Available online www.jsaer.com

***Journal of Scientific and Engineering Research*, 2021, 8(8):144-152**

Review Article

ISSN: 2394-2630
CODEN(USA): JSERBR

Big Data in Industry: An Overview

Matthew N. O. Sadiku[1], Uwakwe C. Chukwu[2], Abayomi Ajayi-Majebi[3], Sarhan M. Musa[1]

[1]Roy G. Perry College of Engineering, Prairie View A&M University, Prairie View, TX, USA
[2]Department of Engineering Technology, South Carolina State University, Orangeburg, SC, USA
[3]Department of Manufacturing Engineering, Central State University, P.O. Box 1004, Wilberforce, OH, USA
Email: sadiku@ieee.org; uchukwu@scsu.edu; ajayi-majebi@centralstate.edu; smmusa@pvamu.edu

Abstract With the rise of digital technology, big data has remarkably revolutionized the industry standards. Big data refers to large, hard-to-manage volumes of data, both structured and unstructured. The marketing and sales of any industry largely depend on how data is used. Industries that are improving their services through big data include healthcare, manufacturing, agriculture, telecommunications, retail, fitness, travel, finance, e-commerce, entertainment, banking, oil and gas, sports, and social media. This paper is an overview on the application of big data in industries.

Keywords big data, data analytics, industry, big data in industry, industrial big data

Introduction

Digital transformation has placed data at the center of all industries, of all sizes. The rapid development of Internet of things and cloud computing have led to the explosive growth of data in almost every industry [1]. Increasing demand for natural resources such as minerals, oil, gas, metals, has led to an increase in the volume, complexity, and velocity of data. The volume of data generated in an industrial system is far beyond the capability of traditional methods. Such massive amounts of information are called big data BD and they that can work wonders. In layman's terms, BD is basically the science of collecting massive amounts of data and using analysis tools to analyze, find patterns and trends to gain actionable insights. Big data consists of huge complex sets of unstructured, semi-structured, or structured data obtained from multiple sources. As illustrated in Figure 1, big data is everywhere [2].

Figure 1: Big data is everywhere [2]

Big data has become a big game-changer and an emerging trend in most modern industries. It has changed just about every industry, from manufacturing to banking. Industries need an effective system for gathering, storing, and analyzing data from several heterogeneous sources to extract and act on valuable insights. One can take data from any source and analyze it to find answers that enable: (1) cost reductions, (2) time reductions, (3) new product development and optimized offerings, and (4) smart decision making [3]. The big data market includes infrastructure providers, data centers, data-as-a-service providers, and other vendors. Big data infrastructure will require strategic governance and optimized security. Figure 2 shows some examples of big data usage [4].

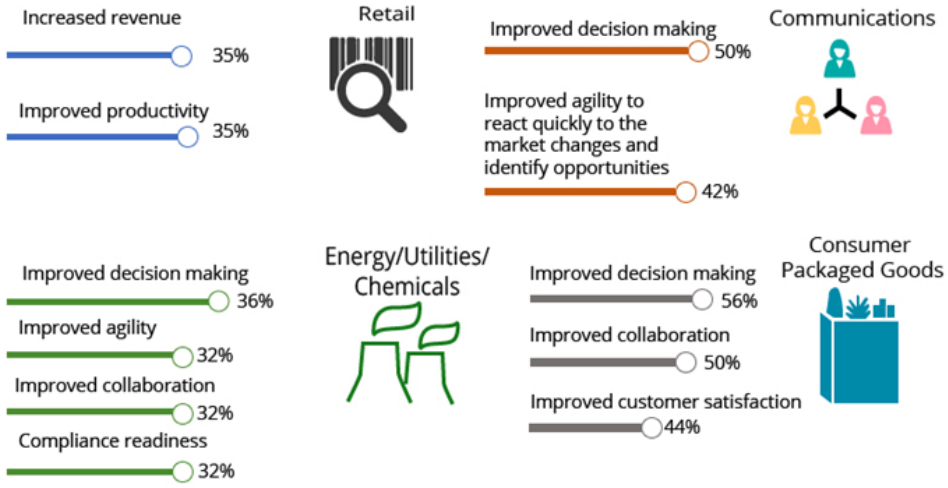

Figure 2: Examples of big data usage [4]

Review on Big Data

Big data (BD) refers to a collection of data that cannot be captured, managed, and processed by conventional software tools. It is a relatively new technology that can help many industries. The three main sources of big data are machines, people, and companies. Big data can be described with 42 Vs [5]. The first five Vs are volume, velocity, variety, veracity, and value [6].

