Big Data Application Analysis: A Review

M.U. Issath Jahan[1], M.S. Faathima Fayaza[2]

[1,2]*Department of Information Technology, South Eastern University of Sri Lanka*

[1]issathjahan1996@seu.ac.lk, [2]fayaza@seu.ac.lk

Abstract

The term "big data" describes datasets that are not only large but also have a high level of diversity and velocity, making it challenging to manage them with conventional tools and methods. Through enhancing decision-making and vision searching, the big data explosion is reshaping lifestyles in terms of working and thinking. This paper conducts an analysis of recent research and studies projects in a variety of sectors that make use of big data. This paper studies 45 previously published articles and conducted a systematic literature review. According to the survey, many fields achieve various benefits, mainly when they apply big data technology. Further, selected algorithms perform better with specific domain data. This paper summarizes the different techniques used in the various domains and their benefits. Also, the article discusses the limitations in this study and limitations in big data applications.

Keywords: Big data analysis, Big data applications, Big data technologies

I. INTRODUCTION

Big Data is becoming a global trend. Although it does not yet have a widely accepted academic or scientific definition, it continues increased commercial expansion for its surrounding industries and related study fields (Hernández-Leal, Duque-Méndez and Moreno-Cadavid, 2017) (Soomro et al., 2019). Big data analytics is seeing patterns, trends, and correlations in massive amounts of unprocessed data in order to make data-driven choices. These processes use well-known statistical analysis techniques, such as clustering and regression, to bigger datasets utilizing contemporary technologies. NASA scientists invented the phrase Big Data in 1997 to characterize the challenge of presenting data sets that are too huge to fit in a computer's main memory, restricting analysis of the data set as a whole (Austin and Kusumoto, 2016). It is particularly formidable when managing large amounts of data and has had much more success in various fields (Li *et al.,* 2019).

The word "big data" refers to data sources that are too large or complicated, and conventional data processing methods are insufficient to process them. It is crucial to consider the data's size, complexity, and velocity in big data. The big data set is expanding in size at an exponential rate. This data is too big, too unstructured, or too "raw" compared to traditional data. In relational databases, processing methods employed still do not worry about the amount of data but are rather concerned the quality of the data (Gorodov and Gubarev, 2013) (Huda et al., 2016). The amount of data in one big data collection currently ranges from a few dozen terabytes (TB) to many petabytes (PB). Because of this, gathering, storing, searching, sharing, analyzing, and displaying massive data can be challenging. Enterprises are currently examining vast amounts of highly detailed data in order to learn something they did not know previously (Du et al., 2021). A Variety of data is generated from multiple sources such as social media, machine logs, sensors, transactional data, etc.

The demand for big data processing technology has increased in the age of big data. Big data brings a quick and significant rise in data resources and a greater difficulty in extracting useful information that customers require (Tulasi, 2013). Processing data is becoming increasingly more complicated. Traditional data analysis techniques have glaring flaws, which raises new demands on big data-based resource services for digital libraries (Li *et al.,* 2019) (Ying, Chan and Qi, 2020).

This paper aims to give an analysis of the big data analytics research that is currently accessible. Some of the numerous big data tools, approaches, and technologies and their potential in various decision fields are examined. In this review, the authors discussed what big data is, what the field big data is used for, how to analyze big data applications, and what obstacles big data faces.

II. LITERATURE REVIEW

Willis et al. discuss the ethics and analytics of big data in higher education (Willis, Campbell and Pistilli, 2013). This paper discusses ethical questions such as the "role of big data in higher education", 'role of big data in student experience", "application of big data in student retention" and "what is the impact on successful outcomes". Authors conclude that the actual issue in higher education when applying big data is that statistical likelihood within an academic prediction matrix has significant ramifications for both organizations and specific students. (Kamilaris, Kartakoullis and Prenafeta-Boldú, 2017) discuss big data in agriculture and smart farming using big data.

Using new healthcare-specific big data technologies, data from numerous sources studies to understand the industry best practices and give high-quality insights. Most countries, especially the United States, use big data to enhance their medical facilities (Batarseh and Latif, 2016). Wang et al. (Wang et al., 2018) discuss big data analytics in healthcare. This paper demonstrates a link between big data analytics capabilities, IT-enabled transformation practices, benefit dimensions, and business values through a big data analytics-enabled transformation model constructed from a practice-based approach. The growing requirement for healthcare management to handle the influx of clinical data supporting evidence-based medical practice is recognized as a potential solution in big data analytics (Wang et al., 2018).

