IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 5.731
Volume 11, Issue 1, November 2021

An Overview of Data Science and Analytics from the Perspective of Data-Driven Smart Computing, Decision-Making, and Applications

Ashish Chauhan[1], Bagoria Kuber[2], Vishwakarma Rohit[3]
Asst. Professor[1] and TYCS[2,3]
Uttar Bhartiya Sangh's Mahendra Pratap Sharda Prasad Singh College of Commerce & Science, Mumbai, Maharashtra

Abstract: *In the era of the Fourth Industrial Revolution (Industry 4.0 or 4IR), the digital world has a vast amount of data, including internet of things (IoT) data, business data, health data, mobile data, urban data, security data, and more. The data can be analyzed to derive valuable knowledge or insights, which can then be utilized for intelligent decision-making in many application fields. In the field of data science, the utilization of advanced analytics techniques such as machine learning modeling can yield actionable insights and enhance understanding of data. This enables the automation and intelligence of the computing process. This paper provides a thorough examination of "Data Science," encompassing a range of sophisticated analytical techniques that can be utilized to improve the intelligence and functionality of an application by making intelligent decisions in various situations. In addition, we analyze and condense ten possible practical areas of application, such as business, healthcare, cybersecurity, urban and rural data science, and others. This analysis is based on the consideration of data-driven intelligent computing and decision-making. Therefore, we now emphasize the difficulties and possible areas of research that fall within the scope of our study. This document seeks to provide a comprehensive resource on data science and advanced analytics for researchers, decision-makers, and application developers. It specifically focuses on data-driven solutions for real-world challenges..*

Keywords: Data science, Advanced analytics, Machine learning, Deep learning, Smart computing, Decision-making, Predictive analytics, Applications of data science

I. INTRODUCTION

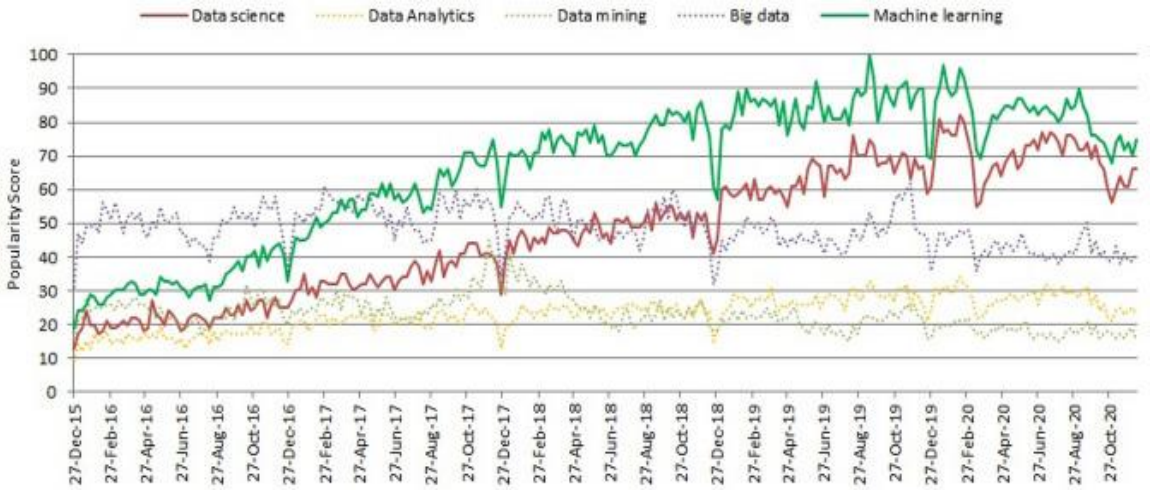
In the current era of "data science and advanced analytics," nearly every aspect of our everyday lives is digitally documented as data. The present digital landscape is abundant with several forms of data, including corporate data, financial data, healthcare data, multimedia data, internet of things (IoT) data, cybersecurity data, social media data, and more. The data can be categorized as organized, semi-structured, or unstructured, and its volume is always growing. Data science is a discipline that aims to integrate statistics, data analysis, and their associated methodologies in order to comprehend and examine real-world phenomena using data. Cao et al. define data science as the scientific study of data, encompassing the creation of data products that can include discoveries, predictions, services, suggestions, insights for decision-making, thoughts, models, paradigms, tools, or systems.

