Data Science Strategy: Exploring the Patterns and Rules in the Data-Intensive Discipline

Geetanjali Tyagi and Susmita Ray

July 7, 2022

DATA SCIENCE STRATEGY: EXPLORING THE PATTERNS AND RULES IN THE DATA-INTENSIVE DISCIPLINE

Geetanjali Tyagi1[0000-0003-3715-3695] and Dr.Susmita Ray2[0000-0002-6073-6509]

[1]Phd Scholar, Dept.of CST,Manav Rachna University Faridabad, India

Email: Git101288@gmail.com

[2]Professor, Dept.of CST, Manav Rachna University Faridabad, India

Email: susmita@mru.edu.in

Abstract. Several researchers have looked at the difficulties associated with adopting different components of Data Science in a variety of business sectors and specialties. An abundance of research has been conducted in the past on the popular acceptance of data science in its entirety as a phrase that encompasses all of its constituent parts. Using a detailed examination of this literature based on the comprehensive Data Science along with the enablers and obstacles to data technology strategies in agencies, this research explores patterns and rules in data science along with evaluating the deployment and relevance of data science strategies in various organizations.

Keywords: Data science, digital transformations, digitalization, digitization, datanature, patterns, rules.

1. Introduction

Businesses across industries and regions are facing a digital transformation. Both conventional and creative firm models must harness digital technology in order to address existential concerns and realize game-changing possibilities. The digitization revolution escalated as a result of the Covid-19 outbreak, which drove firms to review their business practices. Large-scale digital transformations are taking place as a result of the mass availability of digital technology and Internet access, allowing organizations to rethink their business strategies Wang et al.[1]. The digital transformation process is split into two stages: digitization and digitalization Verhoef et al.[2]. The process of digitization involves converting analogue data into digital format. Using digital technology to redesign company structures in order to take advantage of digital business prospects is what is meant when using the term "digitalization.".

Data science has made considerable research expenditures in advanced analytics, data model improvement, and algorithm design. However, few writers have confronted the organizational and socio-technical issues inherent in running a data science project: These include a lack of vision and defined goals, an overemphasis on technical concerns, a low degree of maturity for ad-hoc initiatives, and the uncertainty of responsibilities in data science. In the literature, there are only a few techniques to dealing with these sorts of challenges; some date back to the mid-1990s and, as a result, are out of date with the current paradigm and recent breakthroughs in big data and machine learning technology.

Additionally, fewer models offer a comprehensive foundation for managing people, projects, and data and information. In this post, the researcher will analyze the necessity of taking a more holistic approach to data science operations. The study begins by examining approaches for working on data science projects that have been published in the literature and categorizing them in line with their attention: assignment, crew, facts, and information management. Lastly, the study presents a conceptual framework that describes the important criteria that a methodology for holistically managing data science projects should have. Other researchers might utilize this framework as a guide for constructing new data science approaches or adjusting existent ones.

1.1 Background of the Study

The data explosion is the quick increase of data in cyberspace, propelling humanity into the era of big data. The meaning of data has varied over time. Scientific data are no longer bound by qualitative or quantitative variables, measurement results, or scientific observations and experiments. Additionally, data comprises everything found on the internet. Unconsciously, datanature (all data in cyberspace) develops and changes Tang[3]. It is becoming increasingly common for data-nature to create data that has no natural analogue, such as virus attacks, online games, and junk data. The information generated by datanature has gradually surpassed the facts inherent in nature and evolved its own patterns.

Experts have been utilizing and dealing with data continually since the creation of the computer. The natural world's facts are mapped as data and archived in computers for future use. However, the way data is utilized has grown from simple data access to huge data analysis, especially in the area of research (e.g., life science). This provides new demands and barriers for data technology, leading in data-related study, such as how to examine life using DNA data. The reason for which data is utilized is also changing.

A data analyst does not only address real-world difficulties; he or she also researches the data in order to determine the phenomena and laws behind the data (e.g., identifying the growth patterns and predicting the scale of cyberspace data ten years in the future). Data technologies and techniques can be applied to natural and social sciences, and researchers can and should investigate datanature, to help speed up the transition to data science. We are living in the age of data science, whether we wish to acknowledge it or not; whether we want it or not. Data scientists may have already accomplished the position if they have been doing data science research.