Big data is a term used to represent the increasing expansion and accessibility of organized and unorganized data. It may be just as significant to industry and society as the World wide web has been (Singh Jain et al., 2017). Big data analytics using Deep learning performs exceptionally well nowadays using huge data and the introduction of potent computer hardware like the GPU. It can fully use enormous volumes of unprocessed data and fully automatically find abstract knowledge. Voice recognition, picture categorization, and other industries have effectively used great learning capacity (Peng et al., 2017).

Govindan et al. (Govindan *et al*., 2018) discuss big data in logistic and supply chain frameworks by applying the capability maturity model. Researchers used Twitter data to analyze the food supply chain management problems. Further, they used a text analytics method that uses clustering techniques, support vector machines, and multiscale bootstrap resampling to analyze the content of the Twitter data. The paper suggests that a big cluster of phrases can help decision-makers learn how to improve different parts of the logistics and food distribution chain (Govindan et al., 2018).

Data is regarded as a competitive resource and a new way to provide value for organizations, and the notion of big data is receiving a lot of attention in both the industry and the scholarly literature (Müller and Jensen, 2017). The quantity of data that can now be collected and used is increasing in today's smart cities. The ability to quickly collect, analyze, and store large amounts of data from various types of quantitative and qualitative domain-specific information sources has been made possible by recent advancements in hardware and software components, including social media, the IoT, monitoring devices, mobile technologies, data storage, and cloud computing (Iqbal et al., 2020).

Big data applications enable knowledge service providers to imbibe, procedure, analyze, and distribute content, transforming enormous amounts of data into insightful and compelling understandings. However, it is important to note that big data poses challenges to the methods and means of handling and analyzing data (Maqsood Ahmad Sandhu, Ahm Shamsuzzoha, 2018). Big data are data that, due to their size, complexity, and difficulty, demand novel management strategies, processing methods, algorithms, and analyses. Data collection, processing, storage, analysis, and visualization problems that were initially quantitative problems with data become qualitative problems as the scale of the data exceeds a certain threshold (Lee and Yoon, 2017).

III. METHODOLOGY

This study was conducted using a qualitative approach known as systematic literature review, based on earlier research and review papers over the last five years. The collected data were analyzed qualitatively to examine Big data's Fields, Technologies, and Purpose.

Table 01: Summary of big data application fields and techniques

References	Application Domain	Tools/Techniques and Data used	Purpose/Benefits
(Kamilaris, Kartakoullis and Prenafeta-Boldú, 2017), (Guo and Wang, 2019),(Horita *et al.*, 2017)	Agricultural	Big Data sources - Weather stations, Remote sensing (satellites, synthetic aperture radar, airplanes), geospatial data, historical datasets (land characterization and crop phenology, rainfall and temperature, Ground sensors (salinity, electrical conductivity, moisture), cameras (optical) **Techniques for big data analysis** - Machine learning (scalable vector machines, K-means clustering, random forests, extremely randomized trees), statistical analysis, modelling, cloud platforms, MapReduce analytics, GIS geospatial analysis, NDVI vegetation indices	Improve the accuracy of analyzing and forecasting disaster management, improve farmers productivity, weather forecasting
(Lee and Yoon, 2017),(Stieb, Boot and Turner, 2017),(Bofill-De Ros *et al.*, 2019), (Luo *et al.*, 2016),(Alyass, Turcotte and Meyre, 2015),(Tetko *et al.*, 2016)	Medical	Big Data sources – clinical data, sensors data, wearable medical devices data, DNA/RNA sequence data, biological image data **Techniques for big data analysis –** Machine learning (Decision trees, logistic regression, naive Bayesian approaches, Bayesian networks), Data mining, Neural network, Pattern Recognition, Natural Language Processing	personalized medicine. early detection of diseases(heart disease, cancer), identification of chronic diseases,
(Yan *et al.*, 2018),(Zeng, 2015),(Xie, Zhou and Li, 2016)	Transport	Big Data sources – IoT sensors data, Automatic data(automated passenger counts (APCs), camera video, automated vehicle location (AVL)), Global Positioning System (GPS) **Techniques for big data analysis –** machine learning, data mining, crowdsourcing	Traffic analysis to identify roads with a high risk of accident, transport route optimization
(Bajpai and Mani, 2017),(Willis, Campbell and Pistilli, 2013),(Kim and Ahn, 2016),(Zhu *et al.*, 2019)	Education	Big Data sources - social networking sites (like Facebook, Twitter, Blogs) Course Management systems (CMS), Learning Management System (LMS) and physical world data like library usage **Techniques for big data analysis –** machine learning, data mining, crowdsourcing	Prediction of performance of students, improve the resource management, providing feedback, course recommendation, and students behavior analyze.
(Muliawaty *et al.*, 2019),(Yu and Zhou, 2019),(Jordan, 2014)	Bureaucracy	Big Data sources – social media(Twitter, Facebook, blog and news website) **Techniques for big data analysis-** classification technique for sentiment analysis (Naive Bayes algorithm, Decision Tree, Artificial Neural Network)	Big data can used by local government agencies to enhance administrative and public services by knowing the public opinions.
(Padma and Ahn, 2020)	Tourism	Big Data sources - Satellite images, Geo-tagged images, conventional map, Google Analytics **Techniques for big data analysis –**data mining, crowdsourcing, Statistical and spatial analysis, Regression analysis, Sentiment analysis, Content analysis	Forecast tourist's arrival and predict the tourist volume, identify the potential area of tourism, identify which places are popular among different nationalities
(Du, Liu and Lu, 2021),(Ouyang, Wu and Huang, 2018),(Ishika and Mittal, 2021)	Security	Big Data sources – IoT devices data, sensor data **Techniques for big data analysis –** machine learning, data mining, crowdsourcing	Used to enhance the neural network's performance in order to increase the precision of early warning and Internet credit prevention
(Zhang, Zhan and Yu, 2017)	Business	Big Data sources – Sensor data (detect location in store), transaction logs **Techniques for big data analysis –** machine learning, data mining, crowdsourcing	Big data minimize resource waste and the inventory of the automotive industry. Reduce inventory and operational cost

