ISSN: 2710-2599

SvedbergOpen
DISSEMINATION OF KNOWLEDGE

International Journal of Data Science and Big Data Analytics

Publisher's Home Page: https://www.svedbergopen.com/

Research Paper

Open Access

Conceptual, Philosophical, and Epistemological Aspects of Data Analysis

Sidharta Chatterjee[1*] and Mousumi Samanta[2]

[1]Visiting Researcher, School of Economics, Andhra University. E-mail: sidharta123@yahoo.com
[2]Visiting Researcher, Central Library, Jadavpur University. E-mail: mousumisamanta0@gmail.com

Article Info

Volume 4, Issue 1, May 2024
Received : 02 February 2024
Accepted : 22 April 2024
Published : 05 May 2024
doi: 10.51483/IJDSBDA.4.1.2024.52-57

Abstract

This research is an exploration into the realms of epistemology and philosophy of data science and data analysis. It introduces the conceptual aspects of data analysis in data science research. We describe the philosophical implications of data analysis and the role that epistemology and the expertise of analysts play in analysing data, big or small. The process of data analysis is a complex job as it involves advanced statistical tools that play a crucial role in deciphering what the data tells. It also perceives the value of data analysis as a knowledge resource and wealth for organizations that are in the business of data analytics.

Keywords: *Data analysis, Philosophy of data analytics, Statistical tools, Data science*

© 2024 Sidharta Chatterjee and Mousumi Samanta. This is an open access article under the CCBY license (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

1. Introduction

"We can know more than we can tell."

– Michael Polanyi

The process of evaluating data is called data analysis. Data: something that's given. Driven by the modern abundance, prominence, and explosion of data, data science has become an independent, standalone domain of enquiry allowing evaluation, contextual analysis, and assessment of the value of data that transcend different boundaries of science (Desai *et al.*, 2022). Today, we dwell in a data-driven economy where most decisions are made based on evaluation of information obtained from analysis of data. To understand what the data says, one must understand the processes underlying data analysis. How do you make data answer a relevant question? This is accomplished by data scientists—who is someone who uses various tools and techniques to extract meaning from data to solve problems. The thought process that goes into data analysis has much to do with describing the core problems involved in conducting it and the framework that underlines the applications of methods in conducting analysis to derive coherent results (Peng and Matsui, 2015). This paper has a specific aim to achieve, i.e., understanding the "philosophy and the metaphysics" of data analysis

* *Corresponding author: Sidharta Chatterjee, Visiting Researcher, School of Economics, Andhra University. E-mail: sidharta123@yahoo.com*

2710-2599/© 2024. Sidharta Chatterjee and Mousumi Samanta. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

that explains the very details of the entire process. There exists a diverse plethora of frameworks and methods that are employed to analyse data, but we are not delving into them. Rather than that, our goal is simply directed towards describing what data analysis is and characterising the elements of the process. That is, to understand it as an art in order to infuse a creative spark into the analytic machines of data analysis. Really, one may ask, is data analysis art or a science? Our answer is both.

