# A COMPARATIVE STUDY ON DATA ANALYTICS

**Neha Singh, Jyoti Srivastava, Kriti Mishra, Vibha Srivastava**
*Department of Computer Science & Engineering , Ajay Kumar Garg Engineering College Ghaziabad*
*singhneha@akgec.ac.in, jyotisrivastava@kashiit.ac.in, mishrakriti@akgec.ac.in, srivastavavibha@akgec.ac.in*

***Abstract*—A portion of information that has been formatted using a particular method is called data. Generally speaking, data can take many different forms, such as words or numbers on paper, bits or bytes in computer memory, or facts in a person's brain. We have existed with some data ever since the dawn of human civilization. Whether for a purpose or not, we engage with data on a daily basis. Data is crucial to the expansion of any kind of organization since it allows for the correct planning and execution of future projects as well as the structure of those projects. Data-driven choices will help businesses grow and have a high possibility of success. Business intelligence and data analysis make it easy to analyze the data and provide insights into the dilution. Both business intelligence and data science are based on data. Despite being the bigger pool with more information, data science can be seen as a part of the whole. This research paper identifies and compares the knowledge domains and proficiencies that characterize data analytics, data science, and big data. In this work, we have discussed a number of data processing methods as well as several data types.**

***Keywords*— Data Analytics, Data Science, Big Data, Business Intelligence, Data Analysis**

## I. INTRODUCTION

Data is a portion of information that has been formatted using a certain method. Data generally can be in the form of words or numbers on paper, bits or bytes in computer memory, or facts in a human brain [1]. Additionally, data exists everywhere in the universe. The amount of data that is available throughout the universe has no boundaries. With some facts, humans have existed since the dawn of human civilization. We deal with some data in our daily lives, whether for a purpose or not [2]. Any kind of data can be stored, such as a playlist with thousands of songs or the ability to recall someone's name. However, as human civilization has advanced, the data has already begun to rise rapidly. A novel approach to data manipulation, known as data analytics, was introduced to human civilization in the last century with the goal of simplifying life and boosting corporate profitability while dealing with massive amounts of data [3]. We received some data with a large volume, high speed, and wide diversity as a result of the widespread usage of the internet. That data was too big for the relational database to handle and process. As a result, Big Data, a new kind of data with distinct ideas and technology, was introduced. The following categories can be used to group the data according to the behaviour:

- *Structured Data:* The information that can be kept in a row-column structure in a relational database table. As the name implies, structured data has a certain structure, and the organization's data model defines that structure. The data model, or a model that will enable the storage, processing, and access of organizational data, is initially defined by the organization. The characteristics of the data that will be saved must be specified by the model. Among the data's attributes are its kind (numerical, alphabetic, name, date, etc.) and some limitations (data size, character count, etc.). The primary benefit of structured data is its ease of analysis and storage. Relational databases are the only efficient means of storing and processing structured data because of their high cost, storage space constraints, and processing methods. In essence, Structured Query Language (SQL) is used to manage structured data. A computer language called SQL is used to manage the data in RDBMSs [4].
- *Semi-Structured Data:* It is information that is structured but does not conform to the data models established for structured data. Relational databases and other data table formats cannot hold semi-structured data; instead, it must be saved in a particular kind of file with tags. The data must be kept in a hierarchy, and the tags or markers are separated according to certain semantic constraints. Since the advent of the web, this type of data has grown significantly, and many data formats and applications require information-sharing platforms like XML and JSON [4].
- *Unstructured Data:* Unstructured data is information that lacks a defined structure and, as a result, cannot be kept in a standard database's row-column format. Unstructured data is the reverse of structured data, as the name implies. As a result, it cannot be kept in database fields. Example: Web sites, text documents, images, audio files, and videos, to name a few. The amount of unstructured data is increasing so quickly these days that handling and analysing it is becoming increasingly challenging. Therefore, greater expertise with some cutting-edge technology is needed to analyse unstructured data [4].

This study is structured into multiple sections, each of which addresses a distinct facet of the subject. A literature review is presented in Section 2. Data Science, Data Analytics and Big Data are contrasted in Section 3. Different analytical tools are covered in Section 4, and papers main points are summarized in Section 5.

## II. LITERATURE REVIEW

Numerous applications, methods, tools and other resources have been used in studies by different researchers. To obtain understanding, a number of research studies have been examined and evaluated.

