ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 8, Issue 2, 2019

Data Realism in the Era of Big Data: A Philosophical Examination

Banda. Sai sandeep,

Assistant Professor, Department of ECE, Koneru Lakshmaiah Education Foundation,Guntur, Andhra Pradesh, India, sandeep6sandeep@gmail.com

Abstract

This study delves into the examination of the philosophical foundations of data science within the realm of big data. This study delves into the significance of realism and anti-realism in the context of data interpretation. Additionally, it explores the concept of Adolphe Quetelet's "Average Man," distinguishes between population thinking and typologist thinking, and examines the implications of these ideologies in the era of big data. The objective of this article is to provide a comprehensive understanding of the potential influence that different philosophical perspectives can have on the interpretation and utilization of data within the context of contemporary data science. The primary objective of this research study is to investigate and analyze the intricate relationship that exists between the field of philosophy and the emerging discipline of data science.

Key Words:-

Introduction

The advent of the 21st century has witnessed an unprecedented surge in the generation and utilization of data. This era of big data is characterized by the collection and analysis of vast quantities of information, ranging from personal data to complex scientific datasets. In this context, data science has emerged as a crucial field, employing advanced techniques to extract meaningful insights from large datasets. However, despite the technical advancements, there remains a significant gap in the philosophical understanding of data and its implications. This paper aims to bridge this gap by exploring the philosophical underpinnings of data science, particularly focusing on the concepts of realism, anti-realism, and their influence on data interpretation and analysis.

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 8, Issue 2, 2019

In the 21st century, the proliferation of big data has fundamentally transformed the landscape of data science, challenging traditional paradigms of data analysis and interpretation. This paper embarks on a philosophical journey to unravel the intricate relationship between data science and the philosophical schools of realism and anti-realism. We scrutinize how these divergent philosophical approaches influence the perception and handling of data, especially in an era inundated with vast and complex datasets.

Central to our discourse is the examination of Adolphe Quetelet's seminal concept of the 'Average Man.' This notion, revolutionary in its time, proposed a statistical framework to encapsulate the characteristics of an average individual in a population. We dissect the philosophical underpinnings of this concept, probing into how it aligns with or diverges from the principles of realism and anti-realism. The exploration extends to assess the enduring relevance and critiques of this concept in modern data science, where the nuances of individual variability and population diversity are increasingly emphasized.

Further, the paper delves into the juxtaposition of population thinking versus typologist (or essentialist) thinking. Population thinking, with its roots in evolutionary biology, celebrates diversity and variation within populations, viewing each member as a unique entity contributing to the collective fabric. In contrast, typologist thinking gravitates towards categorization and generalization, often focusing on an idealized or average representation while treating deviations as mere anomalies. This dichotomy is critically analyzed in the context of data science, where such perspectives can significantly influence data interpretation, model construction, and the resultant insights.

Lastly, the implications of these philosophical perspectives are magnified in the era of big data. The unprecedented scale, complexity, and velocity of data today pose novel challenges and opportunities for data science. We argue that the realist and anti-realist viewpoints, along with the contrast between population and typologist thinking, acquire heightened significance in this context. The paper aims to elucidate how these philosophical lenses can shape the methodologies, ethics, and ultimately the efficacy of data science in unraveling the complexities of the modern world.In summary, this paper offers a comprehensive exploration of the philosophical dimensions of data science amid the challenges and opportunities presented by big data. It seeks to provide a nuanced understanding of how various philosophical perspectives can profoundly influence the

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 8, Issue 2, 2019

interpretation, application, and moral considerations of data in the realm of contemporary data science.