Data analysis (DA) also involves the process of deep learning. This is more relevant in the context of Big data analytics which not only process but work with large volumes of data (Najafabadi *et al.*, 2015), and for the Internet of Things (Saleem and Chisti, 2019). The authors mention that data analytics involve analysing every segment of data to identify trends and extract meaning that becomes information having value to the organisations and businesses who are reliant on data for making decisions. The role that analysts play here is more than just simple analysis of the data. Analysts learn from analysing data and feed the software with relevant data to derive useful insights. To some extent, the intuitions of the analysts do play significant roles that help them build models for the future, i.e., sales forecasting models using past data, marketing metrics, consumer data, etc. Analysts churn this data to derive patterns and insights. Various methods of data analysis exist, which help run tests in order to extract meaning and infuse sense into the data. For example, analysts use "recommender systems," which are entirely based on data analysis and utilise collaborative and content-based filtering (Chow *et al.*, 2013) to generate leads and understand patterns of consumer preferences. Based on such a system, certain recommendations are forwarded to prospective clients. Most often, data analysts strive to do their best to predict future product purchases using past and present data. They generate certain trends and patterns that evolve, which the keen eyes of the interpreter can decode to produce coherent results (Napoletani *et al.*, 2011). In fact, in an informative review paper, the authors (Napoletani *et al.*, 2011) have described the modus operandi of data analysis and the use of diverse data in order to answer the most relevant questions concerning biological, physical, and social phenomena. They stressed the role of mathematics in data analysis and called their approach agnostic science. Data—as we know—is the fact that is given and, based on analysis, the interpretations that one may obtain. Now, it takes time to develop expertise and gain competency in data analysis—similar to other professions. It takes years to grow and become a highly capable data analyst, as it involves application of technical and statistical models. Therefore, technical knowledge adds to the competency of an analyst. Learning the *core elements* of data is a good beginning to understanding the processes of data analysis, as they involve multiple steps. You don't become an expert overnight. Leave that to the machines. But, ironically, the act of writing a data analytic program demands a high level of technical competency and mastery in software programming.

2. Organization of Data and Data Structures

Numbers are far easier to read but more difficult to remember than words (Erickson and Nosanchuk, 1992). Most data consists of numbers. They are hard to retain and present as incoherent and ambiguous patterns—patterns that need to be unfolded for understanding following rigorous analysis of the data. Hence, even if you try very hard to just read and extract meaning from it, without applying the tools of analysis, you won't get too far. For the reason that the human brain is usually trained to retain and comprehend written words in sentences and phrases more than numbers, save mathematics. But that too—mathematics—is taught and understood in a systematic process, in a in a methodical manner, and through ordered reasoning. Mathematics uses the tools of reason and logic in a much similar way that data analysts use certain tools for routine data analysis. Similarly, analysts must be well skilled in the methods used to arrange and organize data to fulfil the criteria of analysis. This issue has been highlighted by several researchers where a competency model is necessary to close the skills gap in order to improve quality and competitiveness of the data analysts and the data science workforce (Hattingh *et al.*, 2019).

According to Carr (1985), any set of data must represent a problem, either a general problem area or a specific phenomenon of interest. That is to say, for example, consider a scenario in which you are given to analyse a set of data from several local McDonald outlets about the time taken for each customer on the queue to receive an order placed, which is a specific problem, against the average time a customer has to wait for her turn at all McDonald outlets across the country. Thus, it is important to understand what kind of problem the data reflects. Secondly, what variables and observations are to be brought forward for analysis? Thirdly, a "model" must be specified that underlies the statistical techniques to be employed to analyse a particular set of data. Here, some degree of judgement is indispensable from the part of the analyst concerning the model to be employed that'll fit the data, or the data that'll fit the model thus chosen. Therefore, getting into the data is a tricky job for a data analyst to extract the meanings assigned to the analytically derived representation of data (Carr, 1985).

Basic data analysis is easy and fun. One can learn a lot from analyzing numbers (data). Just by using judgement and rationality, one can go far in numerical analysis of data using statistical tools as aids to processing numbers. This yields hidden patterns, and such patterns can be interpreted into information and further into knowledge. Analysis of data provides clues and cues, hints, and configurations in numerical form. You'll need a keen eye to detect such clues and understand patterns so that you can inform and give meaning to your findings (results of data analysis). But before you do anything else, you must be able to handle data and understand its structure and format. It needs background knowledge.