A keyword-based search for articles and conference papers was conducted in the first stage using the databases IEEE, Emerald, Sage, and ScienceDirect and works indexed in Google Scholar.

The keywords "Big data Application" and "Big data technologies analytics" from scholarly publications from journals and conference proceedings were found using Google Scholar. The possessed publications included qualitative investigations, and the search period was set from 2017 to 2022.

A. Criteria for selecting and excluding articles

A first assessment of the recovered records was carried out by one of the writers. After analyzing individual titles and abstracts, duplicate articles were deleted, and more records were discarded. The included studies were then examined by a second author, who assessed the full-text papers or eligibility.

The authors selected 45 research articles from 75 based on the following criteria and all papers published in conferences or journals. The authors believed these two venues were more likely to contain current and relevant scientific papers related to this study.

- Majority of the Published papers between 2017 and 2022
- Only full papers
- Peer-reviewed papers
- Articles were an open access

B. Research Question

The following are the research questions (RQ) used in this study to collect data analysis.

- Which fields are most frequently utilized in big data applications?
- What emerging techniques support big data technologies across various domains?

IV. RESULTS AND DISCUSSION

Big data has a wide range of possible effects across many disciplines. This study aimed to understand the Techniques and the purposes for implementing Big data technology.

In the agriculture domain, there are some opportunities and berries to apply big data. For example, the lack of expert human resources and reliable infrastructures to collect data are some barriers. Further, there is a huge gap in structure and governance related to agricultural big data. Big data can be used to recommend guidance to farmers based on their crop conditions, weather and responsiveness to fertilizers. Moreover, by integrating agricultural data with the supply chain framework farmers can gain more from planned harvesting.

Nowadays, big data play a crucial role in biomedical and healthcare. By combining data from many sources, practitioners can present a new perspective on patient care that takes a patient's entire health status from DNA to behaviors. Also, the accessibility of cutting-edge mobile health devices enables more accurate real-time data collection and saves many lives by early detection of the seriousness of diseases.

Today businesses use big data for multiple purposes. Researchers conclude that having a positive review on social media for a product effect on the market of the product. Further big data can use to identify market trends, customer satisfaction, develop personal recommendation systems and etc.

Big data applications impact the field of transportation in numerous ways. Understanding passenger behaviors helps in decision making. Further, statistical methods such as Multiple linear regression, factorial analysis of variance (ANOVA) and etc. are used to analyze the tourism data. In all references include various fields and the benefits of big data technology for these studies. Finally, to obtain the best possible outcome from our study, authors may propose big data technology applications uses various fields of applications.

Only 45 articles, 2% of the Google Scholar results for the phrases "Big data Applications" and "Big data Technology Analytics" were examined. Another limitation of this assessment is that articles written outside of English were not taken into consideration. Because only Google Scholar was searched, academic databases and journal articles were also excluded from the search. Data collection from online websites only has possible drawbacks. It was unable to adopt data from a variety of sources, including first-hand information obtained from interviews and questionnaires (Ying, Chan and Qi, 2020).

