



Big Data Characteristics (V's) in Industry

Nagham Saeed*, Laden Husamaldin

School of Computing and Engineering, University of West London, London – UK

Article information

Article history:

Received: May, 24, 2021

Accepted: June, 01, 2021

Available online: June, 14, 2021

Keywords:

Big data,

Big data characteristic,

V's,

Big data analytics characteristics

*Corresponding Author:

Nagham Saeed

nagham.saeed@uwl.ac.uk

Abstract

In the new digital age, Data is the collection of the observation and facts in terms of events, thus data is continuously growing, getting denser and more varied by the minute across multiple channels. Nowadays, consumers generate mass amounts of data on a daily basis. Hence, Big Data (BD) emerged and is evolving rapidly, the various types of data being processed are huge, and ensuring that this data is being used efficiently is becoming increasingly more difficult. BD has been differentiated into several characteristics (the V's) and many researchers have been developing more characteristics for new purposes over the past years. Therefore, it is shown from observation that there is a clear gap between researchers about the current status of the BD characteristics. Even after the introduction of newer characteristics, many papers are still proposing the use of 3 or 5 V's, while some researchers are far more progressed and has reached up to 10V's. This paper will provide an overview of the main characteristics that have been added over time and investigate the recent growth of Big Data Analytics (BDA) characteristics in each industry sector which will provide some detailed and general scope for most researchers to consider and learn from.

DOI: xxxxxxxxxxxxxxxxxxxxxxxxx ©2021, Corporation of Research and Industrial Development/ Ministry of Industry and Minerals – Iraq

1. Introduction

Big data [1] is at the core of modern science, business and big data analysis, which has been created from various channels, such as online transactions, e-mails, videos, audios, images, streams of clicks, logs, posts, search queries, health records, interactions with social networking, science data, sensors and mobile phones and their applications. These data sets sizes are extremely huge, and it is amazing how many data these channels store and use every minute in real time. Big data are big sized data sets that are used to analyse and visualize core information, which will help organizations and their departments make big business decisions. For example, larger new data sources are being generated from mobile devices and large companies like Google, Apple, Facebook, Yahoo, Twitter who are beginning to look carefully at this data to find useful patterns to improve user experience. It is undeniable that millions and millions of data sets are now being generated from technology devices and, as a growing field, so many digital apps such as Siri, Alexa, Cortana are also contributing to this phase of data collection. Every time we ask these devices to perform a task for us (e.g. "Check weather", "Call mum", "Nearest store", etc.), the input we

provide to such data is processed in the cloud and used for different reasons. It illustrates how much the method of producing data has changed and how customers today are generating data for companies without even realising. All these generated data sets are very important, and they are then used for a number of business reasons, such as marketing, stock level, level of employment, and help the business to make better decisions. It's clear that the data collected is proportional to the growth of big businesses and many companies rely on the big data analytics to increase their profits [2]. As Ghasemaghaei [3] stated, by processing large volumes of different data types in a timely manner, firms would be presented with more opportunities to make evidence-based decisions that allow firms to compete in the marketplace. As mentioned, the Big Data is so huge that data analysts and consultants like to differentiate this data into different characteristics. Many researchers are continuously generating new characteristics and developments in these categories and have discussed the Vs of Big data.

With those many Vs of BD were identified and mentioned in literature, how many were considered, referred to and widely used by industry sectors and recent research articles. Therefore, two questions are asked in this paper:

1. Identify the main BD characteristics to date.
2. What are the recent BD characteristics that each industry sector is referring to?

The answers, based on the literature are in the coming sections.

2. Big Data Characteristics: Definition and Growth

The main purpose of big data analytics is to provide accurate, reliable, and understandable information to support decision making. Thus, the datasets should have some basic characteristics before starting the analytics. This section covers the definition and growth of the BD characteristics from three V's to ten V's.

Three Vs

Several researchers have addressed the three main Vs of big data characteristics and how significant they are in the big data analytics world. Initially, the differentiation involved three main characteristics that include Volume, Velocity and Variety [4].

Volume refers to the large amount of the data (size) which is growing continuously for each sector, it is immense and provides better prediction for the future;

Velocity refers to how quickly the data can be analysed in order to make decisions (access), data is increasingly coming in big waves and it is important to obtain useful analytic information in real time.

