BIG DATA ANALYTICS

National Collaborating Centre for Infectious Diseases
Centre de collaboration nationale des maladies infectieuses

Big data's fast growth, high volumes and wide variety add complexity, which requires new analytical approaches, as well as massive computational resources with different storage solutions. This summary of big data analytics describes analytical approaches, processes, and uses which can benefit Public Health.

WHAT ARE BIG DATA ANALYTICS?

Big data analytics (BDA) have been defined as the use of analytic techniques to collect, examine, process, and analyze large datasets in preparation for their use in decision-making.[1–4]

BDA can be classified into four main types that address different questions.

WHAT HAPPENED

Descriptive analytics

Use data aggregation and data mining to examine past and present data, helping to understand a current situation, discern patterns and provide alerts.[4–6]

E.g., Know the hospital occupancy rates or the average length of stay.

WHY DID IT HAPPEN

Diagnostic analytics

Use exploration and analysis of the existing data to discover the causes and effects of a situation, helping to understand its nature and impact.[4–6]

E.g., Diagnostic algorithms can correlate symptoms with known diseases to understand why patients are admitted for a specific illness.

Predictive analytics

Use statistical and mathematical methods to forecast and determine future patterns and possibilities based on previously collected data.[4–6]

E.g., Know the expected amount of a drug to stock in anticipation of an outbreak or forecast the spread of a seasonal disease.

Prescriptive analytics

Use modeling structures and analyses, such as machine learning algorithms, to predict outcomes and simulate various approaches, suggesting the best course of action.[4–6]

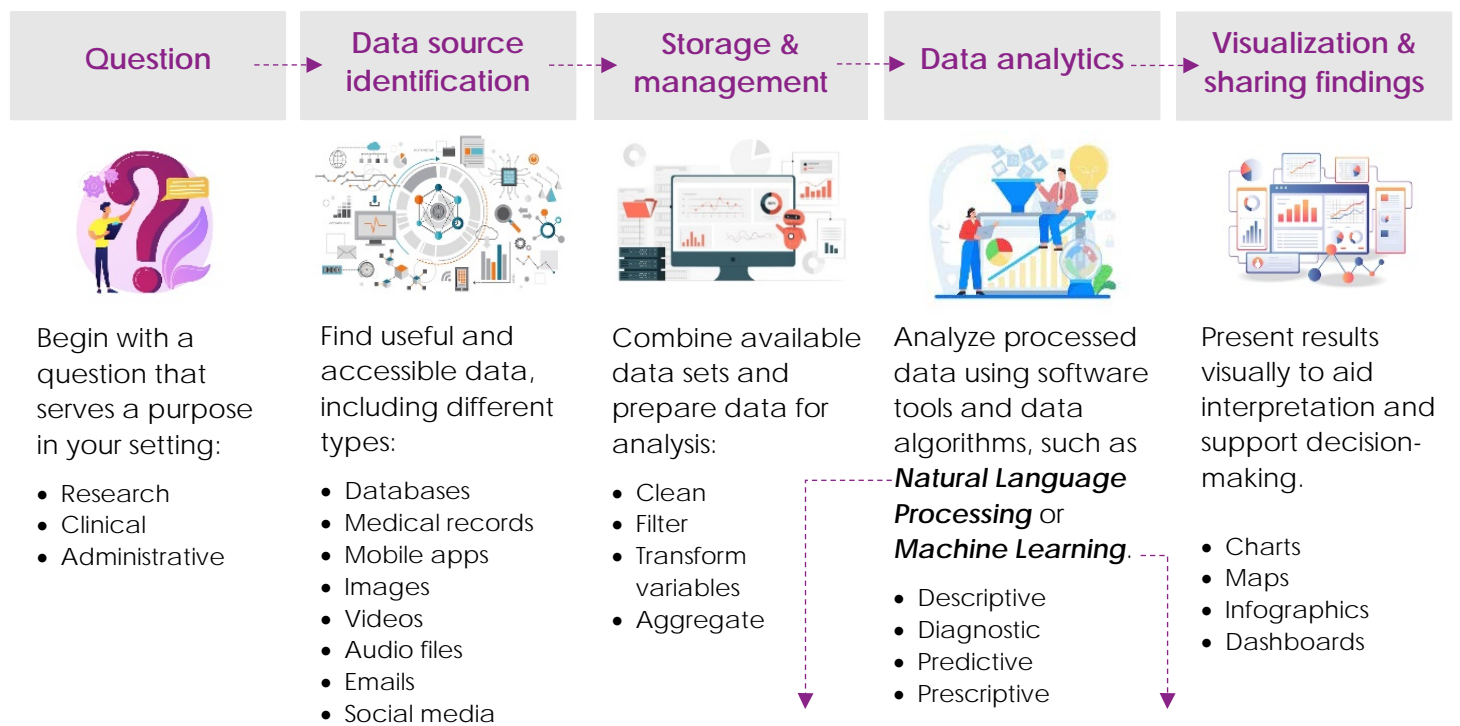**E.g.,** Implement specific preventative treatment plans considering pre-existing conditions and risk factors.