1.2 Problem Statement

As a first step in determining the present status of data science methodologies, the authors of this research have opted to perform a complete evaluation of the literature in this field. It gives an overview and review of the most recent research conducted on the issue, as well as recommendations. Acknowledgements are offered in honor of those individuals who have made the most important contributions to the field of study. This results in the aims and objectives of the research being placed in their proper perspective, as well as the discovery of previously unknown knowledge gaps. Because knowledge about the application of data science methodology is dispersed across many sources, such as scientific journals and books, as well as blogs, white papers, and open internet platforms, conducting a critical assessment rather than a systematic study is a sound business decision in the long run. Informal sources of information were necessary for gaining a grasp of the point of view of real data science activities in order to get a better understanding of their viewpoint.

1.3 Research Aim and Objectives

Specifically, the goal of this study is to compile a list of the current features of data science techniques that are acceptable for use in big data analytics, as well as to identify the essential qualities that impact the effectiveness of deep learning approaches in big data analytics. The following objectives may assist in achieving this goal:

- To establish a field of data science study, in addition to the volume and kind of research and discoveries that take place inside it.
- To investigate the features that have an influence on data science approaches for the purpose of optimizing the performance of big data analytics.
- To carry out study on the patterns and rules that under pin data science strategies.

2. Data Science

In general, researchers who develop data science from well-established fields agree on the domains that feed and grow data Examples of such definitions include Agarwal and Dhar [4], who define data science as the opposite of computer science and business engineering; statistics; data mining and machine learning; job research; six sigma; automation and domain technology; and Gibert et al.[5], who define data science as a multidisciplinary intersection of mathematical expertise with business acumen and hacking skills.

A range of skills, ranging from traditional computer science to art and design, are required to be successful in data science, according to Lv et al. [6]. A Venn diagram depicting data science is provided by Semeler, Pinto, and Rozados (2019), in which the synthesis of hacking talents, math and statistics skills, as well as domain expertise, is demonstrated to be the case. According to the author Suthaharan[7], many computational difficulties are statistical engineering problems requiring distributed computing and machine learning methods in addition to statistical modeling.

Although few authors have examined the fundamental goal of data science, doing so is essential for understanding data science's role in business and industry, as well as its many different domain applications, among other things. Analyzing data in order to resolve issues and explore possibilities is defined by Cao[8] whereas Martinezet al. [9] define data science as "the application of statistical and machine learning techniques on large multi-structured data in a distributed computing environment in order to identify correlations and causal relationships, classify and predict events, identify patterns and anomalies, and infer probabilities, interest, and sentiment" Data science, in their opinion, is a combination of talents in software development, data management, and statistical analysis, among other disciplines. According to Strukova [10]data science is the study of the computational ideas, strategies, and systems that are used for extracting and organising information from large amounts of information.

Additional interest has risen in describing the job of data scientists as well as articulating the key skills required to become a data scientist, with the goal of gaining a clearer understanding of the role of data scientists relative to traditional employment titles. Data scientists, according to McQuillan[11], are "better at statistics than any software engineer and better at software engineering than any statistician," while Van der Aalst (2016) describes them as "unicorns," capable of handling everything from data collection to data processing at scale, visualization, and writing up as a narrative.

3. Digital Transformation and Data Science: Interrelation

Organizations must have a unified and carefully designed digital transformation strategy supported by technology and customer interaction in order to effectively execute digital transformation activities and define future products and services, according to the report Varshney[12]. This approach necessitates the use of digital capabilities to ensure its success. Any digital strategy should incorporate the examination, implementation, and broad usage of technologies such as Social, Mobile, Analytics, Cloud, and Internet of Things (SMACIT), Virtual Reality, Blockchain, 3D Printing, Drones, and Augmented Reality, to name a few examples. Adopting the proper digital strategy helps businesses to remain competitive, overcome the challenges of digitalization, and reap the advantages of digitalization Bailie and Chinn [13].

The digital strategy complements the organizational strategy by moving the focus away from prediction and planning and toward testing and adjusting. As an added benefit, it facilitates flexible planning as well as inclusive responsibility and the optimal utilization of information technology capabilities. According to Gökalp et al. [14], the formation of 'Data Science' competences is often included in the strategy for digital solutions. The rapid growth of digitalization is offering possibilities For the big-scale gathering, storage, and evaluation of enterprise records, which, whilst blended with proper 'data technology' methodologies and era, has the capacity to carry great financial advantages. As a result, information science is essential in taking pictures and studying statistics from all organizational touchpoints that allows you to generate enterprise insights throughout a variety of domain names, such as studies and development (R&D), manufacturing (marketing), operations (supply chain management), client dating control (CRM), method formula (strategic planning), finance (monetary making plans), and human resource management (HRM), among others.