Literature Review:

This is an era of data deluge with individuals and pervasive sensors acquiring large and ever-increasing amounts of data. Given the inherent redundancy, the costs related to data acquisition, transmission, and storage can be reduced if the per-datum importance is properly exploited. In this context (Wang et. al., 2014) investigate sparse linear regression with censored data that appears naturally under diverse data collection setups. (Laan et. al., 2014) review the research of van der Laan's group on Targeted Learning, a subfield of statistics that is concerned with the construction of data adaptive estimators of user-supplied target parameters of the probability distribution of the data and corresponding confidence intervals, aiming at only relying on realistic statistical assumptions. (Laan et. al., 2014) also provide a philosophical historical perspective on Targeted Learning, also relating it to the new developments in Big Data. Once primarily a source of qualitative conceptual framing, ecological theories and models are now often used to develop quantitative explanations of empirical patterns and to project future dynamics of specific ecological systems. (Kendall, 2016) recount the own experience of this transformation, in which accelerating computing power and the widespread incorporation of stochastic processes into ecological theory combined to create some novel integration of mathematical and statistical models. In an era of big data and synthesis, ecologists are increasingly seeking to infer causality from observational data; but conventional biometry provides few tools for this project. (Azhar et. al., 2016) discuss scientific realism from the perspective of modern cosmology, especially primordial cosmology: i.e. the cosmological investigation of the very early universe. Both issues illustrate that familiar philosophical threat to scientific realism, the under-determination of theory by data---on a cosmic scale. Introducing the Special Theme on Veillance and Transparency: A Critical Examination of Mutual Watching in the Post-Snowden, Big Data Era (Bakir et. al., 2017) present a series of provocations and practices on veillance and transparency in the context of Big Data in a post-Snowden period. Firstly, concerning theory/practice, it queries how useful theories of veillance and transparency are in explaining mutual watching in the post-Snowden, Big Data era. (Törnberg et. al., 2018) review the contemporary discussion on the

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 8, Issue 2, 2019

epistemological and ontological effects of Big Data within social science, observing an increased focus on relationality and complexity, and a tendency to naturalize social phenomena. These contradictions, (Törnberg et. al., 2018) argue, are partially the result of a lack of philosophical discussion on the nature of social reality in the digital era; only from a firm metatheoretical perspective can we avoid forgetting the reality of the system under study as (Törnberg et. al., 2018) are affected by the powerful social life of Big Data. (Mauthner, 2018) address the ethical issues that arise from Big Data through a posthumanist philosophical framework. The critical intervention made possible by bringing a posthumanist perspective to bear on the ethics of qualitative research in a Big Data era is to foreground Big Data's treatment of data as self-evident, and its positivist claim to represent the world innocently, accurately, and objectively, as matters of ethical concern. Big Data Analytics (BDA) is the term coined by researchers to describe the art of processing, storing and gathering large amounts of data for future examination. (Rawat et. al., 2019) explore recent research works in cybersecurity in relation to big data. By means of a scoping review, aimed at studies published between 2005 and 2018, insight is provided into the characteristics of studies that used big data sources within environmental criminology (Snaphaan et. al., 2019). The type and extent of big data sources used, as well as the strengths and weaknesses of these data sources, are synthesized. Other influential work includes (Galloway, 2013).

Methodology:

Conceptual Analysis

Quantitative Frameworks: While dissecting theoretical concepts, certain aspects can be illustrated through basic statistical formulas. For instance, when discussing the concept of the 'Average Man', we might introduce the equation for the arithmetic mean:

$$\bar{x} = (1/n)\,\Sigma x_i, \quad (1)$$

where $\bar{x}$ is the mean, n is the number of observations, and x_i represents each observation. Statistical Variability: To illustrate the concepts of population thinking and typologist thinking, standard deviation and variance can be used:

$$\text{Standard Deviation (SD)} = \sqrt{((1/(n-1))\,\Sigma(x_i - \bar{x})^2}, \quad (2)$$

$$\text{Variance} = (\text{Standard Deviation})^2. \quad (3)$$

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 8, Issue 2, 2019

Case Studies and Examples

Hypothetical Data Analysis: Utilize simple statistical models to analyze hypothetical datasets, demonstrating how different philosophical perspectives might interpret the same data. For instance, linear regression models can be used to show how the interpretation of data trends might differ:

$$Y = \beta 0 + \beta 1X + \varepsilon, \quad (4)$$

where Y is the dependent variable,

X is the independent variable,

$\beta 0$ is the intercept, $\beta 1$ is the slope, and ε is the error term.