Data science is the discipline that involves the application of sophisticated analytical techniques and scientific principles to extract valuable commercial insights from data. The primary focus of advanced analytics lies in the anticipation of data utilization to identify trends and predict future occurrences. Basic analytics provides a general overview of data, but advanced analytics goes beyond that by providing a more thorough comprehension of data and assisting in the analysis of detailed and specific data, which is of particular importance to us. Data science encompasses various types of analytics, including descriptive analytics, which focuses on explaining what happened; diagnostic analytics, which aims to understand why it happened; predictive analytics, which forecasts future events; and prescriptive analytics, which provides recommendations on what actions to take. These analytics methods are discussed in more detail in the book "Advanced analytics methods and smart computing". Advanced analytics and decision-making, which are based on machine learning techniques, are a crucial component of artificial intelligence (AI). These

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 5.731
Volume 11, Issue 1, November 2021

capabilities can have a substantial impact on the Fourth Industrial Revolution (Industry 4.0) by enabling smart computing and automation through their ability to learn.

While the field of "data science" encompasses a wide range of topics, our primary focus is on extracting valuable insights using sophisticated analytical techniques. These insights are then utilized to make intelligent judgments in numerous practical domains. To do this, a range of sophisticated analytics techniques, including machine learning modeling, natural language processing, sentiment analysis, neural networks, and deep learning analysis, can be employed. These methods can offer profound insights into data, enabling the development of intelligent applications driven by data. In our study, we consider regression analysis, classification, clustering analysis, association rules, time-series analysis, sentiment analysis, behavioural patterns, anomaly detection, factor analysis, log analysis, and deep learning, which is derived from artificial neural networks.



The article "Advanced analytics methods and smart computing" provides a concise overview of these machine learning-based advanced analytics methods. Therefore, it is crucial to comprehend the principles of the aforementioned advanced analytics methodologies and their suitability for implementation in different real-world domains. In our previous publication, Sarker et al., we explored the application of data science and machine learning modeling in the field of cybersecurity. We highlighted the importance of these techniques in making informed decisions and delivering intelligent security services based on data analysis. This paper provides a comprehensive overview of data science application areas and real-world problems in ten potential domains. These domains include business data science, health data science, IoT data science, behavioral data science, urban data science, and others. The discussion of these domains is briefly presented in the section titled "Real-world application domains".

This paper provides a comprehensive overview of "Data Science," which involves using advanced analytics methods to extract valuable insights from data and make data-driven decisions. The focus is on the importance of machine learning modeling in enhancing the intelligence and capabilities of applications. The primary focus of this work is to comprehend the modeling of data science, elucidating diverse analytical techniques for a solution-oriented approach and their suitability in distinct real-world applications discussed previously. The main objective of this work is to serve as a fundamental guide or reference for those in academia and industry who are interested in studying, researching, and developing automated and intelligent applications or systems using smart computing and decision making in the field of data science.

1.1 Paper's primary contributions can be summarized as follows:

Our study aims to establish the boundaries of data-driven smart computing and decision-making in real-world scenarios. We also provide a concise analysis of the notion of data science modeling, which involves the transformation of business challenges into data products and automation. This analysis aims to comprehend the practicality of data science modeling and its potential to offer intelligent services in real-world situations.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 5.731
Volume 11, Issue 1, November 2021

• The aim is to present a thorough understanding of data science, which encompasses sophisticated analytics techniques that may be utilized to improve the intelligence and functionality of an application.
• The objective is to examine the suitability and importance of machine learning-based analytics techniques in different practical domains. In addition, we provide a concise overview of ten possible practical domains, ranging from commercial to individual applications in our everyday existence, where the utilization of sophisticated data analysis techniques combined with machine learning algorithms can be employed to attain the desired results.

• The purpose of our study is to emphasize and provide a concise overview of the difficulties and possible areas of research that fall within its scope.