- *Volume*: This refers to the size of the data being generated both inside and outside organizations and is increasing annually. Some regard big data as data over one petabyte in volume.
- *Velocity*: This depicts the unprecedented speed at which data are generated by Internet users, mobile users, social media, etc. Data are generated and processed in a fast way to extract useful, relevant information. Big data could be analyzed in real time, and it has movement and velocity.
- *Variety*: This refers to the data types since big data may originate from heterogeneous sources and is in different formats (e.g., videos, images, audio, text, logs). BD comprises of structured, semi-structured or unstructured data.
- *Veracity*: By this, we mean the truthfulness of data, i.e. weather the data comes from a reputable, trustworthy, authentic, and accountable source. It suggests the inconsistency in the quality of different sources of big data. The data may not be 100% correct.
- *Value*: This is the most important aspect of the big data. It is the desired outcome of big data processing. It refers to the process of discovering hidden values from large datasets. It denotes the value derived from the analysis of the existing data. If one cannot extract some business value from the data, there is no use managing and storing it.

On this basis, small data can be regarded as having low volume, low velocity, low variety, low veracity, and low value. Additional five Vs has been added [7]:

- *Validity:* This refers to the accuracy and correctness of data. It also indicates how up to date it is.
- *Viability:* This identifies the relevancy of data for each use case. Relevancy of data is required to maintain the desired and accurate outcome through analytical and predictive measures.

- *Volatility:* Since data are generated and change at a rapid rate, volatility determines how quickly data change.
- *Vulnerability:* The vulnerability of data is essential because privacy and security are of utmost importance for personal data.
- *Visualization:* Data needs to be presented unambiguously and attractively to the user. Proper visualization of large and complex clinical reports helps in finding valuable insights.

Figure 3 shows the 10V's of big data. In addition, the 10V's above, some suggest the following 5V's: Venue, Variability, Vocabulary, Vagueness, and Validity) [8]. The future of big data will bring more Vs.

To thrive in today's complex business environment, businesses must adopt a data-driven culture and leverage analytics platforms to make key decisions that improve productivity. Industries that benefit from big data include the healthcare, financial, airline, travel, restaurants, automobile, sports, agriculture, manufacturing, and hospitality industries.