Increased ability to gather and manage huge amounts of data in order to extract business insights has generated a rush to deploy 'Data Science' initiatives, which has been propelled by technological developments, in order to extract business insights. It is estimated that early adopters of Artificial Intelligence (AI) would share a global profit pool of $1 trillion by 2030 O'Meara[15]. While the importance of 'Data Science' is widely acknowledged, the bulk of businesses are still in the early stages of using it in their operations. It is estimated that between 75 and 80 percent of businesses have failed to appropriately use "Data Science" Steinwandter et al. [16].

In 2015, more than 75% of businesses made investments in 'Data Science' or planned to do so in the near future Deighton [17]. However, according to Prüfer and Prüfer[18], over 95 percent of firms across all sectors and regions have not employed artificial intelligence to reinvent their business processes. Despite the fact that more than 80 percent of enterprises believe that adopting artificial intelligence has strategic potential, just 23 percent have done so, and another 23 percent have begun with trial projects in artificial intelligence Jöhnk et al.[19]. Despite a 270 percent surge in adoption over the previous four years, just 37 percent of organizations had used artificial intelligence (AI) or had short-term plans to do so by 2019. Arrieta et al. [20]. A considerable majority of those who have used 'Data Science' technologies have not seen the expected return on their investments (ROI) or sustained success as a result of its adoption Reddy et al. [21].

Organizations today majorly focus on data science and artificial intelligence capabilities. They are addressing the need to change the commonly-held view that data is the output of a company's processes. Building artificial intelligence and machine learning in any form relies on vast swathes of data science, acquired across a myriad systems and months or years of historical performance. Sophisticated analytical tools turn that data into rules and patterns that give deep insight into the consequences of various actions and inputs. Success of these integrated digital transformations is thus directly proportional to the ability of an organization to capture, store and analyze data generated from their past transactions.

4. State of the Art

Recently, the science of data has received a great deal of attention. To argue for the renaming of "computer science" to "datalogy," Nasution et al. [22] invented the term "datalogy" (also known as "science of data") and advocated for it to be renamed. It was during the 1990s when the term "data science" gained widespread Aria et al.[23]. Founded in 2002 by CODATA (the Committee on Data for Science and Technology), an organization representing the scientific data research community, the term "data science" was coined to refer to the analysis of data from a variety of scientific research fields, and the term was put into practice with the establishment of The Data Science Journal the following year. The term "data science" does not have an official meaning; instead, only specific research material, breadth, and topics have been recognized. Data science was discussed in detail by Grover et al. [24], who examined various aspects of it, such as technologies, companies that perform data science work, and specialized skill sets. They argued that data science should enable the creation of data products rather than simply being considered a data application, and they provided examples of data science products. The term "data science" was first used by Karpatne et al.[25] to define a new discipline with the goal of research being data in 2009.

The field of data science is entering a new age. Several new data science research organizations have been established, including those in the United States of America (USA), Canada (Canada), Australia (Australia), China, the United Kingdom (UK), Japan (Japan), and Korea (Korea). Journals and proceedings have also been published. In the business world, the role of data scientist is fast gaining in popularity and demand. In addition to developing a data science community, the EMC Corporation performed a Survey of the worldwide records technology community (Pereira, Cunha and Fernandes, 2020). LinkedIn, the world's largest professional network, was responsible for the formation of the data science team. There are several companies searching for data scientists to join their data science teams in order to maintain their competitive advantage in the age of big data, including Google, Facebook, IBM, PayPal, and Amazon.

A data science action plan was published by Bell Labs in 2001, with the purpose of broadening the scope of statistics in the company Donoho [26]. CODATA published its first peer-reviewed journal, the Data Science Journal, in 2002, marking the organization's official start. With time, the Journal has developed into a meeting place for data scientists and other professional data scientists Virkus and Garoufallou [27]. In addition to the Journal of Data Science, which is published by Columbia University, is another excellent resource. In 2009, the first monograph on data science, The Science of Data Logging and Data Analysis, was published by the American Society for Data Science. Cao (2017) defines formalized EPJ Data Science was first published as a Springer Open Journal in 2012, thanks to a collaboration between Springer and EPJ.org.