*Nti et al.* [5] provides a thorough mini-literature evaluation of machine learning in BDA. A total of 1512 published publications were found through a keyword search. Based on the study's suggested novel taxonomy, 140 articles were selected for screening. The results of the study demonstrate the widespread use of various machine learning techniques in BDA.

*Bousdekis et al.* [6] examined the topic of industrial quality research from the standpoint of data analytics. In particular, we looked at the body of existing research, clarified the Quality 4.0 study area, summarized the outcomes of the literature survey, and determined the research gaps and difficulties. They also suggested new study directions on top of these.

*Mišić et al.* [7] analyse how data analytics is now being used in three important areas of operations management: supply chain management, revenue management, and hospital operations. Next, determine which future strategies show promise.

*Moustaka et al.* [8] aims to "connect the pieces" between data science and smart cities (SC) by identifying the important topics, services, and methods utilized in SC data monitoring through a comprehensive literature study. The survey focuses on data mining and data gathering practices throughout several SC data cycles. A survey technique is utilized to reach both quantitative and semantically meaningful elements. These exemplary scenarios are described, and the outcomes are quite favourable for systemizing.

*Lee et al.* [9] Following the identification of each article's configurations, a presentation of the literature is given. As a result, we also look at the process-focused morphological matrix's applications and make recommendations for future lines of inquiry.

*Anitha et al.* [10] aims to provide an overview of the adoption of Data Analytics (DA) capabilities as part of a "next generation" architecture by working with sales data to construct a linear regression model. The study also examines the ways in which supply chain data can be stored, processed, managed, interpreted, and visualized using big data techniques.

*Oladeinde et al.* [11] use a review methodology to investigate many subjects, including the evolution of information technology throughout history, the revolutionary function of data analytics in decision-making, and governance challenges within the IT sector. This academic study examines the developing fields of data analytics and IT innovations in Nigerian businesses, emphasizing the strategic ramifications for corporate operations, governance issues, and their revolutionary power. The study aims to understand the intricacies and potential of IT and data analytics in transforming the corporate environment of the country, with a backdrop of Nigeria's dynamic economic landscape.

*Raji et al.* [12] explores the vast field of real-time data analytics (DA) in retail, providing a thorough analysis of methods used both domestically and internationally. Retail behemoths in the US have used real-time data analytics to learn more about consumer preferences, behaviour, and market trends than ever before. This study acknowledges challenges such integration complexity, data protection issues, and the requirement for qualified personnel. Notwithstanding these difficulties, there is no denying real-time data analytics' revolutionary influence on the retail industry. This study offers the dynamic and developing field of real-time data analytics in retail, shedding light on the tactics used by US and international companies to prosper in a period of swift technological advancement.

*Gončarovs et al.* [13] These days, the data flood problem has replaced the data scarcity problem. Customer relationship management (CRM) experts and marketers have access to extensive consumer behaviour data. The choice of suitable data analytics methods and the efficient use of these data in CRM procedures present the current problem. Hidden patterns in data can be found with the aid of data analytics techniques. This article's goal is to examine a few of the many analytics tools and methods that may be used to enhance the CRM process over time. To do this, a thorough literature review has been done. The review's findings emphasize the CRM procedures that are most commonly considered when discussing data analytics.

*Bonthu et al.* [14] analyses large amounts of data, has become more and more popular as a way to uncover hidden patterns, correlations, and other insights. Consequently, many data analysis software programs were developed. Prior to using the product, organizations need to know how it will fit into their broader company objectives. Because data analytics practitioners' needs are always changing, certain tools are becoming less significant while many new ones are entering the market. The most popular tools now on the market are reviewed in this article to help analytics practitioners choose

the appropriate solution for the job. The present study reviews seven popular tools.

*Alsubari et al.* [15] aims to create an intelligent system that uses the reviewer's sentiment scores and n-grams of the review text to identify fraudulent reviews on e-commerce sites. This study's suggested methodology employed for feature extraction and representation, as well as a review dataset for experimentation and data pretreatment techniques. However, a dataset gathered from the Trip Advisor website was used for training and testing, and four distinct supervised machine-learning approaches were used in the studies.