For example, we are aware of how newspaper headlines are used to catch readers' attention. In the words of the noted philosopher *Alfred North Whitehead*[1], "Newspaper headlines are billboards to sell an article." Now, if we take the statement that seems like a metaphor seriously, of course we could derive meaning from it. Newspaper headlines are deliberately made beguiling through use of catchphrases and larger bold fonts. Boldness attracts attention…these create sensations in the minds of the readers. Consider doing a newsgathering task to obtain data from them, i.e., skimming the headlines of several notable newspapers, like for example, Times London, Times of India, or Boston Herald, Washington Post. Make note of the impacts that they have made over the years, say a ten-year period, the nature of content and subjects of such news headlines, word length of each of the frontline (FrontPage) articles, while gathering data from them with these variables in mind. You will get data, of course, and much more, which you can structure and give shape to them for allowing further analysis. It must be remembered: reality is rich data. And the world of things are represented in reality. But it is also true going by the thoughts of Descartes[2], that everything is not precise in the real world of existence. Therefore, the goal of data analysis—to a large extent—is a scientific one: That is, to give meaning to incoherent nature of observed points of references we call data, and prove by methods of analysis the certainty of things in existence. Certainty is striking in mathematics—hence data analysis utilises mathematical and statistical sophistication (using models) to infer truth from data. Data and facts tell us about something that exists or has happened. Now, the Philosophy of Descartes tells us that, something that can be "counted" and obtained as data or facts must exist. Only its true nature is revealed through analysis of the data and facts.

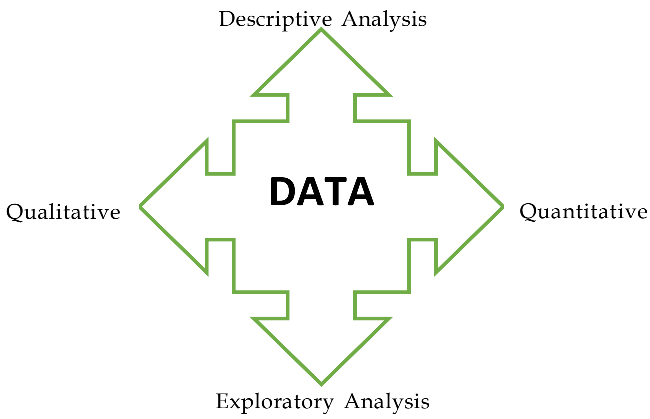

Figure 1: Components of Data Analysis

3. Epistemology and Etymology of Data Analysis

Figure 1 above depicts the four fundamental typologies of data analysis. Getting the truth out of data is of primary concern for any data analyst. What is the purpose of data analysis? To get the meaning out of it and understand what the data informs or tells us about. There are conceptual aspects of data analysis that a data analyst must be aware of if she is to get something of value from the data. Today, data is a highly valued

1 See, Whitehead, A.N. and Price, L. (2001). Dialogues of Alfred North Whitehead (Vol. 84). David R. Godine Publisher.
2 René, D. (1901). Discourse on Method: Meditations on the First Philosophy. *Principles of Philosophy*. Forgotten Books.

commodity, for it is not only difficult to obtain but also costly to produce. Data is obtained from various sources, e.g., experimental data, data obtained from empirical analysis, observation, exploratory analysis, scientific data from questionnaires and surveys, consumer data and data obtained from archaeological explorations, medical and healthcare data, patient information systems, hospital records, institutional and financial data from financial bourses, data from crime bureaus, economic data, the Bureau of National Statistics, census data, and market data mostly obtained from marketing surveys and customer interactions.

There may be various other sources of data, e.g., business intelligence data, market metrics, marketing intelligence, sales data, corporate and legal data, etc. All these means of procuring data indicate one thing: the value of data today as a commodity is well established. In today's data-driven knowledge organisations, data has become one of the most essential capital resource/asset for them. Enterprise systems using latest information technology tools are heavily reliant on data and the outcomes of data analysis, as they collect, analyse, store and curate data to aid management decision-making (Wang *et al.*, 2022). Data in organisations have taken the centre stage and are utilised by enterprise systems. Data is generated from various enterprise and organisational/business activities, such as through production, manufacturing, servicing, work logs, consumer visits to e-commerce sites, order placement, sales, inventory, transactions, among others as mentioned by the authors (Appelbaum *et al.*, 2017; Wang *et al.*, 2022). The question is, how do you get good data? How do we classify and categorise data? And where lies the creativity in data analysis? Although data analysis is mostly a monotonous and mechanistic job, there is some degree of art and fineness of creativity attached to the entire process. The analyst must give their data a form and shape for understanding what it tells. They often go on at great length to interpret data using advanced statistical tools of analysis. Of course, the nature of the data determines the models to be applied and the methods to be adopted for the analysis of the data.

