World Journal of Advanced Research and Reviews
eISSN: 2581-9615 CODEN (USA): WJARAI
Cross Ref DOI: 10.30574/wjarr
Journal homepage: https://wjarr.com/

(REVIEW ARTICLE)

Public health dentist as a big data scientist: A review article

Praveena. S *, Indhu Malar. C, Bhuvaneswari. M, Shyam Sivasamy and Aravinth. V

Department of Public Health Dentistry, Faculty of Dentistry, Meenakshi Ammal Dental College and Hospital(MAHER), Alapakkam main road, Maduravoyal, Chennai-95,

World Journal of Advanced Research and Reviews, 2024, 22(03), 1302–1305

Publication history: Received on 11 May 2024; revised on 19 June 2024; accepted on 22 June 2024

Article DOI: https://doi.org/10.30574/wjarr.2024.22.3.1847

Abstract

This paper is to showcase the possibility of public health dentists as big data scientists to analyse large sets of data. Any data collected at national or state level is complex and enormous in nature and hence can be handled only by using advanced software. Professionals who gathers, accumulate data from disparate sources, clean, organise & process it and then develop methods to improve data mining statistically are called big data scientist. The required skills to become a data scientist are programming machine learning techniques, data visualisation and reporting, data mining, and cleaning. Acknowledging the utility of big data science and predictive analytic in public health surveillance, the possibility of training a public health dentist in big data science should be explored. Being an only speciality in dentistry which has biostatistics as one of its core tools, dental public health can be related to big data science. Bringing a change in the curriculum to explore the young graduates in this upcoming field can be a value addition to get competencies and pave way to entry into world of big data science.

Keywords: Big data; Electronic health records; Biostatistics; Health care; Public health

1. Introduction

Most people are aware of how "data" is altering our world. Big data and large data sets were initially created in the early 1960s and 1970s. In the early 2000s, in response to the explosion of internet data sources, new approaches for capturing, archiving, processing, and managing this potentially valuable information were sought. Data science derives knowledge and insights from noisy, structured and unstructured data and applies them to a wide range of applications using scientific methods, procedures, algorithms and systems. Data is classified in two types as structured and non-structured data. Information that can only be examined and understood by humans, such as a picture or the meaning of a block of text, is referred to as "human-readable" information (sometimes referred to as *unstructured data*). Machine-readable information, sometimes called *structured data*, is data that can be processed by computer programmes. Programmes provide data manipulation instructions. We are currently dealing with a situation where we are overrun with data from every angle. The development of technology has aided in the production of an increasing amount of data. For the handling of huge data, it is therefore necessary to have big data scientists who are competent and trained.

2. Big data

Big data is used to describe massive amounts of data that conventional software cannot store or process. According to Douglas Laney, Big data was expanding in three areas: Volume, Velocity and Variety.[1] This qualifies as *three-VS*. Data that is abundantly available shows *volume.* The *velocity* refers to the pace and speed at which data are gathered for subsequent processing. The *Variety* encompasses the gathering of many data sets via text, audio and video.[2] The *Veracity* is determining the data's relevance and correctness and using it for the appropriate reasons. *Value* describes the volume of trustworthy and dependable facts. Understanding and correctly interpreting the meanings of raw data are at the

* Corresponding author: Praveena. S

centre of variability.[2] The sheer volume of data is difficult for conventional software to handle. Big data is becoming more complex, thus we need improved methods and clever web apps for effective analysis to produce insightful results. Big data information and insights will be more valuable if they are stored and analysed using the proper tools. It would be better to implement cutting-edge technology like artificial intelligence and machine learning for economic development and for services like healthcare and safety.

3. Uses of big data in healthcare:

- By identifying early signs and minimising potential negative effects, we can improve therapy effectiveness and quality while also increasing earlier diagnosis.
- By recognising risk variables, illness prevention opportunities are expanded.
- Increasing patient safety and pharmacovigilence by giving them more direct information (IV) Outcome prediction.[3]

The use of Big Data can also assist in identifying and promptly intervening on patients with high risk and high costs.

4. Big data analyst

Big Data analysts used to convert discoveries into useful business plans. Data scientists are people that acquire clean, process, analyses and find new sources in order to build and enhance statistical data mining utilizing cutting-edge technologies.

5. Required skills to become a big data scientist

For data science, **communication** is essential. Data scientists should rely on explaining the facts to people without technical background. The person must be able to interpret data so that decisions can be made in a clear and succinct manner.