*Gudivada et al.* [16] offers a thorough and cohesive understanding of the foundations of data analytics. This research examines data science as the field that forms the basis of the most recent generation of data analytics solutions. Issues with data quality and data lifetime are described. The list includes open-source tools and resources for creating data analytics systems. Emerging trends in data analytics are highlighted in the chapter's conclusion.

*Nguyen et al.* [17] suggests an integrated framework and a set of common terminology for higher education. Researchers may find new contexts and fields of investigation by taking the framework into consideration. As a Gestalt-like activity, educational stakeholders may find the framework and the articulation of information analytics helpful when making decisions at the individual student, class, curriculum, school, and system levels.

*Settibathini et al.* [18] outlines prospective developments and future trends in AI-driven data analytics (DA). In addition to offering a plagiarism-free examination of the complex relationship between DA and artificial intelligence, this analysis and strategic review are an invaluable resource for researchers, executives, and policymakers. It also offers recommendations for businesses looking to effectively utilize these technologies.

All things considered, the literature shows how data analytics is becoming more and more recognized across a range of businesses. With an emphasis on predictive strategies and procedures that increase automation, efficiency, and accuracy, key findings emphasize the significance of selecting the appropriate tools and approaches to improve decision-making.

## III. STUDY OF DATA SCIENCE, DATA ANALYTICS AND BIG DATA

Using data to solve problems and get insightful information is the main goal of studying data science (DS), data analytics (DA), and big data (BD). Data analytics is concerned with using data to inform decisions, data science is focused on solving complicated problems and predictive modelling, and big data is focused on managing massive datasets using sophisticated tools. When combined, these domains enable businesses to use data to spur innovation and expansion. Table 1 shows a comparison of Big Data, Data Analytics, and Data Science.

Table 1: Comparison of Data science, Data analytics and Big data [19].

| **Data Science** | **Data Analytics** | **Big Data** |
|---|---|---|
| Data science involves breaking down large data sets and employing statistical, mathematical, and technological methods to identify meaningful patterns and trends. | The goal of DA is to offer operational insights into challenging business scenarios. | The term BD describes vast amounts of data of several kinds, including unstructured, semi-structured, and structured data. |
| Finding the truths buried in the intricate network of unstructured data is the responsibility of data scientists so that they can be utilized in business decision-making. | The idea of BD has been around for a while, and most firms now realize that they can apply analytics and gain a lot of value from it if they can collect all the data that enters their operations. | Numerous digital platforms, including social media, e-commerce websites, mobile devices, and the Internet, are used to generate this data. |
| Data scientists carry out the aforementioned task by creating models and heuristic algorithms that have important future applications. | Globally, data analytics has grown so rapidly that it is anticipated that the BD market will soon see a 50% increase in revenue. | Since its introduction, big BD has proven to be quite helpful as businesses have begun to recognize their value for a variety of business objectives. |
| Impact a variety of sectors, including as web development, digital advertisements, e-commerce, utilities, telecommunications, banking, and internet search. | Impact several sectors, such as retail, research, healthcare, energy management, financial analysis, and travel and transportation. | Impact on various sectors, such as operational analysis, fraud detection and analysis, retail, banking, and investments, as well as customer-centric applications. |

## IV. OVERVIEW OF ANALYTICAL TOOLS

All data analysis tools share the endless arguments over why their preferred programming language is superior, more sophisticated, quicker, more holy, etc. The debates about the finest statistical programming language in the data science community today seem to go on forever, with proponents of SAS, SPSS, R, Python, and other languages facing off against one another on all Internet forums. In order to compare various tools and determine which is the best, this overview

provides us with a basic understanding of each instrument. Below is a summary of the top five statistical tools for DA that are currently available.