When we are presented with some data, the first thing we should be doing is to understand the "perspective" of the data, i.e., whether they could be related to one another possessing characteristics, for example, a linear perspective, among others. To derive wisdom from a set of data, we must pursue it from different angles, which will provide diverse vantage points. Indeed, 'we can know more than we can tell' or do with data. But in data analysis, we should be able to do certain things with the data itself without cooking it to falsify results. That is, in order to be able to explain how rigorous analysis could be done using robust models in a step-by-step manner so that others could carry out similar analysis. This is where one puts 'tacit knowledge' in it, and thus can definitely instill creativity in analysing and interpreting data. All these functions are performed to get the maximum "meaning" out of data.

4. Conceptual Aspects of Data Analysis

Analysing data results in the production of value. Data contains information that can be decoded using the tools of data analysis and the human mind. The form and character of the analytic activity depend on the nature of the data. Analysis involves the expenditure of energy and effort. It is a productive activity that aims to inform through inquiry and investigation. There are certain conceptual aspects of the data analysis process open to anyone for learning that become apparent when working with data, big or small. It is mostly obvious to those who have attained a high degree of expertise and supremacy in data analysis. Advantages and pitfalls exist—for these are attached to working with big or small data samples. The information obtained following data analysis often becomes "knowledge" after testing and validation are done using appropriate theoretical models to understand what the data says. As for any analyst, he or she must 'know' how to analyse data with maximum effectiveness to extract maximum "meaning" out of them. This is a big challenge for most data analysts, as it appeals to the creative domains of the intellect to instill a certain degree of fineness and creativity into the process. To gain meaningful insights into data, analysts most often apply their intuitions as well.

4.1. Consider Economics

There are many theoretical models that continue to be used despite the paucity of data for them (Brooks, 2013). But, then again, data shows many trends, refutes many established beliefs, offsets opinions, and replaces them with reason on the grounds of facts. Data analysis can illuminate many patterns of behaviour that have hitherto remained undetected to the naked eye. However, in all such instances, correct interpretation of data is a must to avoid errors that demand more skilled analysts taking care of raw data and handling them efficiently. One objective for the analyst is to learn how to analyse effectively with fewer data points. This involves some degree of acumen, skills, and knowledge of the models and methods to be adopted for analysing data. It is a learning process, as we learn continuously from analysing data. The teaching of "how" to do it is often tacit (Polanyi Michael, 1966). When the sample size is small, analysts need to apply the most effective methods that fit the data in order to derive meaningful results. Here, it requires some degree of creativity on the part of the analyst to undertake the analyses and present results that would be more "acceptable" to them as well as to the decision-makers who rely on the results obtained from the analysis of data to make informed decisions. Acceptable because it is able to inform, but it may or may not be visually appealing. One has to understand the spirit in which an analysis is made. The power of methods and the impact of models on data tell a lot about the approach that has been taken to analyse the sample data. The visual effects of charting and graphical representation of the results of data analysis are for the general appeal and ease of grasping what the data says. The information that is obtained thus becomes "intellectual wealth" for the analysts or the organisation undertaking data analysis. The result is the objective output from analyses that increases the visual acumen of the decision-maker and further augments her with the much-needed information to make decisions that matter. This enforces the role of management energetics in data analysis, which depends on the power, ability, and expertise of the analysts working with data.