There are still several issues with deploying big data approaches in real-world applications,

including a lack of pertinent supporting policies and an absence of uniformity of standards and norms (Li *et al*., 2019).

V. CONCLUSION

Big data analysis was reviewed in this article. This paper examines the applications of big data in different fields and the most current uses of big data technologies in various areas. Big data applications are used in agricultural, medical, transport, education, bureaucracy, tourism, security and business domains. Further to analyses the big data nowadays, mainly machine learning techniques are used. Additionally, some algorithms perform far better in some specific fields. For example, decision trees, logistic regression, naive Bayesian approaches, Bayesian networks, neural networks, etc., perform better in medical applications. Big data applications in all industries have a bright future since they develop new data values. Pertinent research and technology, the collaboration between research institutions and businesses, and strong government encouragement help further bring considerable data value.

REFERENCES

Alyass, A., Turcotte, M. and Meyre, D. (2015) 'From big data analysis to personalized medicine for all: Challenges and opportunities', BMC Medical Genomics, 8(1), pp. 1–12. doi: 10.1186/s12920-015-0108-y.

Arena, F. and Pau, G. (2020) 'An overview of big data analysis', Bulletin of Electrical Engineering and Informatics, 9(4), pp. 1646–1653. doi: 10.11591/eei.v9i4.2359.

Austin, C. and Kusumoto, F. (2016) 'The application of Big Data in medicine: current implications and future directions', Journal of Interventional Cardiac Electrophysiology, 47(1), pp. 51–59. doi: 10.1007/s10840-016-0104-y.

Bajpai, S. and Mani, S. (2017) 'Big Data in Education and Learning Analytics', TechnoLearn: An International Journal of Educational Technology, 7(1and2), p. 45. doi: 10.5958/2249-5223.2017.00005.5.

Batarseh, F. A. and Latif, E. A. (2016) 'Assessing the Quality of Service Using Big Data Analytics: With Application to Healthcare', Big Data Research, 4(October), pp. 13–24. doi: 10.1016/j.bdr.2015.10.001.

Bofill-De Ros, X. et al. (2019) 'QuagmiR: A cloud-based application for isomiR big data analytics', Bioinformatics, 35(9), pp. 1576–1578. doi: 10.1093/bioinformatics/bty843.

Du, G., Liu, Z. and Lu, H. (2021) 'Application of innovative risk early warning mode under big data technology in Internet credit financial risk assessment', Journal of Computational and Applied Mathematics, 386, p. 113260. doi: 10.1016/j.cam.2020.113260.

Gorodov, E. Y. E. and Gubarev, V. V. E. (2013) 'Analytical review of data visualization methods in application to big data', Journal of Electrical and Computer Engineering, 2013. doi: 10.1155/2013/969458.

Govindan, K. et al. (2018) 'Big data analytics and application for logistics and supply chain management', Transportation Research Part E: Logistics and Transportation Review, 114(March), pp. 343–349. doi: 10.1016/j.tre.2018.03.011.

Guo, T. and Wang, Y. (2019) 'Big data application issues in the agricultural modernization of china', Ekoloji, 28(107), pp. 3677–3688.

Hernández-Leal, E. J., Duque-Méndez, N. D. and Moreno-Cadavid, J. (2017) 'Big Data: una exploración de investigaciones, tecnologías y casos de aplicación', TecnoLógicas, 20(39), pp. 15–38. doi: 10.22430/22565337.685.

Horita, F. E. A. et al. (2017) 'Bridging the gap between decision-making and emerging big data sources: An application of a model-based framework to disaster management in Brazil', Decision Support Systems, 97, pp. 12–22. doi: 10.1016/j.dss.2017.03.001.

Huda, M. et al. (2016) 'Innovative teaching in higher education: The big data approach', Turkish Online Journal of Educational Technology, 2016(NovemberSpecialIssue), pp. 1210–1216.

Iqbal, R. et al. (2020) 'Big data analytics: Computational intelligence techniques and application areas', Technological Forecasting and Social Change, 153(December 2017), pp. 0–1. doi: 10.1016/j.techfore.2018.03.024.

Ishika and Mittal, N. (2021) 'Big Data Analysis for Data VisualizationA Review', 2021 9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions), ICRITO 2021, pp. 64–75. doi: 10.1109/ICRITO51393.2021.9596423.