Finally, Variety refers to the extremely heterogeneous data sources both at the level of the schema (structured, unstructured and semi-structured) including text, sensor, audio, video, graph, and more. Hence, there is a lot of variety between data collected and this provides a better analysis of big data.

In other words, when considering a data set, the provision of reliable information should be large enough, easy to access and in different structures. Which means large varied of data needs to be easily accessible to make real-time decisions from the data sets.

The authors mentioned above are not the only ones, there are many other academics and researchers who investigated the three Vs of big data. The views are reasonably similar, but the characteristics were applied in different sectors. For example, Dong and Srivastava's [5] study is one of the examples that was done on BD Vs to provide a clear understanding, the authors analysed and explained each characteristic from their perspective.

Four Vs

Big Data analytic characteristics have grown over the years due to the rapid technological change such as the Internet of things, machine learning, artificial intelligence, robotics, 3D printing, biotechnology, nanotechnology, renewable energy technologies, satellite and drone technologies. Therefore, scholars have addressed this issue by revisiting the three BD characteristics and a newly added fourth characteristic (Veracity).

One of the fundamental difficulties the fourth characteristic was introduced for, stated by Berti-Equille and Ba [6] is that data can be biased, noisy, obsolete, erroneous, misleading and therefore unreliable. The problem is exacerbated by conflicting data from various sources which in turn leads to the veracity of the data having to be evaluated.

Hence, Veracity refers to the various values, with significant differences in data coverage, accuracy and timeliness. It is important to ensure that the data set dealt with is sufficiently accurate since very important business decisions are made on the basis of those data sets. Kepner et al. [7] addressed the confidentiality, credibility and availability of their data which resulted in the introduction of veracity.

Five Vs

While many researchers believed in three and four Vs for BD characteristics, Nguyen [8] discussed the data characteristics in his paper and reviewed the five big Vs. According to Nguyen, there are five Vs for Big data; he adds Value to the existing four Vs (Volume, Velocity, Variety and Veracity).

Value is described as the individual or organizational capability of turning big data into real values, which includes an ability to collect and then leverage the data to achieve specific goals. Furthermore, Ghasemaghahi has raised the significance of the fifth V of BD in her [3] paper Where, Value refers to the social and economic value (mostly monetary) that big data may generate. The end goal of data analytics is to provide information and facts that will help the big organizations make better decisions that add value to their companies, and, in most cases, the value is monetary income. A lot of analysts and consultants are employed by businesses to evaluate several variables from different departments to understand, research and then make decisions that improve their businesses. For example, Debattista et. al. [9] discussed the potential of Linked Data methods to tackle all five Vs. Hence, value is one of the dimensions that are critical to business decisions and to the essential interest of the BD world.

Seven Vs

Another concept that needs to be addressed to complete this research is the discussions that were taking place between researchers on the seven Vs of big data. There are many scholars debating seven Vs, and one of the best examples is Khan, Uddin, and Gupta's paper [10]. This paper concisely revealed the seven Vs. The seven characteristics according to them are: Volume, Velocity, Variety, Veracity, Value, Validity, and Volatility. The two characteristics added are Validity and Volatility where:

Validity is the accuracy and correctness of the data with respect to the intended use. Although it sounds so similar to veracity, they are two different concepts. A data set may have no problem with veracity, but it may still not be valid. In other words, without validating it, we cannot simply take a data set, and trust it.

Volatility is the responsibility for rapid and unexpected transitions. There are many businesses that have openly admitted that they are not storing older data that have no value. Online companies, for example, may not want to keep older consumer purchasing history, since the warranty may expire. It is important to ensure the volatility of a data set to allow complete reliability of the final outcome.

Ten Vs

Interestingly, Ranjan [11] is one of the researchers who discussed ten Vs of BD in his paper. The 10 Vs are: Volume, Velocity, Variety, Veracity, Value, Validity, Volatility, Variability, Visualization, and Vulnerability. They are BD characteristics to explore whether those properties hold the key to successful implementation of Big Data projects. Ranjan added Variability, Visualization and Vulnerability to this list of characteristics to the previous studies, where:

Variability may sound similar to Variety but is used over time to calculate the accuracy of incoming data. In Big Data Analytics, Variability refers to the inconsistencies in the results. Due to the multitude of data dimensions resulting from multiple disparate data types and sources, Big data can be also variable. An example of this is that airlines have a moderate level of variability since social customer data may be highly variable, flight data is fairly streamlined except in situations such as bad weather, engine problems, etc. In order for any meaningful analytics to occur, anomaly and outlier detection methods should be applied.