Together, **machine learning** and statistics play a crucial role in predicting outcomes since statistics uses sample data to derive generalizations about the population, while machine learning uses Artificial Intelligence (AI). Machine learning tools analyze data sets, present into data-driven algorithms and statistical models to deduce conclusions from patterns found or make predictions based on them. Algorithms learn from the data as they run against it, unlike conventional rules-based analytics systems, which follow explicit instructions. Python and SQL (Structured Query Language) are two **programming languages** that aid in the creation of code for performing numerical and statistical analysis. Data management encompasses **data mining**, which is the process of information extraction.

After data acquisition, the data needs to be cleaned or scrubbed to ensure its accuracy, correctness, consistency, relevance, and purity. A clear and engaging **visualization** of data with charts, heat maps and histograms to illustrate contrasting figures, along with correct labelling to eliminate confusion, can make it easier for us to absorb information and apply it.

6. Public health perspective

Therefore, from a public health perspective, the collection of enormous amounts of data represents an invaluable resource to be used in epidemiological research, analysis of the population's health needs, evaluation of population-based interventions, and informed policy making. Big data and predictive analytics can help precision public health by improving public health surveillance and assessment. While using big data, there are ethical and legal challenges including the risk of compromising privacy and individual autonomy, as well as the effect on public demand for transparency, trust, and fairness.[4]

7. Uses of big data in medical field

Health professionals are responsible for a range of information including patient's medical history (including diagnoses and prescriptions), medical and clinical data (such as imaging and lab analyses) and other personal health information. Previously, such medical records were typically kept either as handwritten notes or as typed reports for a patient.[5] A paper file system was used to store even the results of a medical examination. In fact, the earliest evidence of this practice comes from an Egyptian papyrus text dating back to 1600 BC.[6] Stanley Raiser explains, "Clinical case records describe an episode of illness as a story in which patient, family, and doctor are all involved".[7]

8. Electronic health record

The electronic health record (EHR) is defined by Murphy, Hanken and Waters in 1999 as "a computerized record of an individual's past, present or future physical or mental condition that resides in an electronic system(s) used to obtain, store, retrieve, link, and manipulate multimedia data for the primary purpose of providing healthcare and health-related services". The information contains demographics, clinical narratives, laboratory test results, allergies, prescription information, and medical diagnoses. In this way, medical conditions are recognized and treated more quickly due to a reduction in the lag time of previous tests. Challenges associated with healthcare for big data management and analysis are continuously being developed, primarily for real-time data streaming, capturing, aggregation, analytics (using ML and predictive), and visualization solutions that can help to integrate EMRs (Electronic Medical Records) into healthcare.

9. Responsibilities for a public health dentist

- Surveillance of oral health
- Evaluation of dental interventions, programmes, and services
- Developing and implementing dental policy and strategies
- Leading and collaborating to improve health by improving oral health
- Public health and safety
- To provide and oversee dental services of the highest calibre
- Public health intelligence
- Dental public health in academia
- Provide suggestions to health authorities in making appropriate decisions and judgements on Public health.
- Ethical understanding and legal responsibilities
- Role within the health system to deliver oral care.
- Continuing education for enhancing one's knowledge.[8]

Compiling, categorizing, and tabulating numerical data and expressing the results mathematically is the purpose of statistics. Biostatistics is the only branch of dentistry that deals with them. Within the next 10 years, most of the data science work will be automated. There is a clear need for professionals who can design and implement data-driven solutions. In dentistry, as a part of curriculum, the public health dentists are trained to handle large datasets in research. Hence they can be qualified to perform roles of big data scientists when trained with further advanced software skills.

10. Conclusion

By obtaining the proper training, public health dentists can take up the role of big data scientists, and suggest appropriate paths towards the success of the WHO 2022 theme- *Our Planet, Our Health.*

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] Laney D. 3D Data management: Controlling data volume, velocity and variety. Technical report Meta Group. 2001 Feb: 1-4.

[2] Mauro AD, Greco M, Grimaldi M. A formal definition of big data based on its essential features. Libr Rev. 2016; 65(3):122-35.

[3] Osterreich G. Study on big data in public health, Telemedine and healthcare. European Comission, Directorate-General for Health and Food Safety. Brussels. 2016 Dec.

[4] Ienca M, Ferretti A, Hurst S, et al. Considerations for ethics review of big data health research: a scoping review. PLoS One 2018; 13: e0204937.

[5] Gillum RF. From papyrus to the electronic tablet: a brief history of the clinical medical record with lessons for the digital Age. Am J Med. 2013; 126(10):853–7.

[6] Reiser SJ. The clinical record in medicine part 1: learning from cases. Annals of Internal Medicine. 1991 May 15; 114(10):902-7.

[7] Reisman M. EHRs: the challenge of making electronic data usable and interoperable. Pharm Ther. 2017; 42(9):572–5.

[8] Healthhttps://www.gdc-uk.org › docs › dphcurriculum downloaded on12/06/23.