The subsequent sections of the paper are structured as follows. The following section presents the context and previous research, and establishes the boundaries of our investigation. This section introduces the principles of data science modeling for constructing a data-centric application. Subsequently, provide a concise analysis and elucidation of many sophisticated analytics techniques and intelligent computing. The next section provides a discussion and summary of different practical domains where the application of concepts can be observed in real-world scenarios. Subsequently, we emphasize and succinctly outline various research concerns and possible avenues for future exploration. Lastly, the concluding part wraps up this paper.

II. CONTEXT AND RELEVANT RESEARCH

2.1 Glossary of Data Terminology and Definitions

In the sector, there are several important terminology that are closely interconnected and can be easily misunderstood. These phrases include data analysis, data mining, data analytics, big data, data science, advanced analytics, machine learning, and deep learning. In the following, we will provide definitions for these categories and distinguish them from the phrase "Data Science" based on our objective.

Data analysis is the act of processing data using traditional ideas, technologies, and techniques such as statistical, empirical, or logical methods. The goal is to extract valuable information and apply it to actual situations. Data analytics encompasses the ideas, methods, instruments, and processes used to gain a thorough knowledge and investigation of actionable insights from data. The primary focus of this process is the statistical and mathematical examination of the data. "Data mining" is a widely used term in the past ten years that is synonymous with other terms such as extracting knowledge from data, knowledge discovery, knowledge extraction, data analysis, data archaeology, and data dredging.

Han et al. (2021) argue that a more suitable name for the concept should have been "knowledge mining from data". Data mining is the systematic extraction of valuable patterns and knowledge from vast quantities of data. Possible data sources encompass databases, data centers, the Internet or Web, other data repositories, or data that is dynamically streamed across the system. The term "big data" is now popular and has the potential to revolutionize statistical and data analysis methods due to its distinct characteristics of being vast, high dimensional, heterogeneous, complicated, unstructured, incomplete, noisy, and erroneous. Mobile devices, social networks, the Internet of Things, multimedia, and various other emerging applications are sources of big data. Big data is comprehended and characterized using various distinctive attributes, namely volume, velocity, variety, veracity, and value (together known as the 5Vs), as well as complexity.

Basic analytics offers a concise overview of data, whereas "Advanced Analytics" goes beyond that by providing a more profound comprehension of data and facilitating the analysis of detailed information. Advanced analytics refers to the process of autonomously or semi-autonomously analyzing data or material using sophisticated tools and methodologies to uncover profound insights, make predictions, or produce suggestions. This goes beyond the scope of traditional business intelligence or analytics. "Machine learning", a subdivision of artificial intelligence (AI), is a prominent methodology employed in sophisticated analytics that can automate the construction of analytical models. This focuses on the concept that systems have the ability to acquire knowledge from data, identify patterns, and make decisions, with limited human intervention. "Deep Learning" is a branch of machine learning that focuses on algorithms that are influenced by the structure and function of the human brain, specifically artificial neural networks.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 5.731
Volume 11, Issue 1, November 2021

Data science is a comprehensive term that includes advanced data analytics, data mining, machine learning, deep learning modeling, and other related disciplines such as statistics. Its purpose is to extract insights or valuable knowledge from datasets and convert them into actionable business strategies. Cao et al. provided a disciplinary definition of data science as a new interdisciplinary field that integrates statistics, informatics, computing, communication, management, and sociology. Its purpose is to analyze data and its various environments, including domains and contextual aspects such as organizational and social factors. The goal is to convert data into insights and decisions by employing a data-to-knowledge-to-wisdom approach and methodology. The article "Understanding data science modeling" provides a practical overview of the process of data science modeling. It begins by addressing business problems and then explores how data products can be used to help data scientists analyze and solve real-world problems within the field of data science and analytics.

III. LITERATURE REVIEW

Within the region, the researchers have examined multiple papers pertaining to data science and its importance. The authors discuss the emerging discipline of data science and its significance in the broader knowledge landscape. They also highlight some challenges that distinguish data science and informatics difficulties from traditional techniques in information sciences.

Donoho et al. provide a comprehensive overview of 50 years of data science, which includes an analysis of data science in mass media and a discussion on the similarities and differences between data science and statistics. The authors introduce and define the theory-guided data science (TGDS) approach and provide a classification of research themes within TGDS. Cao et al. provide an extensive examination and instructional guide on the essential elements of data science, encompassing the shift from data analysis to data science, the concepts behind data science, and the discipline and proficiency required for data education.

