. Social media data also adds to velocity as the users views, posts, feeds scale up in seconds to enormous amount in case of epidemics/national disasters.

2. WORLDWIDE BDA INITIATIVES

There have been upsurge in the number of big data initiatives around the globe. Some of the initiatives are as follows: New Zealand's Ministry of Health in collaboration with experts from the New Zealand Society for the Study of Diabetes (NZSSD) have used SAS data analysis capabilities, for establishing a Virtual Diabetes Register (VDR) that combines and filters various sources of health information to more accurately determine how many people are diagnosed with the diabetic condition, as well as predicting who is likely to develop diabetes in the future. The Data Science Institute of Columbia University, New York in collaboration with the New York City Department of Health and Mental Hygiene (NYC DOHMH), is working on a project that focuses on the detection of disease outbreaks in New York City restaurants. The goal of the project is to identify and analyze the unprecedented volumes of user- contributed opinions and comments on social media sites such as Twitter, Face book and Yelp, which host massive amounts of content by users about their real-life experiences and opinions about restaurants. It will help to extract

reliable indicators of otherwise-unreported disease outbreaks associated with the restaurants. Seattle Children's Hospital and Regional Medical Center, is a 250-bed children's hospital in the Laurelhurst neighbourhood of Seattle, Washington. It treats 350,000 patients annually. It is using big data analytics as part of its Clinical Standard Work (CSW) program, which defines patient populations and recommends an ideal protocol for each population, allowing ensuring that every patient at the hospital receives the same standard of care. The CSW program gets the enormous data from enterprise data warehouse (EDW) which currently integrates data from 10 sources across the hospital, including electronic medical records (EMRs) and billing systems.

Web Services. The software uses natural language generation techniques, to examine through structured data and automatically present it in story form. US State of North Carolina processes about 88 million claims totalling about \$12 billion annually from 66,000 providers who treat the state's two million Medicaid patients. The State's Department of health and human services in collaboration with IBM used big data analytics to help identify suspicious billing patterns by healthcare providers. Using three years' worth of North Carolina Medicaid claims data, IBM data mining software, which featured special algorithms and modelling capabilities, was applied to detect common fraud and abuse schemes. Almost 90% reduction in fraud was achieved.

3. PHASES IN THE BDA PROCESS

As already mentioned the data is fed to the system through many external sources like clinical data from Clinical Decision Support systems (CDSS), EMR, EHR, machine generated sensor data, data from wearable devices, national health register data, drug related data from Pharmaceutical companies, social media data like twitter feeds, Facebook status, web pages, blogs, articles and many more.[3] This data is either stored in databases or data warehouse. With advent of cloud computing, it is convenient to store such voluminous data on the cloud rather than on physical disks. This is more cost effective and manageable way to store data. The data which has been acquired should be complete and should be in a structured format, for performing effective analysis. Generally it is seen in that healthcare data from flaws like, many patients don't share their data completely like data about their dietary habits, weight and lifestyle. In such cases the empty fields need to be handled appropriately. Another example can be for e.g.: for field like Gender of person, there can be at most one of two values i.e. male or female. In case any other value or no value is present then such entries need to be marked and handled accordingly. The data from sensors, prescriptions, medical image data and social media data need to be expressed in a structured form suitable for analysis. Once the data is cleaned and integrated, the next step is to query the data. A query can be simple query like for eg: What is mortality rate in a particular region? or complex query as how many patients with diabetes are likely to develop heart related problems in next 5 years? Depending on the complexity of the query, the data analyst has to choose appropriate platform and analysis tools.

CONCLUSION

BDA is a process which has many steps like data acquisition, cleaning etc and each step has a unique challenge. We need to address these challenges and formulate new technological standards, protocols so that at least technologically we become competent enough to manage and analyse such volume of complex data. As has been seen in existing studies, the BDA has shown remarkable outcomes in many healthcare organizations. In the future, with even more advancements in the BDA processes we expect that healthcare cost will come down drastically, life expectancy will increase, and we will see much healthier population as compared to now with people taking more accountability and charge of their health using technological

advancements. The future of healthcare is promising.

REFERENCES

- [1] Vivek Wadhwa, "The rise of big data brings tremendous possibilities and frightening perils", April 2014. Available: <http://www.washingtonpost.com/blogs/innovations/wp/2014/04/18/therise-of-big-data-brings-tremendous-possibilities-and-frightening-perils/>
- [2] D. Agrawal et. al, "Challenges and Opportunities with Big Data", Big Data White Paper- Computing Research Association, Feb 2012. Available: <http://cra.org/ccc/docs/init/bigdatawhitepaper.pdf>
- [3] Wullianallur Raghupathi and Viju Raghupathi, "Big data analytics in healthcare: promise and potential", *Health Information Science and Systems* 2014, 2:3, Available: <http://www.hissjournal.com/content/2/1/3>
- [4] Nambiar, R. ; Cisco Syst., Inc., San Jose, CA, USA ; Bhardwaj, R. ; Sethi, A. ; Vargheese, R., "A look at challenges and opportunities of Big Data analytics in healthcare", IEEE Conference 2013. Available: http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=6691753&url=http%3A%2F%2Fieeexplore.ieee.org%2F%2Fxppls%2Fabs_all.jsp%3Farnumber%3D6691753
- [5] Ahmed E. Youssef, "A Framework for Secure Healthcare Systems Based on Big Data Analytics in Mobile Cloud Computing Environments", *The International Journal of Ambient Systems and Applications* 06/2014, Available: <http://airccse.org/journal/ijasa/papers/2214asa01.pdf>
- [6] J. Archenaa, E.A. Mary Anita, "A Survey of Big Data Analytics in Healthcare and Government", *Procedia Computer Science*, Elsevier, Volume 50, 2015, Pages 408–413, Big Data, Cloud and Computing Challenges, Available: <http://www.sciencedirect.com/science/article/pii/S1877050915005220>
- [7] Matthew Herland, Taghi M Khoshgoftaar and Randall Wald, "A review of data mining using big data in health informatics", *Herland et al. Journal of Big Data* 2014, Springer.