IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 3, January 2023
Impact Factor: 7.301

Big Data Analytics: A Literature Review Paper

Miss. Pawar Sakshi[1] and Mrs. Gauri Malwadkar[2]
Student, M.Sc. I.T., I. C. S. College, Khed, Ratnagiri, Maharashtra, India[1]
Asst. Prof., Department of I.T., I. C. S. College, Khed, Ratnagiri, Maharashtra, India[2]

Abstract: *In the age of information, decision-makers now have access to an enormous amount of data. The term "big data" refers to datasets that are not only large but also extremely diverse and fast, making it challenging to deal with them using conventional methods and tools. In order to handle and extract value and knowledge from these datasets, it is necessary to investigate and provide solutions due to the rapid growth of such data. In addition, decision-makers need to be able to draw useful conclusions from such a wide range of data that is constantly changing, such as data from daily transactions, customer interactions, and social network data. Using big data analytics, which is the application of advanced analytics methods to big data, such value can be provided. The purpose of this paper is to look at some of the various analytics tools and methods that can be used with big data, as well as the opportunities provided by using big data analytics in different decision domains.*

Keywords: Big Data, Data Mining, Analytics, Decision Making

I. INTRODUCTION

Imagine a world with no storage for data; a location where every aspect of a person or organization that can be documented, every transaction that is carried out, or every detail is lost immediately after use. As a result, organizations would be unable to provide new opportunities and advantages, conduct in-depth analyses, or extract valuable knowledge. Anything from the names and addresses of customers to the products available, the items purchased, the employees hired, and so on has become necessary for maintaining daily continuity. Any business is built on data, which is its foundation.

Consider the sheer volume of data and information available today thanks to technological advancements and the internet. Huge amounts of data are now easily accessible as storage capacities and data collection methods increase. In order to extract value, an increasing amount of data must be stored and analyzed every second. Additionally, as data storage has become less expensive, businesses must maximize the value of the vast amounts of data they store.

The size, variety, and rapid change of such data necessitate a novel kind of big data analysis, in addition to novel storage and analysis strategies. The massive amounts of big data must be thoroughly analyzed, and relevant data must be extracted.

This paper makes a contribution by analyzing the existing literature on big data analytics. As a result, a number of the various big data tools, techniques, and technologies that can be used are discussed, as well as their applications and opportunities in a variety of decision domains.

For the purpose of our research, the literature was chosen for its novelty and discussion of significant big data-related topics. The majority of the literature focusing on big data dates from 2011 to 2013, with publication years ranging from 2008 to 2013.This is because big data has recently received a lot of attention. In addition, the majority of the research in our corpus comes from leading industry corporations' white papers, conferences, and journals. The majority of papers discussing big data analytics, its tools and methods, and its applications were found to be conference papers and white papers due to journals' lengthy review processes. While big data analytics is being studied in academia, most of the new technologies and developments in industry were discussed in industry papers.

II. BIG DATA ANALYSIS

Datasets that become so large that traditional database management systems become difficult to work with are now referred to as "Big Data." They are large data sets that can't be captured, stored, managed, or processed in a reasonable amount of time using standard software tools and storage systems [12].

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 3, January 2023
Impact Factor: 7.301

The sizes of big data are constantly increasing, and a single data set currently contains anywhere from a few dozen terabytes (TB) to many petabytes (PB).As a result, big data challenges include capture, storage, search, sharing, analytics, and visualization. Today, businesses are looking into huge amounts of highly detailed data in order to learn things they didn't know before [17].

As a result, advanced analytical methods are utilized on large data sets in big data analytics. Business change is revealed and capitalized on by analytics based on large data samples. However, managing a larger set of data becomes more challenging [17].

We will begin this section by discussing the characteristics and significance of big data. Naturally, larger and more complex data sets that necessitate real-time or near-real-time capabilities can often benefit businesses; However, as a result, new data architectures, analytical techniques, and tools are required. As a result, the subsequent section will focus on the big data analytics tools and methods in particular, beginning with the storage and management of big data and moving on to the analysis of big data. After that, it discusses a few of the various big data analyses that have become increasingly popular with big data.

2.1 Characteristics of Big Data

New technical architectures, analytics, and tools are required to enable insights that unlock new sources of business value from big data, which is data whose scale, distribution, diversity, and/or timeliness necessitate the use of these new technologies. Big data has three main characteristics: the three Vs, or lume, variety, and velocity. The size and extent of the data determine its volume. The frequency with which data is created or the rate at which it changes is referred to as velocity. Finally, variety encompasses the various data formats, types, and applications, in addition to the various methods of data analysis [9].