5. Conclusion

No amount of data can fully and truly represent the social reality of our existence. But data contains information that speaks a lot about our activities, behaviours, choices, and preferences. The analysis of data lets us arrive at a sense of truth; it informs us and enables us to make informed decisions. Analysis also lets us understand the relationship between data. Scientific data, when it is completely accessible to us and represents the true reality of the occurrences of events and phenomena under controlled conditions, is the data gathered following astute and repeated observations. It is more precise than those originating from social experiments. This makes scientific truth more accurately and completely accessible to us. But if something is not known correctly and completely, that is, with absolute certainty, how can we claim its nature to be true in order to be able to describe and differentiate it from others? Of course, the data here fills the gap, as it helps describe "differences" in reality. Therefore, it is important to understand what the data tells us. We cannot be silent about interpretations of data, as they must tell something about something. And when it reveals something of value, it becomes a commodity of commerce.

Acknowledgment

The authors acknowledge the kind assistance of the Central Library, Jadavpur University.

References

Appelbaum, D., Kogan, A., Vasarhelyi, M. and Yan, Z. (2017). Impact of Business Analytics and Enterprise Systems on Managerial Accounting. *International Journal of Accounting Information Systems*, 25, 29-44.

Brooks, D. (2013). The Philosophy of Data. *New York Times*, 4(02).

Carr, C. (1985). Getting into Data: Philosophy and Tactics for the Analysis of Complex Data Structures. *For Concordance in Archaeological Analysis*, 18-44.

Chow, R., Jin, H., Knijnenburg, B. and Saldamli, G. (2013, October). Differential Aata Analysis for Recommender Systems. In Proceedings of the 7th ACM Conference on Recommender Systems (pp. 323-326).

Desai, J., Watson, D., Wang, V. *et al.* (2022). The Epistemological Foundations of Data Science: A Critical Review. *Synthese*, 200, 469. https://doi.org/10.1007/s11229-022-03933-2.

Erickson, B. and Nosanchuk, T. (1992). Understanding Data. McGraw-Hill Education (UK).

Hattingh, M., Marshall, L., Holmner, M. and Naidoo, R. (2019). Data Science Competency in Organisations: A Systematic Review and Unified Model. Proceedings of the South African Institute of Computer Scientists and Information Technologists 2019, 1-8.

Najafabadi, M.M., Villanustre, F., Khoshgoftaar, T.M., Seliya, N., Wald, R. and Muharemagic, E. (2015). Deep Learning Applications and Challenges in Big Data Analytics. *Journal of Big Data*, 2, 1-21.

Napoletani, D., Panza, M. and Struppa, D.C. (2011). Agnostic Science. *Towards a Philosophy of Data Analysis*. Foundations of Science, 16, 1-20.

Peng, R.D. and Matsui, E. (2015). The Art of Data Science: A Guide for Anyone Who Works with Data. Skybrude Consulting, LLC.

Polanyi, Michael (1966). The Tacit Dimension, The University of Chicago Press, Ltd., London.

René, D. (1901). Discourse on Method: Meditations on the First Philosophy; Principles of Philosophy. Forgotten Books.

Saleem, T.J. and Chishti, M.A. (2019). Deep Learning for Internet of Things Data Analytics. *Procedia Computer Science*, 163, 381-390.

Whitehead, A.N. and Price, L. (2001). Dialogues of Alfred North Whitehead (Vol. 84). David R. Godine Publisher.

Yu Chung Wang, W., Pauleen, D. and Taskin, N. (2022). Enterprise Systems, Emerging Technologies, and the Data-Driven Knowledge Organisation. *Knowledge Management Research & Practice*, 20(1), 1-13.

Cite this article as: Sidharta Chatterjee and Mousumi Samanta (2024). Conceptual, Philosophical, and Epistemological Aspects of Data Analysis. *International Journal of Data Science and Big Data Analytics,* 4(1), 52-57. doi: 10.51483/IJDSBDA.4.1.2024.52-57.