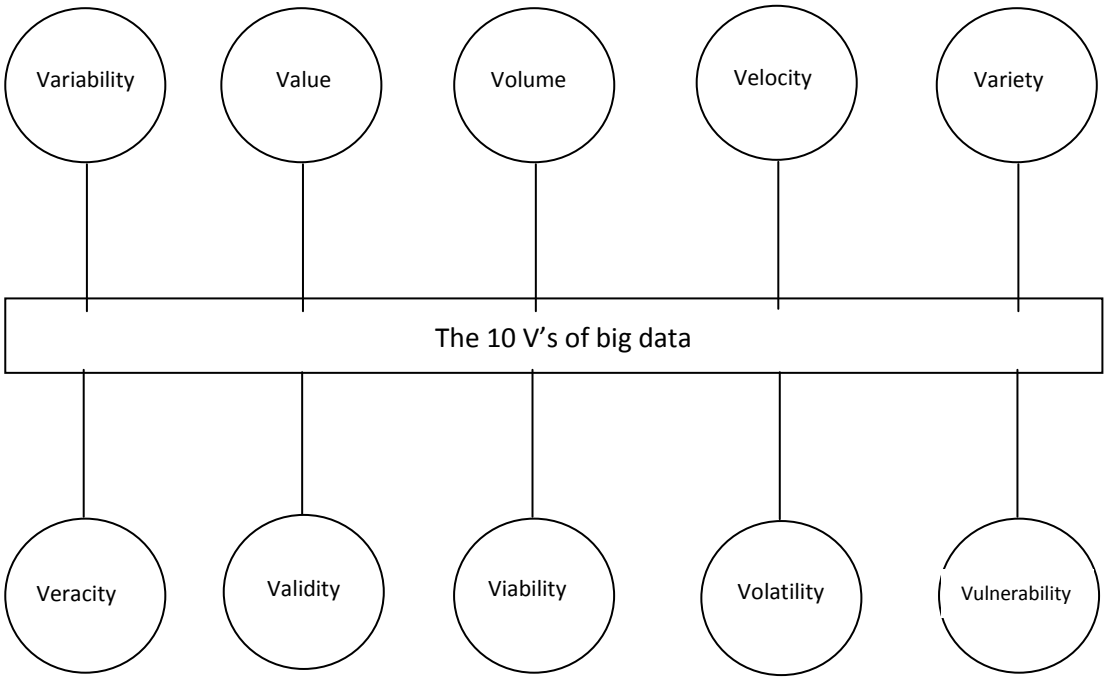

Figure 3: The 10V's of big data

Big Data Analytics

Big data sets can be staggering in size so that its analysis is daunting. Every day, data is growing bigger and bigger, and big data analysis (BDA) has become a requirement for gaining invaluable insights into data such that companies could gain significant profits in the global market. Big data analytics can leverage the gap within structured and unstructured data sources. Once the big data is ready for analysis, we use advanced software programs such as Hadoop, MapReduce, MongoDB, Spark, Cassandra, Apache Storm, and NoSQL databases [9]. Big data analytics refers to how we can extract, validate, translate, and utilize big data as a new currency of information transactions. It is an emerging field that is aimed at creating empirical predictions. Data-driven organizations use analytics to guide decisions at all levels [10].

Big data analytics is capable of processing massive amounts of dirty data and extract the gold information from it. It is crucial to generating more revenue and providing personalized experiences in the digitally-driven industry. How industries use data analytics in their business varies considerably by sector, business size, and access to resources. The use of data analytics delivers significant gains in the following ways [11]:

- *Cost Reduction:* Using big data technologies like Hadoop or cloud-based analytics allows organizations to store large amounts of data in a cost-effective, efficient manner.
- *Better Decision-Making:* Big data's primary value comes from its ability to facilitate smarter decision-making.

- *Improved Products & Services:* Big data analytics applications also allow companies to come up with new and improved solutions and products.

Industrial Big Data

Industrial big data refers to the vast amount of generated data at a high speed by industrial equipment. Its objective is to reveal insights from the massive amount of raw data and turn that information into value. As data from automated industrial equipment and sensors are being generated at high speed and volume, the infrastructure of processing, storing, and managing the large data becomes a major challenge faced by any industry [12].

Enormous amounts of industrial data are been generated by manufacturing companies. Industrial big data analytics helps us understand the nature of machine states and anomaly and predict the future of manufacturing. Compared to big data in general, industrial big data can create value in different sections of manufacturing industry [13].

Applications

Some application areas of big data are shown in Figure 4 [14]. Here is the list of the top industries using big data applications [15]:

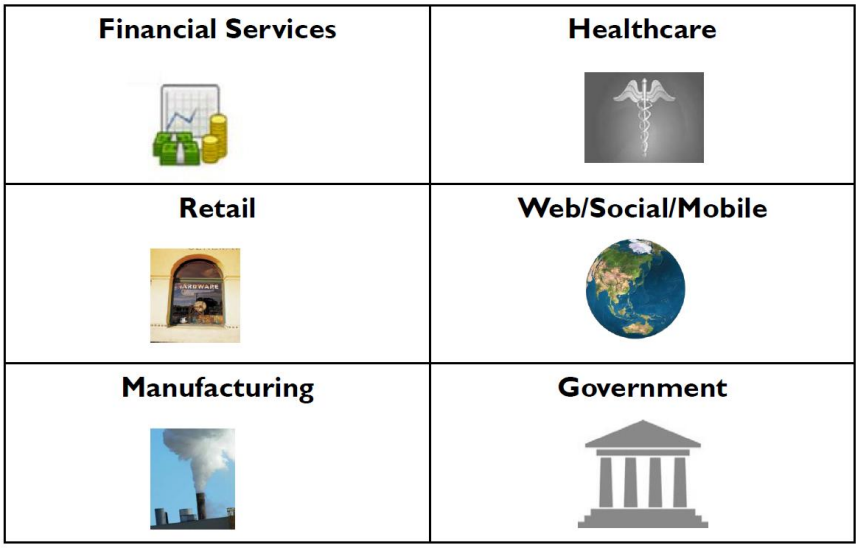

Figure 4: Some application areas of big data [14]

1. Banking and Securities
2. Communications, Media and Entertainment
3. Healthcare
4. Education
5. Manufacturing
6. Government
7. Insurance
8. Retail and Wholesale trade
9. Transportation and Logistics
10. Energy and Utilities
11. Agriculture
12. Oil and gas

We will examine how some of these industries are using big data.

- *Healthcare:* The healthcare sector generates huge data sets. It is one of the largest recent adopters of big data analytics. Data is crucial for doctors seeking to figure out which blood pressure range is

normal or how much sugar you should consume each day. Data collection using devices like smartphones and wearables will help doctors understand their patients more and deliver better care. The healthcare industry now employs data analytics to answer bigger, more complex questions, improve patient outcomes, and provide better experiences [16]. Big data can identify disease trends as a whole based on demographics, geographics, socio-economics, and other factors. IBM Watson enforces and integrates a wide array of healthcare domains to provide meaningful and structured data. IBM Watson has been used to predict specific types of cancer based on the gene.

- *Manufacturing:* Manufacturers have been generating a lot of real-time data for quite some time. Manufacturers of all types of products are finding significant value in big data. In the data-driven economy, turning data into actionable analytics is the best way to boost efficiency, quality, and productivity. Modern manufacturing facilities are data-rich environments that produce manufacturing intelligence. Innovative manufacturers are using big data technologies to focus on customers and operations at the same time. Big data is shaking up the manufacturing industry across the board. Billions of data samples are being produced by every single machine per day in a manufacturing line. The amount of data generated by IoT and today's manufacturing systems must be translated into actionable ideas. Big data enables companies to solve today's manufacturing challenges and to gain a competitive edge. It is helping to integrate isolated systems so that companies can obtain a complete visualization of the manufacturing processes. It is driving automation and addressing the challenges of modern factories which are becoming increasingly complex and interconnected [17].
- *Financial Industry*: The financial sector generates a lot of data. Big data has fundamentally changed the finance industry, particularly stock trading. Big data analytics in finance allows companies to create convenient, personalized products and services, without compromising consumer security. Big data analytics can evaluate stock prices alongside social trends, economic factors, and the political landscape that might impact the stock market. Today, banks and credit card companies can use data analytics to detect anomalies, instantly freeze consumer accounts, detect and prevent credit card fraud, making it safer for consumers to make purchases online [11].
- *Industry 4.0:* This is essentially the fusion of the real world with the virtual world. An important aim of Industry 4.0 is the realization of intermittent manufacturing at mass production's productivity and specific cost. Big data is basically at the core of Industry 4.0. Major technologies that adopt Industry 4.0 successfully include advanced technologies such as big data, artificial intelligence, robotics, Internet of things, cloud computing, cyber-physical system, and 3D printing. Although these technologies differ from each other, it is difficult to talk about one without the other. Industry 4.0 has been adopted across a wide range of industry sectors including aerospace, defense and security, automotive, electronics, and industrial manufacturing. However, without big data, there would have been no Industry 4.0 and the intelligent technologies that support it. Industry 4.0 is paving the way for widespread big data analytics. Industry 4.0 big data analytics can uncover patterns that predict machine failures before they occur. Big data will provide further advancement in Industry 4.0 and play an efficient role in its successful adoption. Manufacturers need solutions from providers who are part of the Industry 4.0 revolution [18,19]. Industry 4.0 is expected to change production in the future.
- *Agriculture*: Today, data is transforming one of the world's oldest industry: agriculture. Data were always regarded as an important source of knowledge for farmers, agricultural professionals, and policy-makers. Agriculture has become increasingly high-tech and data-driven over the years. Data analytics are now crucial for agriculture. They help farmers to grow out of the land so that they feed a growing world population. US agricultural manufacturer John Deere has adopted big data practices. Big data is paving its way to revolutionize the food industry. To stay competitive and not left behind, food and beverage companies should consider implementing data analytics tools. Various food retailers such as McDonald's, TacoBell, Pizza Hut, Dominos, and KFC have used data analytics [20,21]. Figure 5 shows some data analysts in agricultural industry [16].