The primary characteristic of big data is its volume. In addition to the number of records, transactions, tables, or files, the size of big data can be measured in TBs or PBs. In addition, the fact that big data comes from a wider variety of sources than ever before—such as logs, clickstreams, and social media—is one of the factors that makes it so enormous. When these sources are used for analytics, semi-structured data like eXtensible Markup Language (XML) or Rich Site Summary (RSS) feeds and unstructured data like text and human language join common structured data. Additionally, there is data, which is difficult to classify because it comes from audio, video, and other devices. A data warehouse can also be used to extract multidimensional data to provide historical context for big data. As a result, variety is just as important as volume in big data.

Furthermore, the velocity or speed of big data can be used to define it. This basically refers to the frequency with which data is generated or delivered. Streaming data, which is collected in real time from websites, is at the cutting edge of big data [17].The addition of a fourth V, or veracity, has been discussed by some organizations and researchers. The quality of the data is the focus of veracity. Due to data inconsistency, incompleteness, ambiguity, latency, deception, and approximations, this classifies big data quality as either good or undefined [22].

2.2 Big Data Analytics Tools and Methods

There is a growing need for faster and more effective methods for analyzing such data as a result of the development of technology and the increasing volumes of data that are entering and leaving organizations on a daily basis. It is no longer sufficient to have a mountain of data available to make effective decisions at the right time.

Traditional data management and analysis methods and infrastructures no longer work well with such data sets. As a result, big data analytics-specific tools and methods, as well as the infrastructure needed to store and manage such data, are required. As a result, the rise of big data has an impact on all aspects, including data collection, processing, and ultimately extracted decisions.

As a result, the Big – Data, Analytics, and Decisions (B-DAD) framework, which incorporates the methods and tools of big data analytics into the decision-making process, was proposed [8].The framework connects the various stages of the decision-making process to the various tools for big data storage, processing, analytics, and visualization and evaluation. As a result, the changes brought about by big data analytics can be seen in three main areas: big data analytics, which can be used for knowledge discovery and informed decision making, data and analytics processing, and big data architecture. In this section, we will go over each area in greater detail. This section, on the other hand,

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 3, January 2023
Impact Factor: 7.301

focuses on providing a general idea rather than a list of all potential opportunities and technologies because big data is still evolving as an important field of research and new findings and tools are constantly being developed.

2.3 Big Data Analytics

Nowadays, people don't just want to collect data; they also want to know what the data means, why it matters, and how to use it to help them make decisions. Algorithms are used to analyze sets of data and extract both known and useful patterns, relationships, and information in data analysis [1].In addition, important relationships between stored variables and previously unknown, useful, valid, and hidden patterns and information can be discovered using data analytics on large data sets. As a result, analytics have had a significant impact on research and technologies because decision makers are increasingly interested in gaining a competitive advantage by learning from previous data [21].

Additional analyses have become common with big data in addition to association rules, clustering, classification and decision trees, regression, and some of the most common advanced data analytics techniques.

For instance, social media has recently gained prominence for content sharing and social networking. However, social media websites produce a vast amount of content that is largely untapped. Social media analytics, on the other hand, can be used to look at such data and make insightful inferences and predictions [2].Utilizing informatics frameworks and tools to collect, monitor, summarize, analyze, and visualize social media data is the foundation of social media analytics. In addition, in addition to what people share on social media websites, social media analytics makes it easier to comprehend the responses and conversations that people have in online communities. It also makes it easier to extract useful patterns and intelligence from their interactions [24].

Social Network Analysis (SNA), on the other hand, focuses on the connections between social entities as well as the patterns and implications of these connections [23].In order to comprehend what facilitates the flow of knowledge between interacting parties, such as who knows who and who shares what knowledge or information with whom and how [19], an SNA maps and measures both formal and informal relationships.

SNA, on the other hand, attempts to capture the social relationships and patterns that exist between people's networks, in contrast to social media analysis. Social media analysis, on the other hand, looks at what people say on social media to find useful patterns, information about the users, and feelings. Traditionally, text mining or sentiment analysis are used to accomplish this.

On the other hand, text mining is used to look at a document or set of documents to figure out what the information means and what it says about the content. Due to the fact that the majority of information stored, excluding audio, video, and images, consists of text, text mining has become extremely important in today's world. Text has unique characteristics that basically follow a non-relational form [18], whereas data mining deals with structured data.

In addition, the exponential growth of online opinion data, such as product reviews, blogs, and social data from social media sites like Twitter and Facebook, is increasing the importance of sentiment analysis, also known as opinion mining. Text mining makes it possible to conduct sentiment analysis, which focuses on deciphering emotions from subjective text patterns. It is useful for classifying viewpoints as positive or negative because it identifies the opinions and attitudes of individuals regarding particular subjects. Sentiment analysis makes use of text analytics and natural language processing to identify and extract information by locating words that are indicative of a sentiment and relationships between words in order to accurately describe sentiments.