- *R:* Revolution is an open-source software environment and programming language for statistical computation, data mining (DM), DA, and graphics. In addition to a vast array of statistical, machine learning (ML), and DM techniques, it has nearly hundreds of different "packages" that are publicly available. It is capable of producing a working ML application with an only forty lines of code. Its focus is on data analysis, graphical models, and better, more accessible statistics. The majority of R's applications have been in academics and research. RStudio is the IDE for R. R is more likely to be considered an interpreted language than an environment [20].
- *Python:* It is a popular high-level, general-purpose programming language. Programming styles supported by this language include object-oriented and structural. It is also possible to employ other styles. Python has "nice" syntax, which makes coding and debugging simple. The code's meaning is impacted by its indentation. When it comes to doing something new that has never been done before, Python is incredibly versatile. It can also be used by developers to script websites and other apps [20].
- *SPSS:* SPSS stands for Statistical Package for the Social Sciences. In 1979, SPSS jeopardized the tax-exempt status of the University of Chicago. In 2009, IBM paid US$1.2 billion to acquire SPSS. SPSS was intended to be easier to use than other statistical software such as S-Plus, R, or SAS. SPSS is a great tool for non-statisticians because of its easy-to-use interface and straightforward drop-down options. Like Excel, SPSS is popular outside of the data science field. In essence, SPSS is a statistical software that includes a variety of statistical tests, regression models, correlations, and factor analyses. In other words, SPSS will more than satisfy our demands because it is a versatile tool that allows for a large variety of analyses, data conversions, and output formats.
- *WEKA:* The GNU general public license governs the Waikato Environment for Knowledge Analysis, or Weka. In 1997, it was first made available. Weka is a Java-based collection of machine learning programs that originated at the University of Waikato. It has a vibrant community and is widely used in both business and academia. Weka has three graphical user interfaces: the Explorer, which facilitates exploratory data analysis and supports pre-processing, attribute selection, learning, and visualization; the Experimenter, which provides an experimental environment for testing and assessing machine learning algorithms; and the Knowledge Flow, which is used to visually design KDD processes and is inspired by a new process model [20].
- *SAS:* Statistical Analysis System is what SAS stands for. Jim Barr and Jim Goodnight are the creators of SAS. It started off as an agricultural research analysis study at North Carolina State University. As the need for these software features increased, SAS was established in 1976 to serve clients across a range of sectors, including banks, pharmaceutical corporations, government agencies, and academic institutions. Over the following few decades, SAS—both the software and the business—grew and prospered. Because the software could run on various platforms, its development reached unprecedented heights in the industry. SAS is quite good at handling data. Its software upgrades are well vetted because they are released in a controlled setting. SAS is a costly option [20].

## V. CONCLUSION AND FUTURE SCOPE

To sum up, data is the cornerstone of well-informed decision-making, propelling advancement and creativity in a variety of fields. BD, DS, and DA all play different but related roles that highlight how crucial they are for drawing conclusions and developing data-driven strategies. Through the efficient utilization of many forms of data and sophisticated instruments, businesses can unleash substantial possibilities for expansion and prosperity. The knowledge areas and competencies that characterize DA, DS, and BD are examined in this study. It looks at the many kinds of data as well as the tools and methods used for processing and analyzing them. By comprehending these components, this research seeks to draw attention to the connections and differences between different domains, providing a thorough knowledge of their functions in contemporary information systems. Future research should focus on incorporating cutting-edge technologies like AI, ML, and IoT to enhance data processing and analysis. It will also be necessary to look into scalable frameworks for huge data management and predictive model building in order to address emerging problems and expanding data-driven applications across industries.

## REFERENCES

[1] Sharma, K., Shetty, A., Jain, A. and Dhanare, R.K., 2021, January. A Comparative Analysis on Various Business Intelligence (BI), Data Science and Data Analytics Tools. In 2021 International Conference on Computer Communication and Informatics (ICCCI) (pp. 1-11). IEEE.

[2] Radovilsky, Z., Hegde, V., Acharya, A. and Uma, U., 2018. Skills requirements of business data analytics and data science jobs: A comparative analysis. Journal of Supply Chain and Operations Management, 16(1), pp.82-101.

[3] Sahu, S.K., Jacintha, M.M. and Singh, A.P., 2017, May. Comparative study of tools for big data analytics: An analytical study. In 2017 International Conference on Computing, Communication and Automation (ICCCA) (pp. 37-41). IEEE.