Visualization involves the presentation of data of almost any type in a graphical format that makes it easy to understand and interpret data. Choosing the wrong visual aid or simply defaulting to the most common type of data visualization could cause confusion with the viewer or lead to mistaken data interpretation. As one of the struggles that slows down reporting and analysis is understanding what types of graphs to use and why.

Vulnerability relates to the security measures that need to be put in place so that the data collected is processed in accordance with the legislation and the wishes of the customer. Ranjan [11] has also cleverly illustrated how each industry could be susceptible to a different level.

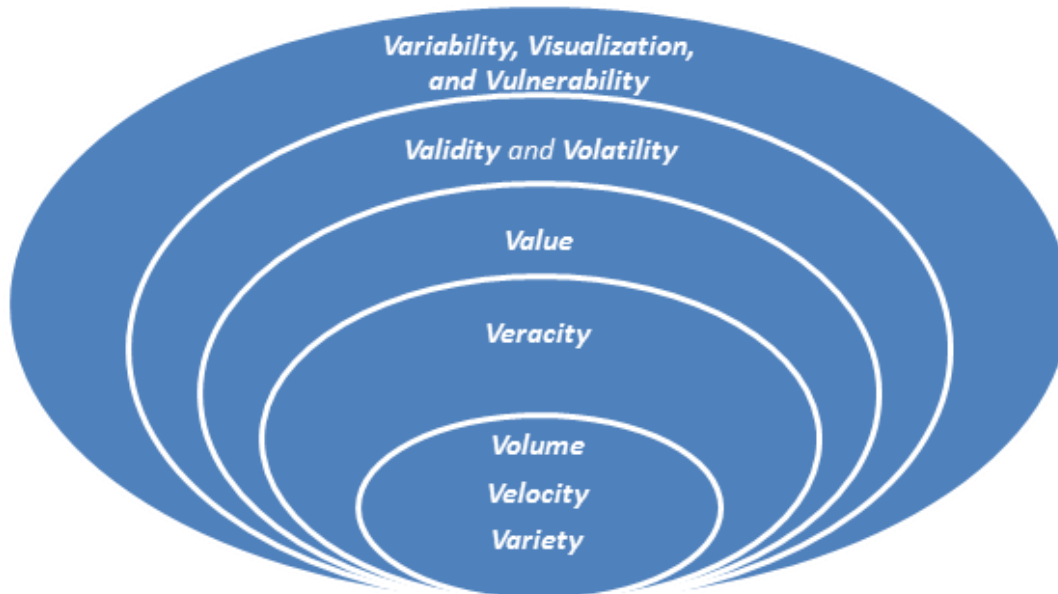


Figure (1). Big Data Characteristics growth.

Figure (1) presents the growth in the V's mentioned in this section. In the study conducted by Husamaldin and Saeed [12] to conceive a correlation taxonomy of existing methods and associated techniques. They introduced more Vs such as vision, verification, confirmation, uncertainty, location, vocabulary and vagueness, complexity and immutability. This shows that there are many more characteristics which are not officially specified yet.

Even though researchers had introduced many new Vs, it is, however, unbeatable for researchers to agree on the three main characteristics, which are the Volume, Velocity and Variety. Such characteristics were all first identified in the researchers work and regularly highlighted as the first three.

3. Methodology

A systematic review was conducted for capturing relevant analytics literature from different academic sources, focusing on the following objectives:

1. To explore the growth in the numbers of Big Data Analytics characteristics.
2. To identify the BDA characteristic numbers used in different industry sectors.

This investigation has supported the understanding of the overall number of Big Data Analytics characteristics needed in various domains.

Academic information collection resources:

After thorough review, research for articles was made on the following databases: ScienceDirect, IGI/global, IEEE Xplore and Taylor & Francis, Google Book, SAGE pub, researchgate, Springer.