Data wisdom exploration facilities such as the Institute forInformation Sciences and Engineering at Columbia university and Shanghai Key Laboratory of statistics technological know-how, each positioned at Fudan college, are examples of institutions that are creating data science research facilities. The University of California, Berkeley, introduced an Introduction to Data Science course in 2012. The Introduction to Data Science course at Columbia University was first offered in 2011. In the meantime, recent data science workshops and conferences have been organized, including Data Sciences Summer Institutes (DSSI) at UIUC (University of Illinois at Urbana-Champaign) in 2011 and 2012; and Fudan University's yearly data science workshops, which have been running since 2010. The primary global conference on data technological know-how become sponsored with the aid of China in 2014 in Beijing (ICDS).

5. Current Challenges

In addition to the analytical issues that arise when data science is used in a commercial environment, other obstacles are created. The case examples in this post's introduction are solely intended to draw attention to the current difficulties associated with conducting data science and big data initiatives. When it comes to data science initiatives, experts have taken the effort to compile a list of the most common difficulties and pain spots that they've encountered.

5.1 Building Data Analytics Teams

It is usual for data science projects to be ad hoc in nature, necessitating a great deal of back-and-forth between team members as well as trial-and-error to determine the most appropriate tools, programme, and settings to use in order to analyze the data. Because of the experimental nature of these endeavors, establishing correct expectations, developing realistic project timelines, and estimating project duration may be difficult to accomplish. According to Benaroch et. al.[27], the scope of the project as well as its business objectives may be troublesome prior to the project's commencement date.

A lack of clearly defined business goals, an inadequate return on investment or business case, and an unclear project scope are all concerns that must be addressed, as pointed out by Martinez et al.[9]. According to Mikalef et al. , organizations have been limited in their capacity to successfully employ data analytics as a result of a disproportionate focus placed on technological issues in the past. When it comes to benchmarking efforts, data scientists have traditionally placed a greater emphasis on achieving the best possible results than on addressing the underlying business problem. The pursuit of improved performance, on the other hand, may result in models that are much too complicated to be of any value. This method is advantageous to data science competitions such as Kaggle Prüfer & Prüfer[18], but it is not beneficial to the industrial sector. Kaggle tournaments, on the other hand, are wonderful for machine learning education, but they may lead to unrealistic assumptions about what is necessary in real-world commercial situations Mueller and Massaron[31].

Putting together the correct project team, as Thomas[32] points out, is tough, and Succi and Canovi[33] point out that there is a scarcity of analytically-minded individuals. Several new big data, analytics, and data science degrees have been introduced at every major institution in response to a scarcity of highly trained analytical workers Song and Zhu[34]. According to Donoghue, Voytek, and Ellis, a multi-disciplinary team is required for data science initiatives to succeed in order to be successful (2021). Because of the immaturity of the process and the absence of a robust team-based approach, for example, Wang et al. [31] say that data science teams are heavily reliant on the lead data scientist.

5.2 Communication, Collaboration, and Coordination

As a subject, data science is moving away from working in isolation and toward working in collaborative teams. According to Sablis, Smite, and Moe, "the management of dependencies across job activities" is the most significant problem for data science endeavors in general (2021). Processes that are not coordinated result in inefficiency, mistakes, and confusion. Observe that this lack of collaboration exists both inside and between data analytics teams. This is critical to understand Dubey et al.[36].

Several researchers, like Patricio et al. [37] have noted that there is a lack of open communication among three main stakeholders : the company (client), the analytics team, and the information technology department. Data science teams face a variety of challenges, including those highlighted by Hong et al., such as deploying to production, collaborating with IT, and collaborating with business partners, among others. However, Ferraris et al.[38] demonstrate that even in the presence of sufficient business input or domain knowledge information, there is insufficient support to provide satisfactory outcomes. A general lack of effective communication between information technology and business agents, as well as the data analytics team, seems to be a concern.

According to the authors of Wang, Kung, and Byrd[39] as well as Vilminko-Heikkinen and Pekkola[40], there is a lack of proper management and top-level sponsorship in connection to data analytics governance, which they describe as follows: According to Hekkala et al.[41], working in a chaotic, disorganized environment can demotivate team members and make it difficult for them to focus on the project's goals and objectives. Moreover, working in a chaotic, disorganized environment can make it difficult for team members to communicate effectively.