2.4 Big Data Analytics and Decision Making

The significance of big data for decision-makers lies in its capacity to provide value information and information on which to base decisions. Over the years, the managerial decision-making process has been a significant and extensively researched subject in research.

Decision-makers are increasingly relying on big data as an asset. Scanners, mobile phones, loyalty cards, the internet, and social media platforms all provide the opportunity to provide organizations with significant benefits from large volumes of highly detailed data. This is only possible if the data are properly analyzed to reveal useful insights. This makes it possible for decision-makers to take advantage of the opportunities presented by the abundance of historical and real-time data generated by supply chains, production processes, customer behaviors, and other activities.4].

In addition, businesses are currently accustomed to analyzing internal data like inventory, sales, and shipments.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 3, January 2023
Impact Factor: 7.301

However, there is a growing demand for analyzing external data, such as supply chains and customer markets, and big data can provide cumulative value and knowledge. It becomes necessary to make more informed decisions based on drawing meaningful inferences from the data as the sizes and types of unstructured data available increase [7].

As a result, the B-DAD framework, which incorporates big data tools and techniques into the decision-making process, was created [8].The purpose of this framework is to improve the quality of the decision-making process when dealing with big data. The intelligence phase of the decision-making process is the first step in gathering data from both internal and external data sources that can be used to identify opportunities and problems. The sources of big data must be identified during this phase, and the data must be gathered from various sources, processed, stored, and transferred to the end user. Following the definition of the data sources and types of data required for the analysis, the selected data is acquired and stored in any of the big data storage and management tools previously discussed. After the big data is acquired and stored, it is organized, prepared, and processed across a high-speed network using ETL/ELT or big data processing tools, which have been discussed in the preceding sections. This is accomplished using big data processing tools.

2.5 Customer Intelligence

Retail, banking, and telecommunications are just a few of the industries that can greatly benefit from big data analytics' potential for customer intelligence. Big data has the potential to increase transparency and make pertinent data more quickly and easily accessible to stakeholders [14].Customers can be profiled and segmented based on a variety of socioeconomic characteristics using big data analytics, which can also improve customer satisfaction and retention [4].They may be able to recognize sales and marketing opportunities and make more informed marketing decisions as a result of this [17]. They may also be able to market to various segments based on the preferences of those segments. Additionally, businesses can learn from their customers' social media posts what they like and dislike. Companies can anticipate when customers are turning against them or switching to other products before they happen by conducting sentiment analysis on this data [7].

Additionally, organizations can use SNAs to identify influential people, monitor brand perceptions among customers, and respond to trends through direct marketing. Predictive models for customer behavior and purchase patterns can also be built using big data analytics [4] to improve overall profitability. Real-time micro-segmentation of customers, for example, is one of the more advanced big data techniques that even established businesses are beginning to use to target promotions and advertising [14].As a result, big data analytics can help businesses by enabling more targeted social influencer marketing, defining and predicting trends from market sentiments, and analyzing and comprehending customer churn and other behaviors [17].

2.6 Supply Chain and Performance Management

Big data analytics can be used to forecast changes in demand and match their supply in supply chain management. Manufacturing, retail, transportation, and logistics can all reap the benefits of this. Organizations can automate replenishment decisions by analyzing stock utilization and geospatial data on deliveries. This will cut down on lead times, costs, delays, and process interruptions. Analyzing supplier performance data can also be used to make decisions about switching suppliers based on quality or price competitiveness. Alternate pricing scenarios can also be tested right away, which can result in lower inventories and higher profit margins [4].As a result, big data can help improve planning and forecasting and reveal the underlying causes of costs [17].

2.7 Quality Management and Improvement

Big data can be used for quality management, particularly in the manufacturing, energy and utilities, and telecommunications industries, to improve the quality of goods and services provided and thus increase profitability and decrease costs. For instance, predictive analytics on big data can be used in the manufacturing process to minimize performance variability and provide early warning alerts to prevent quality issues. This can cut down on scrap rates and speed up time to market because it can save a lot of money to catch production delays before they happen [4].Big data analytics can also improve manufacturing lead generation [17].In addition, managers may be able to make decisions regarding quality management more quickly with the assistance of real-

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 3, January 2023
Impact Factor: 7.301

time data analyses and machine log monitoring. In addition, big data analytics may make it possible to forecast bandwidth in response to customer behavior and monitor network demand in real time.

Additionally, by communicating and integrating patient data across various departments and institutions while maintaining privacy controls, healthcare IT systems can enhance both the efficiency and quality of care [4].Electronic health records can be analyzed to create a huge dataset that can be used to predict and compare treatments and outcomes as well as to improve patient continuity of care. As a result, there is an opportunity to mine the de-identified patient information for evaluating the quality of healthcare, disease management, and health services [22] with the growing use of electronic health records and advancements in analytics tools.