Jordan, S. R. (2014) 'Beneficence and the Expert Bureaucracy: Ethics for the Future of Big Data Governance', Public Integrity, 16(4), pp. 375–394. doi: 10.2753/PIN1099-9922160404.

Kamilaris, A., Kartakoullis, A. and Prenafeta-Boldú, F. X. (2017) 'A review on the practice of big data analysis in agriculture', Computers and Electronics in Agriculture, 143(January), pp. 23–37. doi: 10.1016/j.compag.2017.09.037.

Kim, Y. H. and Ahn, J. H. (2016) 'A Study on the Application of Big Data to the Korean College Education System', Procedia Computer Science, 91(Itqm 2016), pp. 855–861. doi: 10.1016/j.procs.2016.07.096.

Kolisetty, V. V. and Rajput, D. S. (2020) 'A review on the significance of machine learning for data analysis in big data', Jordanian Journal of Computers and Information Technology, 6(1), pp. 41–57. doi: 10.5455/jjcit.71-1564729835.

Koseleva, N. and Ropaite, G. (2017) 'Big Data in Building Energy Efficiency: Understanding of Big Data and Main Challenges', Procedia Engineering, 172, pp. 544–549. doi: 10.1016/j.proeng.2017.02.064.

Lee, C. H. and Yoon, H. J. (2017) 'Medical big data: Promise and challenges', Kidney Research and Clinical Practice, 36(1), pp. 3–11. doi: 10.23876/j.krcp.2017.36.1.3.

Li, S. et al. (2019) 'Research on the application of information technology of Big Data in Chinese digital library', Library Management, 40(8–9), pp. 518–531. doi: 10.1108/LM-04-2019-0021.

Li, Y. et al. (2019) 'Deep learning in bioinformatics: Introduction, application, and perspective in the big data era', Methods, 166(2019), pp. 4–21. doi: 10.1016/j.ymeth.2019.04.008.

Luo, J. et al. (2016) 'Big Data Application in Biomedical Research and Health Care: A Literature Review', Biomedical Informatics Insights, 8, p. BII.S31559. doi: 10.4137/bii.s31559.

Maqsood Ahmad Sandhu, Ahm Shamsuzzoha, P. H. (2018) '기사 (Article) 와 안내문 (Information) [', The Eletronic Library, 34(1), pp. 1–5.

Muliawaty, L. et al. (2019) 'The concept of big data in bureaucratic service using sentiment analysis', International Journal of Sociotechnology and Knowledge Development, 11(3), pp. 1–13. doi: 10.4018/IJSKD.2019070101.

Müller, S. D. and Jensen, P. (2017) 'Big data in the Danish industry: application and value creation', Business Process Management Journal, 23(3), pp. 645–670. doi: 10.1108/BPMJ-01-2016-0017.

Ouyang, Q., Wu, C. and Huang, L. (2018) 'Methodologies, principles and prospects of applying big data in safety science research', Safety Science, 101(March 2017), pp. 60–71. doi: 10.1016/j.ssci.2017.08.012.

Padma, P. and Ahn, J. (2020) 'Guest satisfaction & dissatisfaction in luxury hotels: An application of big data', International Journal of Hospitality Management, 84(February 2019). doi: 10.1016/j.ijhm.2019.102318.

Peng, L. et al. (2017) 'The Advances and Challenges of Deep Learning Application in Biological Big Data Processing', Current Bioinformatics, 13(4), pp. 352–359. doi: 10.2174/1574893612666170707095707.

Singh Jain, A. D. et al. (2017) 'Application of Big Data in Supply Chain Management', Materials Today: Proceedings, 4(2), pp. 1106–1115. doi: 10.1016/j.matpr.2017.01.126.

Soomro, K. et al. (2019) 'Smart city big data analytics: An advanced review', Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 9(5), pp. 1–25. doi: 10.1002/widm.1319.

Stieb, D. M., Boot, C. R. and Turner, M. C. (2017) 'Promise and pitfalls in the application of big data to occupational and environmental health David Stieb, Cecile Boot, Michelle Turner, Osmar Zaiane', BMC Public Health, 17(1), pp. 4–7. doi: 10.1186/s12889-017-4286-8.

Tetko, I. V. et al. (2016) 'BIGCHEM: Challenges and Opportunities for Big Data Analysis in Chemistry', Molecular Informatics, 35(11–12), pp. 615–621. doi: 10.1002/minf.201600073.

Tulasi, B. (2013) 'Significance of Big Data and Analytics in Higher Education', International Journal of Computer Applications, 68(14), pp. 21–23. doi: 10.5120/11648-7142.