IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

OVERVIEW OF DATA SCIENCE AND ADVANCED ANALYTICS

[1]Sameer Nagar

[1]Student

[1]Artificial Intelligence and Data Science – Computer Science & Engineering

[1]Acropolis Institute of Technology and Research, Indore, India

Abstract: Data science is an interdisciplinary academic field that uses statistics, scientific computing, scientific methods, processes, algorithms and systems to extract or extrapolate knowledge and insights from noisy, structured, and unstructured data Analytics is the systematic computational analysis of data or statistics. It is used for the discovery, interpretation, and communication of meaningful patterns in data. It also entails applying data patterns toward effective decision-making. I present an overview of "Data Science" and “Advanced Analytics”. It is to provide everyone with an overview of data science and advanced analytics,

Index Terms - Data Science, Advanced Analytics, Machine Learning, Deep Learning, Smart Computing, Predictive Analysis, Big Data, Data Mining, Data Warehousing

I. INTRODUCTION

Data science is the practice of using analytics techniques and scientific ideas to extract useful business data. Advanced analytics is a step forward in providing a deeper understanding of data and assisting in the analysis of granular data, which is what I are interested in. Basic analytics only provide a general description of the data.

There are types of analytics in the field of data science, such as "Descriptive analytics," which provides an answer to the question "what happened." "Diagnostic analytics," which provides an answer to the question " "predictive analytics," which makes predictions about the future; also "Prescriptive examination" which recommends what move ought to be made, talked about momentarily in "Cutting edge examination techniques and shrewd figuring".

Even though "data science" is a huge field, our main focus is on using advanced analytics to get useful insights that can be used to make smart decisions in a variety of real-world application areas. In order to accomplish this, advanced analytics techniques like machine learning modeling, natural language processing, sentiment analysis, neural network analysis, and deep learning analysis can provide a deeper understanding of data and, as a result, can be utilized to create intelligent applications that are driven by data.

II. OBJECTIVES

1. To define the scope of research into smart computing and decision-making based on data in everyday life. In order to comprehend its applicability and provide intelligent services in real-world scenarios, we also have a brief discussion on the concept of data science modeling from business problems to data product and automation.
2. To present a complete picture of data science, including advanced analytics techniques that can be used to boost an application's intelligence and capabilities.
3. To talk about how machine learning-based analytics can be used in a variety of real-world applications and how important they are. I additionally sum up ten potential certifiable application regions, from business to customized applications in our day to day existence, where cutting-edge examination with AI displaying can be utilized to accomplish the normal result.

4. To draw attention to and provide a summary of the difficulties and potential directions for future research within the scope of our study.

III. LITERATURE REVIEW

Data mining has the same meaning as "knowledge mining from data," "knowledge extraction," "knowledge discovery from data" (KDD), "data/pattern analysis," "data archaeology," and "data dredging." Another popular term in today's world is "big data," which has the unique characteristics of "massive, high dimensional, heterogeneous, complex, unstructured, incomplete, noisy, and erroneous". Mobile devices, social networks, the Internet of Things, multimedia, and numerous other new applications can generate big data. To comprehend and describe big data, a number of distinctive characteristics, such as volume, velocity, variety, veracity, value (5Vs), and complexity, are utilized.

Types of Real-World Data

The availability of data is typically the most important factor when developing a data-driven real-world system in a particular field. The data can be of a variety of types, including

(i) Structured, which has a well-defined data structure and is arranged in a standard way. Examples of structured data include names, dates, addresses, credit card numbers, stock information, geolocation, and so on;

(ii) Unstructured: has no predetermined format or organization; examples include sensor data, emails, blog posts, wikis, word processing documents, audio files, videos, images, presentations, web pages, and so on;

(iii) Semi-organized — has components of both the organized and unstructured information containing specific authoritative properties, models are HTML, XML, JSON archives, NoSQL data sets, and so on.;

(iv) Metadata, which is information about the data and includes things like the author, file type, file size, creation date and time, and last modification date and time, among other things.