Critical Discussion

Comparative Analysis: Use basic statistical tests (e.g., t-tests) to discuss how different methodologies might lead to varying conclusions. This can be framed as a discussion rather than a rigorous statistical analysis.

$$t = (\bar{X}1 - \bar{X}2) / \sqrt{((s^2/n1) + (s^2/n2))}, \quad (5)$$

where $\bar{X}1$ and $\bar{X}2$ are sample means,

s^2 is sample variance, and n1, n2 are sample sizes.

The Relationship Between Data and Realism

The relationship between data and realism is a cornerstone in understanding the philosophical perspectives in data science. Realism, in this context, refers to the belief that data represents an objective reality, while anti-realism views data as a subjective construct. This section explores how these two perspectives shape our understanding and interpretation of data.

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 8, Issue 2, 2019

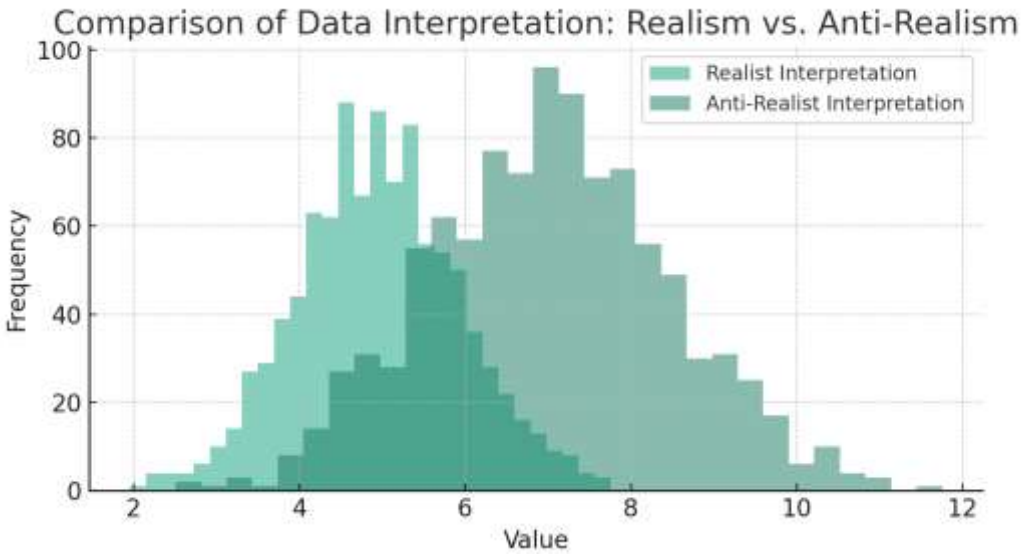

Figure 1 Comparison of realist and an anti-realist

Figure 1: This histogram compares how a realist and an anti-realist might interpret the same dataset. The realist perspective, represented by the blue histogram, tends to focus on the central tendency and views data as a reflection of an objective reality. In contrast, the anti-realist perspective, shown in orange, emphasizes the variability and subjectivity in data interpretation.