[4] Das, A.C., Pani, S.K. and Mohanty, S.N., 2015. A Comparative Study on Data Analytics and Big Data Analytics. International

Journal of Computer Science and Information Technology Research, 4(1), pp.67-75.
[5] Nti, I.K., Quarcoo, J.A., Aning, J. and Fosu, G.K., 2022. A mini-review of machine learning in big data analytics: Applications, challenges, and prospects. Big Data Mining and Analytics, 5(2), pp.81-97.
[6] Bousdekis, A., Lepenioti, K., Apostolou, D. and Mentzas, G., 2023. Data analytics in quality 4.0: literature review and future research directions. International Journal of Computer Integrated Manufacturing, 36(5), pp.678-701.
[7] Mišić, V.V. and Perakis, G., 2020. Data analytics in operations management: A review. Manufacturing & Service Operations Management, 22(1), pp.158-169.
[8] Moustaka, V., Vakali, A. and Anthopoulos, L.G., 2018. A systematic review for smart city data analytics. ACM Computing Surveys (cSuR), 51(5), pp.1-41.
[9] Lee, C., 2021. A review of data analytics in technological forecasting. Technological Forecasting and Social Change, 166, p.120646.
[10] Anitha, P. and Patil, M.M., 2018. A review on data analytics for supply chain management: a case study. International Journal of Information Engineering and Electronic Business, 10(5), p.30.
[11] Oladeinde, M., Hassan, A.O., Farayola, O.A., Akindote, O.J. and Adegbite, A.O., 2023. Review of it innovations, data analytics, and governance in nigerian enterprises. Computer Science & IT Research Journal, 4(3), pp.300-326.
[12] Raji, M.A., Olodo, H.B., Oke, T.T., Addy, W.A., Ofodile, O.C. and Oyewole, A.T., 2024. Real-time data analytics in retail: A review of USA and global practices. GSC Advanced Research and Reviews, 18(3), pp.059-065.
[13] Gončarovs, P., 2017. Data analytics in CRM processes: A literature review. Information Technology and Management Science, 20(1), pp.103-108.
[14] Bonthu, S. and Bindu, K.H., 2017. Review of leading data analytics tools. International Journal of Engineering & Technology, 7(3.31), pp.10-15.
[15] Alsubari, S.N., Deshmukh, S.N., Alqarni, A.A., Alsharif, N., Aldhyani, T.H., Alsaade, F.W. and Khalaf, O.I., 2022. Data analytics for the identification of fake reviews using supervised learning. Computers, Materials & Continua, 70(2), pp.3189-3204.
[16] Gudivada, V.N., 2025. Data analytics: fundamentals. In Data analytics for intelligent transportation systems (pp. 27-66). Elsevier.
[17] Nguyen, A., Gardner, L. and Sheridan, D., 2020. Data analytics in higher education: An integrated view. Journal of Information Systems Education, 31(1), p.61.
[18] Settibathini, V.S., Kothuru, S.K., Vadlamudi, A.K., Thammreddi, L. and Rangineni, S., 2023. Strategic analysis review of data analytics with the help of artificial intelligence. International Journal of Advances in Engineering Research, 26, pp.1-10.
[19] Nadikattu, R.R., 2020. Research on data science, data analytics and big data. International Journal Of Engineering, Science And, 9(5), pp.99-105.
[20] Bansal, A. and Srivastava, S., 2018. Tools used in data analysis: A comparative study. International Journal of Recent Research, 5(1), pp.15-18.

## ABOUT THE AUTHORS

**Dr. Neha Singh** is currently working as an Assistant Professor in Ajay Kumar Garg Engineering College, Ghaziabad. She received the M. Tech. degree in Information Technology from the Madan Mohan Malaviya University of Technology, Gorakhpur, in 2020. She received her Ph.D. degree from Madan Mohan Malaviya University of Technology, Gorakhpur, in 2024. Her research interests include Machine Learning and Deep Learning.

**Kriti Mishra** is currently working as an Assistant Professor in Ajay Kumar Garg Engineering College, Ghaziabad since August 2024. She received the M. Tech. degree in Computer Science Engineering. She believes in exploring new emerging technology like Machine Learning and Deep Learning.

**Vibha Srivastava** is currently working as an Assistant Professor in Ajay Kumar Garg Engineering College, Ghaziabad science August 2024. She received the M. Tech. degree in Computer Science Engineering from the AKTU, Lucknow. She believes in exploring new emerging technology like Machine Learning and Blockchain.

**Jyoti Srivastava** is currently working as an Assistant Professor in Kashi Institute of Technology. She has published 2 SCIE and 2 ESCI papers in International Journals, and more than 15 papers in National and International Conferences. Her research interests include Networks, Machine Learning, Natural Language Processing and Video and Image Processing.