In addition, the authors incorporate a data science study that seeks to offer an accurate portrayal of the use of statistical characteristics and associated data science techniques in the field of bioimage informatics. The authors of the report analyze the main data science algorithms used by central banks and demonstrate the increasing popularity of these algorithms over time. This paper adds to the development of a research direction on the significance of data science in central banking. The authors present a comprehensive and instructional guide on the data-driven design of intelligent wireless networks. The authors present a comprehensive exploration of computational optimum transport, specifically focusing on its application to data science. The authors introduce data science as theoretical advances in information systems through the use of text analytics.

In contrast to the aforementioned recent studies, our focus in this paper is on the understanding of data science, which encompasses advanced analytics methodologies, machine learning models, real-world application domains, and prospective research possibilities within the scope of our study. This paper discusses advanced analytics methods that utilize machine learning techniques. These methods can be employed to improve the capabilities of an application by enabling data-driven intelligent decision making and automation in the final data product or systems.

In summary, data science modeling can effectively facilitate the implementation of modifications and enhancements in business procedures. The intriguing aspect of the data science approach involves acquiring a more profound comprehension of the business challenge that needs to be resolved. Without that, the process of collecting accurate data and extracting the most valuable information from it for decision-making would become far more challenging. Data scientists are primarily responsible for analyzing and manipulating data to discover insights that assist organizations in making unbiased judgments and resolving intricate issues.

IV. CONCLUSION

This paper provides a thorough examination of data science, encompassing a range of advanced analytical techniques that can be utilized to augment the intelligence and functionality of an application. We have also depicted the present popularity of data science and machine learning-driven advanced analytical modeling, and also distinguished these from the pertinent words employed in the field, in order to establish the significance of this paper. This study focuses on the comprehensive analysis of data science modeling, including the different processing modules required to extract meaningful insights from data for a specific business challenge and the resulting data product. Therefore, in alignment

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 5.731
Volume 11, Issue 1, November 2021

with our objective, we have provided a concise overview of the importance of various data modules in a data-driven business solution within the context of the data science process. In addition, we have compiled a comprehensive overview of modern analytical techniques and their corresponding results, as well as machine learning models, which are essential for addressing the related business challenges. The main contribution of this study is the explanation of various advanced analytical methods and how they can be applied in real-world data-driven applications such as business, healthcare, cybersecurity, urban and rural data science. The study focuses on the use of data-driven smart computing and decision making in these areas.

In conclusion, we have identified and examined the difficulties encountered during our study, along with potential areas for further investigation and future paths to explore. The issues revealed present significant research prospects in the sector, which can be investigated with effective solutions to enhance the data-driven model and systems. In summary, our study on advanced analytical solutions utilizing data science and machine learning methods has yielded promising results. It can serve as a valuable reference for future research and practical applications in the field of data science, benefiting both academic and industry professionals.

REFERENCES

[1]. Adnan N, Nordin SM, Rahman I, Noor A. The effects of knowledge transfer on farmers decision making toward sustainable agriculture practices. World J Sci Technol Sustain Dev. 2018.

[2]. Agrawal R, Gehrke J, Gunopulos D, Raghavan P. Automatic subspace clustering of high dimensional data for data mining applications. In: Proceedings of the 1998 ACM SIGMOD international conference on management of data. 1998. p. 94–105.

[3]. Agrawal R, Imieliński T, Swami A. Mining association rules between sets of items in large databases. In: ACM SIGMOD record, vol 22. ACM. 1993. p. 207–16.

[4]. Agrawal R, Srikant R. Fast algorithms for mining association rules. In: Proceedings of the international joint conference on very large data bases, Santiago, Chile, vol 1215. 1994. p. 487–99.

[5]. Aha DW, Kibler D, Albert MK. Instance-based learning algorithms. Mach Learn. 1991;6(1):37–66. doi: 10.1007/BF00153759. [CrossRef]

[6]. Al-Abassi A, Karimipour H, HaddadPajouh H, Dehghantanha A, Parizi RM. Industrial big data analytics: challenges and opportunities. In: Handbook of big data privacy. Springer; 2020. p. 37–61.