Figure 5: Data analysts in agricultural industry [16]

Other areas of applications include sport industry, social media, oil and gas industry, chemical industry, automotive industry, hospitality industry, hotel industry, mining industry, construction industry, telecommunications, and insurance.

BENEFITS

The benefits of big data cut across every industry. Big data is transformational for businesses of any size. Most industries adopting big data projects do so to enhance customer experience, cost reduction, better-targeted marketing, and making existing processes more efficient. With the right tools and strategies, big data can provide a competitive edge. Big data will soon become a key basis of competition. It assists companies in improving their marketing campaigns, increasing productivity and innovation, and developing creative high demand products. In 2014, BMW used big data to detect vulnerabilities in their new car prototypes. General Electric helps power producers use big data at four levels [22].

Other benefits of big data in industry include the following [4]:

- *Cost reduction:* Improved process efficiency and business intelligence
- *Decision-making:* Evaluates past data to make future predictions
- *Fraud detection:* Supports the finance-based detection and prevents fraud activities
- *Managing risk:* Proactively identifies and manages the data exposure to the risks associated with it
- *Improved organizational performance:* Enables you to know your customers and allows to serve them well
- *Better sales*: New competitive strategies are created and real-time analytics keeps you a step ahead
- *Customer relationship management:* Provides in-depth information about your customer trends

Challenges

Although big data is gaining interest in several industries, there are still some major challenges which should to be addressed. In a big data world, policies pertaining to privacy, security, intellectual property, and liability need to be addressed. Other challenges include the following [23,24].

- *Cost*: A major challenge associated with the application of big data' in any industry is the cost of managing, recording, storage, and analysis of the data.

- *Lack of awareness*: In most industries, there is lack of awareness of big data analytics within the industry and business support.
- *Lack of skilled personnel:* It is expedient to hire new skilled tech employees or retrain old ones.
- *Resisting the new technologies:* Some employees will naturally resist a new technology like big data.
- *Ethics:* As big data combines information from diverse sources to create knowledge, it faces some ethical issues arising from reselling consumers' data to the secondary market. Figure 6 depicts some current ethical issues within the big data industry [25].