Steps of Data Science Modeling-:

1. Getting a handle on business issues: Getting a clear understanding of the problem that needs to be solved, its impact on the relevant individuals or organizations, the ultimate objectives for addressing it, and the relevant research paper plan are all part of this
2. Comprehending data: The availability of data is largely what drives data science. To create a data-driven model or system, a solid understanding of the data is therefore required.
3. Exploration and pre-processing of data: In data science, an approach to analyzing datasets to summarize their key characteristics, frequently using visual methods, is referred to as exploratory data analysis. Data exploration is typically used to establish a foundational evaluation of the data's quality, quantity, and characteristics.
4. Evaluation and modeling of machine learning: Data scientists create a model, algorithm, or set of models to address the business issue after the data has been prepared for modeling.
5. Automation and data production: A data product is a guide, data-enabled deliverable, or discovery, prediction, service, suggestion, insight into decision-making, thought, model, paradigm, tool, application, or system that processes data and produces results

IV. CONCLUSION

We have outlined and discussed the literature related to Data Science, Data Mining, Advanced Analytics and Steps to Data Science Modelling.

In the end, I come to the conclusion that my research and shared knowledge has positive implications and can serve as a guide for future data science research and applications by professionals in industry and academia. This is a brief overview I have provided regarding data science and data mining which will help many people understand the concepts briefly and easily.

V. ACKNOWLEDGMENT

I take this opportunity to express my deepest and sincere gratitude to my mentor Dr. Praveen Bhanodia, Computer Science & Engineering Department, Acropolis Institute of Technology & Research, Indore for his insightful advice, motivating suggestions, invaluable guidance, help and support in successful completion of this research paper and also for his constant encouragement and advice.

I express my deep gratitude to Prof. Shiv Shankar Rajput, Computer Science & Engineering Department, Acropolis Institute of Technology & Research, Indore for his regular support, co-operation, and co-ordination.

I express my hearty gratitude to for Dr. Kamal Kumar Sethi, Professor & Head, Computer Science & Engineering Department, Acropolis Institute of Technology & Research, Indore his support, and help provided during the research paper tenure.

I also express my deep gratitude to Dr. S. C. Sharma, Principal, Acropolis Institute of Technology & Research, Indore for all the help they provided for the completion of research paper.

The in-time facilities provided by the department throughout the Master's program are also equally acknowledgeable.

I would like to convey my thanks to the teaching and non-teaching staff of the Department of Computer Science Engineering, and for their invaluable help and support throughout the period of Master's Degree. I am also grateful to all my classmates for their help, encouragement and invaluable suggestions.

Finally, yet more importantly, I would like to express my deep appreciation to our grandparents, parents, sister and brother for their perpetual support and encouragement throughout the Master's degree period.

.

REFERENCES

[1]. The effects of knowledge transfer on farmers' decision-making regarding sustainable agricultural practices, Adnan N, Nordin SM, Rahman I, and Noor A. World J Sci Technol Develop 2018.

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

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

[4]. Fast algorithms for mining association rules, by R. Agrawal and R. Srikant. In: Procedures of the worldwide joint gathering on extremely huge information bases, Santiago, Chile, vol 1215. 1994. p. 487–99.

[5]. Kibler D, Aha DW, and Albert MK. Algorithms for instance-based learning Learn Mach. 1991;6(1):37–66. [[Search Scholar] Al-Abassi A, Karimipour H, HaddadPajouh H, Dehghantanha A, Parizi RM. Big data analytics in the industry: opportunities and challenges In: Handbook on the privacy of big data. Springer; 2020. p. 37–61.

[7]. A review of machine and deep learning strategies for internet of things (iot) security by Al-Garadi MA, Mohamed A, Al-Ali AK, Du X, Ali I, and Guizani M. IEEE Commun Surv Instructor. 2020;22(3):1646–1685.

[8].[[From Google Scholar] Ankerst M, Kriegel H-P, Breunig MM, and Sander J. Optics: ordering points to determine the structure of the clusters. Sigmod Rec. ACM 1999;28(2):49–60.

[9].[[From Google Scholar] Morabito G, Atzori L, and A. Iera. The internet of things: an inquiry. Netw. Computing 2010;54(15):2787–805.

[10]. [Google Scholar] Autoencoders, deep architectures, and unsupervised learning by Baldi P. In: The papers presented at the ICML workshop on transfer and unsupervised learning. 2012. p. 37–49.

[11]. Data Science and Analytics: An Overview from Data-Driven Smart Computing, Decision-Making and Applications Perspective Iqbal H. Sarker 377 (2021)