[7]. Al-Garadi MA, Mohamed A, Al-Ali AK, Du X, Ali I, Guizani M. A survey of machine and deep learning methods for internet of things (iot) security. IEEE CommunSurv Tutor. 2020;22(3):1646–1685. doi: 10.1109/COMST.2020.2988293. [CrossRef]

[8]. Ankerst M, Breunig MM, Kriegel H-P, Sander J. Optics: ordering points to identify the clustering structure. ACM Sigmod Rec. 1999;28(2):49–60. doi: 10.1145/304181.304187. [CrossRef]

[9]. Atzori L, Iera A, Morabito G. The internet of things: a survey. ComputNetw. 2010;54(15):2787–805. doi: 10.1016/j.comnet.2010.05.010. [CrossRef]

[10]. Baldi P. Autoencoders, unsupervised learning, and deep architectures. In: Proceedings of ICML workshop on unsupervised and transfer learning. 2012. p. 37–49.

[11]. Balducci F, Impedovo D, Pirlo G. Machine learning applications on agricultural datasets for smart farm enhancement. Machines. 2018;6(3):38. doi: 10.3390/machines6030038. [CrossRef]

[12]. Box GEP, Jenkins GM, Reinsel GC, Ljung GM. Time series analysis: forecasting and control. New York: Wiley; 2015.

[13]. Breiman L. Random forests. Mach Learn. 2001;45(1):5–32. doi: 10.1023/A:1010933404324. [CrossRef]

[14]. Brettel M, Friederichsen N, Keller M, Rosenberg M. How virtualization, decentralization and network building change the manufacturing landscape: an industry 4.0 perspective. FormaMente 2017;12.

[15]. Canadian institute of cybersecurity. University of new Brunswick, iscx dataset. http://www.unb.ca/cic/datasets/index.html/. Accessed 20 Oct 2019.

[16]. Cao H, Bao T, Yang Q, Chen E, Tian J. An effective approach for mining mobile user habits. In: Proceedings of the international conference on information and knowledge management, Toronto, ON, Canada, 26–30 October. New York: ACM; 2010. p. 1677–80.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 5.731
Volume 11, Issue 1, November 2021

[17]. Cao L. Data science: a comprehensive overview. ACM ComputSurv (CSUR). 2017;50(3):1–42. doi: 10.1145/3076253. [CrossRef]

[18]. Carpenter GA, Grossberg S. A massively parallel architecture for a self-organizing neural pattern recognition machine. Comput Vis Graph Image Process. 1987;37(1):54–115. doi: 10.1016/S0734-189X(87)80014-2. [CrossRef]

[19]. Tsai C-W, Lai C-F, Chao H-C, Vasilakos AV. Big data analytics: a survey. *J Big Data*. 2015;2(1):1–32. doi: 10.1186/s40537-015-0030-3.

[20]. Tuncel KS, Baydogan MG. Autoregressive forests for multivariate time series modeling. *Pattern Recognit*. 2018;73:202–15. doi: 10.1016/j.patcog.2017.08.016.

[21]. Wagstaff K, Cardie C, Rogers S, Schrödl S, et al. Constrained k-means clustering with background knowledge. *ICML*. 2001;1:577–84.

[22]. Wang J, Zhang W, Shi Y, Duan S, Liu J. Industrial big data analytics: challenges, methodologies, and applications. 2018. arXiv:1807.01016.

[23]. Wang L, Zhang J, Chen G, Qiao D. Identifying comparable entities with indirectly associative relations and word embeddings from web search logs. Decis Support Syst. 2021:141.

[24]. Wang W, Yang J, Muntz R, et al. Sting: a statistical information grid approach to spatial data mining. *VLDB*. 1997;97:186–95.

[25]. Waskom ML. Seaborn: statistical data visualization. *J Open Source Softw*. 2021;6(60):3021. doi: 10.21105/joss.03021.

[26]. Wei P, Li Y, Zhang Z, Tao H, Li Z, Liu D. An optimization method for intrusion detection classification model based on deep belief network. *IEEE Access*. 2019;7:87593–605. doi: 10.1109/ACCESS.2019.2925828.