6. Conclusion

This review article analyses the issues presented by data and why data science is needed. Additionally, it analyses how data science varies from current technologies and areas. In addition, it looked at some fundamental difficulties that data science will confront as it grows into an academic field utilizing data as its study object, such as core concepts, innovative methodologies, and research themes. Additionally, the book assesses the accomplishments and challenges of the current state of research and society of data science. Last but not least, it explains how existing knowledge can be transferred to new science.

References

1. Wang, Q., Su, M., Zhang, M. and Li, R., 2021. Integrating digital technologies and public health to fight Covid-19 pandemic: key technologies, applications, challenges and outlook of digital healthcare. International Journal of Environmental Research and Public Health, 18(11), p.6053.
2. Verhoef, P.C., Broekhuizen, T., Bart, Y., Bhattacharya, A., Dong, J.Q., Fabian, N. and Haenlein, M., 2021. Digital transformation: A multidisciplinary reflection and research agenda. Journal of Business Research, 122, pp.889-901.
3. Tang, C., 2021. The Composition of Data Capital. In Data Capital (pp. 109-141). Springer, Cham.

4. Agarwal, R. and Dhar, V., 2014. Big data, data science, and analytics: The opportunity and challenge for IS research. Information systems research, 25(3), pp.443-448.
5. Gibert, K., Horsburgh, J.S., Athanasiadis, I.N. and Holmes, G., 2018. Environmental data science. Environmental Modelling & Software, 106, pp.4-12.
6. Lv, Z., Song, H., Basanta-Val, P., Steed, A. and Jo, M., 2017. Next-generation big data analytics: State of the art, challenges, and future research topics. IEEE Transactions on Industrial Informatics, 13(4), pp.1891-1899.
7. Suthaharan, S., 2016. Machine learning models and algorithms for big data classification. Integr. Ser. Inf. Syst, 36, pp.1-12.
8. Cao, L., 2017. Data science: a comprehensive overview. ACM Computing Surveys (CSUR), 50(3), pp.1-42.
9. Martinez, I., Viles, E. and Olaizola, I.G., 2021. Data science methodologies: Current challenges and future approaches. Big Data Research, 24, p.100183.
10. Strukova, S., Ruipérez-Valiente, J.A. and Mármol, F.G., 2022. Identifying Experts in Question\& Answer Portals: A Case Study on Data Science Competencies in Reddit. arXiv preprint arXiv:2204.04098.
11. McQuillan, D., 2018. Data science as machinic neoplatonism. Philosophy & Technology, 31(2), pp.253-272.
12. Varshney, D., 2020. Digital transformation and creation of an agile workforce: Exploring company initiatives and employee attitudes. In Contemporary global issues in human resource management. Emerald Publishing Limited.
13. Bailie, B. and Chinn, M., 2018, April. Effectively harnessing data to navigate the new normal: Overcoming the barriers of digital adoption. In Offshore Technology Conference. OnePetro.
14. Gökalp, M.O., Gökalp, E., Kayabay, K., Koçyiğit, A. and Eren, P.E., 2021. The development of the data science capability maturity model: a survey-based research. Online Information Review.
15. O'Meara, S., 2019. Will China lead the world in AI by 2030. Nature, 572(7770), pp.427-428.
16. Steinwandter, V., Borchert, D. and Herwig, C., 2019. Data science tools and applications on the way to Pharma 4.0. Drug discovery today, 24(9), pp.1795-1805.
17. Deighton, J., 2017. Rethinking the profession formerly known as advertising: How data science is disrupting the work of agencies. Journal of Advertising Research, 57(4), pp.357-361.
18. Prüfer, J. and Prüfer, P., 2020. Data science for entrepreneurship research: studying demand dynamics for entrepreneurial skills in the Netherlands. Small Business Economics, 55(3), pp.651-672.
19. Jöhnk, J., Weißert, M. and Wyrtki, K., 2021. Ready or not, AI comes—an interview study of organizational AI readiness factors. Business & Information Systems Engineering, 63(1), pp.5-20.
20. Arrieta, A.B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., García, S., Gil-López, S., Molina, D., Benjamins, R. and Chatila, R., 2020. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information fusion, 58, pp.82-115.
21. Reddy, R.C., Bhattacharjee, B., Mishra, D. and Mandal, A., 2022. A systematic literature review towards a conceptual framework for enablers and barriers of an enterprise data science strategy. Information Systems and e-Business Management, pp.1-33.
22. Nasution, M.K., Sitompul, O.S. and Nababan, E.B., 2020, June. Data science. In Journal of Physics: Conference Series (Vol. 1566, No. 1, p. 012034). IOP Publishing.