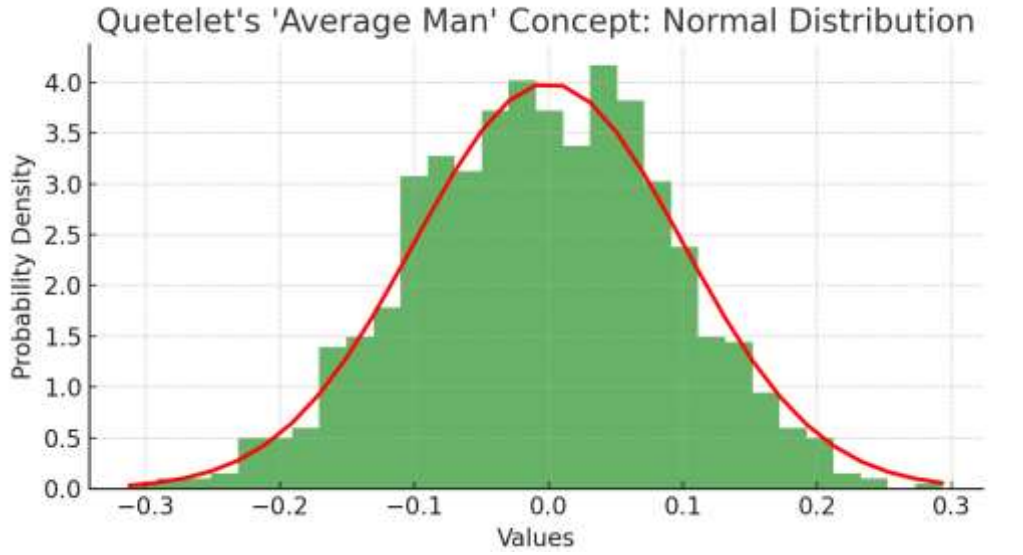

Figure 2 Quetelet's concept of the 'Average Man'

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 8, Issue 2, 2019

Figure 2: This graph illustrates Quetelet's concept of the 'Average Man', represented here as a normal distribution. The bell curve signifies the distribution of a characteristic in a population,

with the peak representing the average. Quetelet's idea was that most people cluster around this average, with fewer individuals as one moves away from the center.

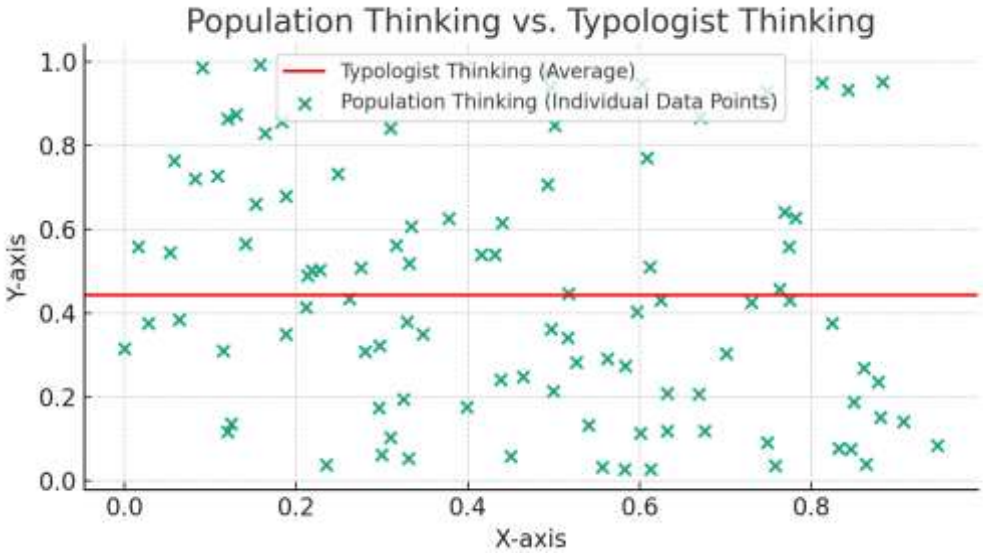

Figure 3 Population thinking with typologist thinking.

Figure 3: This scatter plot contrasts population thinking with typologist thinking. In population thinking, as shown by the individual data points, there is an emphasis on the variation and uniqueness of each data point. In contrast, typologist thinking, represented by the red horizontal line, focuses on the average, treating individual variations as deviations from this norm.