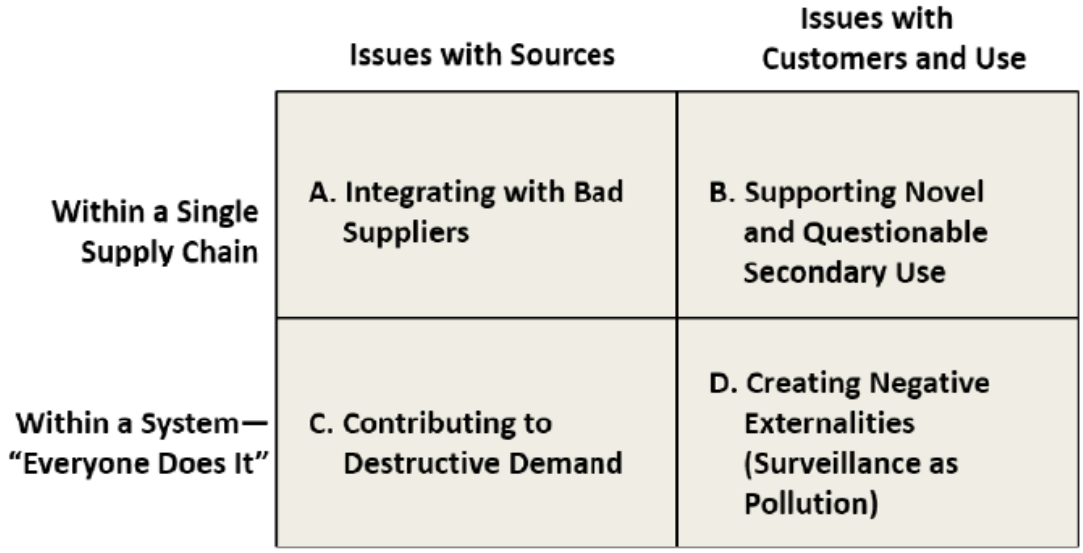

Figure 6: Some current ethical issues within the big data industry [25]

Conclusion

The ability to process massive amounts of data and extract useful insights from it has greatly revolutionized our modern society. This phenomenon—known as Big Data—has applications across a wide range of industries [26]. Big data will definitely continue to grow to petabytes, zettabytes, and beyond. With the rapid spread of Internet of things and other sensors, the volume and velocity of data are only going to grow. Data are now becoming an important factor of production, alongside labor and capital.

Big data is now receiving a lot of press and attention. Companies that are still wondering whether big data analytics fit into their business strategy are losing ground fast. Embracing data-based operations is the key to innovating and staying relevant for any company in the age of big data. More information about big data in industry can be found in the books in [27,28] and the following related journals:

- *Journal of Big Data*
- *Big Data Research.*
- *Big Data & Society.*

References

[1]. X. Jin et al., "Significance and challenges of big data research," *Big Data Research*, vol. 2, no. 2, June 2015, pp. 59-64.

[2]. L. Andrews, "What are examples of big data in manufacturing?" October 2019, https://www.sensrtrx.com/what-are-examples-of-big-data-in-manufacturing/

[3]. "Big data: What it is and why it matters," https://www.sas.com/en_us/insights/big-data/what-is-big-data.html

[4]. V. Agrawal, "Big data in a nutshell," June 2019, https://jaxenter.com/big-data-nutshell-159112.html

[5]. "The 42 V's of big data and data science," https://www.kdnuggets.com/2017/04/42-vs-big-data-data-science.html