23. Aria, M., Misuraca, M. and Spano, M., 2020. Mapping the evolution of social research and data science on 30 years of Social Indicators Research. Social indicators research, 149(3), pp.803-831.
24. Grover, V., Chiang, R.H., Liang, T.P. and Zhang, D., 2018. Creating strategic business value from big data analytics: A research framework. Journal of Management Information Systems, 35(2), pp.388-423.
25. Karpatne, A., Atluri, G., Faghmous, J.H., Steinbach, M., Banerjee, A., Ganguly, A., Shekhar, S., Samatova, N. and Kumar, V., 2017. Theory-guided data science: A new paradigm for scientific discovery from data. IEEE Transactions on knowledge and data engineering, 29(10), pp.2318-2331.
26. Donoho, D., 2017. 50 years of data science. Journal of Computational and Graphical Statistics, 26(4), pp.745-766.
27. Virkus, S. and Garoufallou, E., 2019. Data science from a library and information science perspective. Data Technologies and Applications.
28. Benaroch, M., Lichtenstein, Y. and Fink, L., 2016. Contract design choices and the balance of ex ante and ex post transaction costs in software development outsourcing. MIS quarterly, 40(1), pp.57-82.
29. Mikalef, P., Pappas, I.O., Krogstie, J. and Giannakos, M., 2018. Big data analytics capabilities: a systematic literature review and research agenda. Information Systems and e-Business Management, 16(3), pp.547-578.
30. Cao, L., 2017. Data science: challenges and directions. Communications of the ACM, 60(8), pp.59-68.
31. Wang, D., Weisz, J.D., Muller, M., Ram, P., Geyer, W., Dugan, C., Tausczik, Y., Samulowitz, H. and Gray, A., 2019. Human-ai collaboration in data science: Exploring data scientists' perceptions of automated ai. Proceedings of the ACM on Human-Computer Interaction, 3(CSCW), pp.1-24.
32. Thomas, G., 2017. How to do your research project: A guide for students. Sage.
33. Succi, C. and Canovi, M., 2020. Soft skills to enhance graduate employability: comparing students and employers' perceptions. Studies in higher education, 45(9), pp.1834-1847.
34. Song, I.Y. and Zhu, Y., 2016. Big data and data science: what should we teach?. Expert Systems, 33(4), pp.364-373.
35. Donoghue, T., Voytek, B. and Ellis, S.E., 2021. Teaching creative and practical data science at scale. Journal of Statistics and Data Science Education, 29(sup1), pp.S27-S39.
36. Dubey, R., Gunasekaran, A., Childe, S.J., Roubaud, D., Wamba, S.F., Giannakis, M. and Foropon, C., 2019. Big data analytics and organizational culture as complements to swift trust and collaborative performance in the humanitarian supply chain. International Journal of Production Economics, 210, pp.120-136.
37. Patricio, R., Moreira, A., Zurlo, F. and Melazzini, M., 2020. Co‐creation of new solutions through gamification: A collaborative innovation practice. Creativity and Innovation Management, 29(1), pp.146-160.
38. Ferraris, A., Mazzoleni, A., Devalle, A. and Couturier, J., 2018. Big data analytics capabilities and knowledge management: impact on firm performance. Management Decision.
39. Wang, Y., Kung, L. and Byrd, T.A., 2018. Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations. Technological Forecasting and Social Change, 126, pp.3-13.

40. Vilminko-Heikkinen, R. and Pekkola, S., 2017. Master data management and its organizational implementation: An ethnographical study within the public sector. Journal of Enterprise Information Management.
41. Hekkala, R., Stein, M.K. and Rossi, M., 2018. Metaphors in managerial and employee sensemaking in an information systems project. Information Systems Journal, 28(1), pp.142-174.
42. Hong, T., Gao, D.W., Laing, T., Kruchten, D. and Calzada, J., 2018. Training energy data scientists: universities and industry need to work together to bridge the talent gap. IEEE Power and Energy Magazine, 16(3), pp.66-73.
43. Pereira, P., Cunha, J. and Fernandes, J.P., 2020, August. On understanding data scientists. In 2020 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC) (pp. 1-5). IEEE.