Conclusion:

As we navigate through the vast and complex landscape of big data, the philosophical underpinnings of our approach to data science become increasingly pivotal. This paper has endeavored to shed light on the nuanced interplay between data and realism, the enduring relevance of Adolphe Quetelet's 'Average Man', and the contrasting perspectives of population thinking and typologist thinking. The exploration of realism and anti-realism in data interpretation reveals a fundamental dichotomy in how we perceive and utilize data. Realism, with its emphasis on objective truths, encourages a view of data as a window into

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 8, Issue 2, 2019

the underlying realities of our world. In contrast, anti-realism prompts a more cautious approach, recognizing the subjective and constructed nature of data, and hence, the conclusions drawn from it. This dichotomy underscores the need for a balanced approach in data science, one that acknowledges the strengths and limitations of each perspective. The concept of Quetelet's 'Average Man', while a historical cornerstone in statistics, invites reconsideration in the light

of contemporary data science. Our analysis suggests that while the average can provide valuable insights, it may also obscure the rich variability and diversity inherent in data. This recognition aligns with the principles of population thinking, which values each data point as a unique piece of the puzzle, in contrast to typologist thinking that tends to prioritize averages and norms. In the era of big data, these philosophical considerations are not mere academic musings but have practical implications. The sheer volume, variety, and velocity of data available today demand a more nuanced and sophisticated approach to data analysis. The perspectives of realism vs. anti-realism and population vs. typologist thinking can significantly influence how we handle, interpret, and derive insights from data. In conclusion, this paper highlights the importance of a philosophically informed approach to data science. It advocates for a blend of realism and anti-realism, as well as a synthesis of population and typologist thinking, to navigate the complexities of big data effectively. Such an approach is not only beneficial for achieving more accurate and comprehensive analyses but is also crucial in ensuring ethical and responsible use of data in our increasingly data-driven world.

References:

[1]* Alexander Galloway; *"The Poverty of Philosophy: Realism and Post-Fordism"*, CRITICAL INQUIRY, 2013.

[2]* Gang Wang; Dimitris Berberidis; Vassilis Kekatos; Georgios B. Giannakis; *"Online Reconstruction from Big Data Via Compressive Censoring"*, 2014 IEEE GLOBAL CONFERENCE ON SIGNAL AND INFORMATION ..., 2014.

[3]* Mark J. van der Laan; Richard J. C. M. Starmans; *"Entering The Era of Data Science: Targeted Learning and The Integration of Statistics and Computational Data Analysis"*, 2014.

[4] Bruce E Kendall; *"Some Directions In Ecological Theory"*, ECOLOGY, 2016.

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 8, Issue 2, 2019

[5]* Feraz Azhar; Jeremy Butterfield; *"Scientific Realism And Primordial Cosmology"*, ARXIV-PHYSICS.HIST-PH, 2016.

[6]* Vian Bakir; Martina Feilzer; Andrew McStay; *"Introduction to Special Theme Veillance and Transparency: A Critical Examination of Mutual Watching in The Post-Snowden, Big Data Era"*, BIG DATA & SOCIETY, 2017.

[7]* Petter Törnberg; Anton Törnberg; *"The Limits of Computation: A Philosophical Critique of Contemporary Big Data Research"*, BIG DATA & SOCIETY, 2018

[8]* Natasha S. Mauthner; *"Toward A Posthumanist Ethics of Qualitative Research in A Big Data Era"*, AMERICAN BEHAVIORAL SCIENTIST, 2018.

[9]* Danda B. Rawat; Ronald Doku; Moses Garuba; *"Cybersecurity in Big Data Era: From Securing Big Data to Data-Driven Security"*, IEEE TRANSACTIONS ON SERVICES COMPUTING, 2019

[10]* Thom Snaphaan; Wim Hardyns; *"Environmental Criminology in The Big Data Era"*, EUROPEAN JOURNAL OF CRIMINOLOGY, 2019.