[6]. M. N. O. Sadiku, M. Tembely, and S. M. Musa, "Big data: An introduction for engineers," *Journal of Scientific and Engineering Research*, vol. 3, no. 2, 2016, pp. 106-108.
[7]. P. K. D. Pramanik, S. Pal, and M. Mukhopadhyay, "Healthcare big data: A comprehensive overview," in N. Bouchemal (ed.), *Intelligent Systems for Healthcare Management and Delivery*. IGI Global, chapter 4, 2019, pp. 72-100.
[8]. J. Moorthy et al., "Big data: Prospects and challenges," *The Journal for Decision Makers*, vol. 40, no. 1, 2015, pp. 74–96.
https://www.grandviewresearch.com/industry-analysis/industrial-wireless-sensor-networks-iwsn-market
[9]. M. N. O. Sadiku, J. Foreman, and S. M. Musa, "Big data analytics: A primer," *International Journal of Technologies and Management Research*, vol. 5, no. 9, September 2018, pp. 44-49.
[10]. C. M. M. Kotteti, M. N. O. Sadiku, and S. M. Musa, "Big data analytics," *Invention Journal of Research Technology in Engineering & Management*, vol. 2, no. 10, Oct. 2018, pp. 2455-3689.
[11]. "How big data is transforming industries in big ways," June 2020,
https://www.tiempodev.com/blog/how-big-data-is-transforming-industries-in-big-ways/
[12]. "Industrial big data," *Wikipedia*, the free encyclopedia
https://en.wikipedia.org/wiki/Industrial_big_data
[13]. J. Wang et al., "Industrial big data analytics: challenges, methodologies, and applications," submitted to IEEE Transactions on Automation Science and Engineering.
[14]. R. Oeglar, "Introduction to analytics and big data – Hadoop,"
https://na.eventscloud.com/file_uploads/1e91ec22ad6f4f7c0796d0adabf4be5a_RobPeglar_Introduction_to_Analytics_Big_Data_Hadoop.pdf
[15]. "Top 10 big data applications across industries,"
https://www.simplilearn.com/tutorials/big-data-tutorial/big-data-applications
[16]. C. Tozzi, "Big data industries: 5 industries being reshaped by data analytics," February 2021,
https://www.precisely.com/blog/big-data/big-data-industries-data-analytics
[17]. M. N. O. Sadiku, J. T. Ashaolu, A. Ajayi-Majebi, S. M. Musa, "Big data in manufacturing," *International Journal of Scientific Advances*, vol. 2, no. 1, Jan.-Feb. 2021, pp. 63-66.
[18]. M. Schuldenfrei, "Big data challenges of Industry 4.0," April 2019,
https://www.datanami.com/2019/04/25/big-data-challenges-of-industry-4-0/
[19]. M. Javaid et al., "Significant applications of big data in Industry 4.0," *Journal of Industrial Integration and Management*, March 2021.
[20]. M. Rangaiah, "7 benefits of big data in food industry," April 2020,
https://www.analyticssteps.com/blogs/role-big-data-food-industry
[21]. M. N. O. Sadiku, S. Cui, Y. Wang, S. M. Musa, "Big data in agriculture," *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 9, No. 4, pp. 70-72, April 2019.
[22]. A. Bekker, "Big data in manufacturing: Use Cases + Guide on How To Start,"
https://www.scnsoft.com/blog/big-data-in-manufacturing-use-cases
[23]. K. E.M. Martin, "Ethical issues in the big data industry," *MIS Quarterly Executive*, vol.14, no. 2, April 2015.
[24]. M. Mohammadpoor and F. Torabi, "Big Data analytics in oil and gas industry: An emerging trend," *Petroleum*, vol. 6, no. 4, December 2020, pp. 321-328.
[25]. K. E. Martin, "Ethical issues in the big data industry," MIS Quarterly Executive, vol. 14, no. 2, June 2015, pp. 67-85.
[26]. M. Bilal et al., "Big Data in the construction industry: A review of present status, opportunities, and future trends,"
https://uwe-repository.worktribe.com/output/909586/big-data-in-the-construction-industry-a-review-of-present-status-opportunities-and-future-trends

[27]. C. Tang, *The Data Industry: The Business and Economics of Information and Big Data.* John Wiley & Sons, 2016.
[28]. S. C. Lewis (ed.), *Journalism in an Era of Big Data: Cases, Concepts, and Critiques.* Routledge, 2018.

About Authors

Matthew N. O. Sadiku is a professor emeritus in the Department of Electrical and Computer Engineering at Prairie View A&M University, Prairie View, Texas. He is the author of several books and papers. His areas of research interest include computational electromagnetics and computer networks. He is a fellow of IEEE.

Uwakwe C. Chukwu is an associate professor in the Department of Industrial & Electrical Engineering Technology of South Carolina State University. He has published several books and papers. His research interests are power systems, smart grid, V2G, energy scavenging, renewable energies, and microgrids.

Abayomi Ajayi-Majebi is a professor in the Department of Manufacturing Engineering at Central State University in Wilberforce, Ohio. In 2015 he was honored by the White House as a Champion of Change for his significant contributions to the engineering education of minority students. He is a senior member of both the Society of Manufacturing Engineers and the American Society for Quality.

Sarhan M. Musa is a professor in the Department Electrical and Computer Engineering at Prairie View A&M University, Texas. He has been the director of Prairie View Networking Academy, Texas, since 2004. He is an LTD Sprint and Boeing Welliver Fellow. His areas of research interest include computational electromagnetics and computer networks.