2.8 Risk Management and Fraud Detection

In the area of risk management, big data analytics can be beneficial to industries like insurance and investment or retail banking. Big data analytics can assist investors in making investment decisions by weighing the likelihood of gains against the likelihood of losses, which is important for the financial services industry because risk assessment and management is a crucial aspect of the industry. For a more comprehensive and dynamic evaluation of risk exposures, internal and external big data can also be analyzed [4].As a result, big data can help businesses by making it possible to quantify risks [17].Additionally, enterprise-wide risk profiles can be created by combining department-specific risk profiles with high-performance analytics. Because decision-makers receive a comprehensive view of the various risk types and their interrelationships, this can assist in risk mitigation [4].

By increasing the capacity to scale and capture the necessary data, new big data tools and technologies can manage the exponential growth of network-produced data and reduce database performance issues. Organizations can now incorporate multiple data streams and automated analyses to defend themselves against cyber and network attacks thanks to advances in cyber analytics and data intensive computing solutions [22].

III. CONCLUSION

In this study, we looked at the cutting-edge subject of big data, which has recently attracted a lot of attention due to the perceived unprecedented opportunities and benefits it offers. Every day, vast quantities of high-velocity data are generated in the information age we are living in. These data contain intrinsic details and patterns of hidden knowledge that must be extracted and utilized. By applying advanced analytic techniques to big data and revealing hidden insights and valuable knowledge, big data analytics can be used to leverage business change and improve decision making.

As a result, a literature review was conducted to provide an analysis of the research's big data analytics concepts and their significance to decision-making. As a result, big data, its characteristics, and its significance were discussed. In addition, some of the methods and tools for big data analytics in particular were examined. As a result, big data analytics processing and storage were described in detail. The various advanced data analytics methods were also discussed in greater detail.

This kind of analytics can be applied to big data, and valuable information can be extracted and used to support informed decisions and improve decision-making. As a direct consequence of this, we investigated a few of the various areas in which big data analytics can support and facilitate decision-making. It was discovered that big data analytics can open up a plethora of opportunities in a variety of applications and fields, including supply chain management, fraud detection, and customer intelligence. Additionally, its advantages can benefit a variety of sectors and industries, including manufacturing, healthcare, retail, telecom, and others.

As a result, people and businesses now have a better understanding of the various big data tools, techniques, and technologies that can be used. This provides users with an idea of the necessary technologies and developers with an idea of what they can do to provide more advanced solutions for big data analytics that support decision-making. As a result, the contribution of big data analytics to decision-making was shown.

We believe that big data analytics is very important in this day and age of excessive data flow because it can offer decision-makers unexpected insights and benefits in a variety of areas. Big data analytics has the potential to serve as a foundation for scientific, technological, and humanitarian advancements if properly utilized.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 3, January 2023
Impact Factor: 7.301

REFERNCES

[1]. Adams, M.N.: Perspectives on Data Mining. International Journal of Market Research52(1),11–19(2010)

[2]. Asur, S., Huberman, B.A.: Predicting the Future with Social Media. In: ACM InternationalConference on Web Intelligence and Intelligent Agent Technology, vol. 1, pp. 492–499(2010)

[3]. Bakshi, K.: Considerations for Big Data: Architecture and Approaches. In: Proceedings oftheIEEE Aerospace Conference, pp.1–7(2012)

[4]. Cebr:Dataequity,Unlockingthevalueofbigdata.in:SASReports,pp.1–44(2012)

[5]. Cohen, J., Dolan, B., Dunlap, M., Hellerstein, J.M., Welton, C.: MAD Skills: New Analy-sis Practices for Big Data. Proceedings of the ACM VLDB Endowment 2(2), 1481–1492(2009)

[6]. Cuzzocrea, A., Song, I., Davis, K.C.: Analytics over Large-Scale Multidimensional Data:The Big Data Revolution! In: Proceedings of the ACM International Workshop on DataWarehousingandOLAP,pp.101–104(2011)

[7]. Economist Intelligence Unit: The Deciding Factor: Big Data & Decision Making. In:Capgemini Reports,pp.1–24(2012)

[8]. Elgendy, N.: Big Data Analytics in Support of the Decision Making Process. MSc Thesis,German UniversityinCairo,p.164(2013)

[9]. EMC: Data Science and Big Data Analytics. In: EMC Education Services, pp. 1–508(2012)

[10]. He, Y., Lee, R., Huai, Y., Shao, Z., Jain, N., Zhang, X., Xu, Z.: RCFile: A Fast and Space-efficient Data Placement Structure in MapReduce-based Warehouse Systems. In: IEEEInternational Conferenceon DataEngineering(ICDE),pp.1199–1208 (2011)