WHAT COULD HAPPEN

WHAT SHOULD WE DO

HOW DO BIG DATA ANALYTICS WORK?

Question	Data source identification	Storage & management	Data analytics	Visualization & sharing findings

Begin with a question that serves a purpose in your setting:

- Research
- Clinical
- Administrative

Find useful and accessible data, including different types:

- Databases
- Medical records
- Mobile apps
- Images
- Videos
- Audio files
- Emails
- Social media

Combine available data sets and prepare data for analysis:

- Clean
- Filter
- Transform variables
- Aggregate

Analyze processed data using software tools and data algorithms, such as *Natural Language Processing* or *Machine Learning*.

- Descriptive
- Diagnostic
- Predictive
- Prescriptive

Present results visually to aid interpretation and support decision-making.

- Charts
- Maps
- Infographics
- Dashboards

Natural Language Processing
Enables machines to process, understand, and interpret human language to perform tasks such as extraction, translation, and classification.[9]

Machine Learning
Uses statistical methods and algorithms to teach machines how to learn and improve performance from experience in order to perform data classification or prediction.[3]

Artificial intelligence (AI)

A computer science innovation that creates intelligent machines and computer programs able to solve problems and scenarios previously addressed by humans.[7]

The most common types of AI are:

HOW ARE BIG DATA ANALYTICS USED?

Toronto SickKids applies BDA to vital-sign data from bedside monitoring devices to identify potential signs of infection as much as 24 hours earlier than previous methods, improving outcomes for infants prone to nosocomial infections.[8]

Massachusetts Institute of Technology researchers reported BDA can predict length of hospital stay, number of patients requiring surgery, and patients at risk for sepsis or iatrogenic diseases—information critical to preventing disease and complications.[2]

North York General Hospital uses a real-time dashboard to track trends in admissions to hospital and intensive care and various resources to respond quickly to new upticks in community transmission.[8]

Johns Hopkins School of Medicine researchers used data from Google Flu Trends to predict sudden increases in flu-related emergency room visits at least a week before warnings from the Centres for Disease Control and Prevention.[8]

REFERENCES

1. Galetsi P, Katsaliaki K. Big data analytics in health: an overview and bibliometric study of research activity. Health Info Libr J. 2019 Mar 1;37(1):5–25.
2. Corsi A, de Souza FF, Pagani RN, Kovaleski JL. Big data analytics as a tool for fighting pandemics: a systematic review of literature. J Ambient Intell Humaniz Comput [Internet]. 2021 Oct 1 [cited 2023 Apr 19];12(10):9163. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7595572/
3. Mooney SJ, Pejaver V. Big data in public health: terminology, machine learning, and privacy. Annu Rev Public Health [Internet]. 2018 Apr 1 [cited 2023 Apr 19]; 39:95–112. Available from: https://pubmed.ncbi.nlm.nih.gov/29261408/
4. Gamage C. Big data analytics: significance, challenges and techniques. Int J Appl Inf Syst [Internet]. 12(29):21-29, May 2020 [cited 2023 Apr 19]. Available from: https://www.ijais.org/archives/volume12/number29/1084-2020451858
5. Rai BK, Meshram AA, Gunasekaran A. Big data in healthcare management: a review of literature. Am J Theor Appl Bus [Internet]. 2018 [cited 2023 Apr 19];4(2):57-69. Available from: https://www.semanticscholar.org/paper/Big-Data-in-Healthcare-Management%3A-AReview-of-Senthilkumar-Rai/fc5747192aa13a30ba70cdba54fd10b65d1abe1b
6. Raghupathi W, Raghupathi V. An overview of health analytics. J Health Med Informat [Internet]. 2013 [cited 2023 Apr24]; 4(3):1-11. Available from: https://doi.org/10.4172/2157-7420.1000132
7. Sun Z, Huo Y. The spectrum of big analytics. J Comput Inform Syst [Internet]. 2019 [cited 2023 Apr 19];61(2):154–62. Available from: https://www.tandfonline.com/doi/abs/10.1080/08874417.2019.1571456
8. Raghupathi W, Raghupathi V. Big data analytics in healthcare: promise and potential. Health Inf Sci Syst [Internet]. 2014 Feb 7 [cited 2023 Apr 19];2(1). Available from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4341817/
9. Baclic O, Tunis M, Young K, Doan C, Swerdfeger H, Schonfeld J. Challenges and opportunities for public health made possible by advances in natural language processing. 2020 [cited 2023 Apr 19];46(6). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7343054/