Selection criteria

To capture the literature relevant to the research interest, the articles published in English between 2018 and 2021 had primary emphasis on Big Data Analytics characteristics in retail, agriculture, manufacturing, public sector and public healthcare.

The research procedure started with research for publications on electronic databases containing keywords “big data analytics characteristics” or “big data analytics categories”, and mapped to the previous sectors “retail”, “agriculture”, “manufacturing”, “public sector” and “public healthcare”. Then, a further step was taken to examine the title, abstract and keywords of identified articles on the basis of the selection criteria to ensure only the high-quality cited published research papers are selected. The final step is a thorough study of the chosen papers.

4. Results and Discussion

After the scrutiny of the literature papers Table (1) created to present the recent studies discussing Big Data characteristics in various sectors. Two examples over the last four years for each sector were given. The sectors considered were: retail, agriculture, manufacturing, public sector and public healthcare. The public sector focused on: oil and gas, healthcare, telecommunications, law enforcement, waste management and education.

Table (1) presents the recent and largest BD characteristics considered in literature for different sectors. The tables reveal many interesting points to highlight:

First, it is noticeable that the characteristics have been increased over the years. For example, in the retail sector, in 2019 Anshari et.al [13] have considered the use of 5 V's (Volume, Variety, Velocity, Veracity, and Value) in their research. While, in 2020 Holmlund et.al. [14] have discussed the use of 7 V's (Volume, Velocity, Variety, Veracity, Variability, Visualisation, and Value).

Second, the V's number differs in the sectors. It is noted that some sectors were considering more V's in their research than others. For example, the “Telecommunication” sector considers 3 V's where as “Retail” and “Oil and Gas” consider 7 V's. This could be related to the type of the data treated and needed to improve the decisions in that sector.

Third, the V's used in the same sector could be different. For example, the examples for the “Telecommunication” sector both consider 3 V's but there are different characteristics. Wang and Jones [23] consider Veracity as the third characteristic where Zahid et.al. [24] consider Variety as the third characteristic. Same was noticed in the “Health Care” sector when Saranya and Asha [22] introduce Variability in the 5 V's instead of “Variety” the usual fourth characteristic.

Fourth, Table (2) below. clearly shows that the characteristics that are having a significant effect on many sectors (Agriculture, Manufacturing, Health Care, Law enforcement & Education) are the five V's (Volume, Velocity, Variety (or Variability), Veracity and Value).

Finally, Table 2 ranks the sectors according to their engagement with the growth of the V's. As shown “Retail” and “Oil and Gas” sectors were at the top whereas “Telecommunication” sector was the last with 3 V's.

Table (1). Big Data Characteristics in Sectors.

| No. | Sector | No. of V's | The V's | Studies By |
|-----|-------------------|------------|---|-----------------------------|
| 1 | Retail | 5 V's | (1) Volume, (2) Velocity, (3) Variety, (4) Veracity, and (5) Value. | Anshari et.al. [13] |
| | | 7 V's | (1) Volume, (2) Velocity, (3) Variety, (4) Veracity, (5) Value, (6) Variability, and (7) Visualisation. | Holmlund et.al. [14] |
| 2 | Agriculture | 3 V's | (1) Volume, (2) Velocity, and (3) Variety. | Lioutas and Charatsari [15] |
| | | 5 V's | (1) Volume, (2) Velocity, (3) Variety, (4) Veracity, and (5) Value. | Tseng et.al. [16] |
| 3 | Manufacturing | 3 V's | (1) Volume, (2) Velocity, and (3) Variety. | Moktadir et.al. [17] |
| | | 5 V's | (1) Volume, (2) Velocity, (3) Variety, (4) Veracity, and (5) Value. | Shukla et.al. [18] |
| 4 | Oil and Gas | 5 V's | (1) Volume, (2) Velocity, (3) Variety, (4) Veracity, and (5) Value. | Tewari and Dwivedi [19] |
| | | 7 V's | (1) Volume, (2) Velocity, (3) Variety, (4) Veracity, (5) Value, (6) Variability, and (7) Visualisation. | Xu et.al. [20] |
| 5 | Health Care | 4 V's | (1) Volume, (2) Velocity, (3) Variety and (4) Veracity | Dash et.al. [21] |
| | | 5 V's | (1) Volume, (2) Velocity, (3) Variability, (4) Veracity, and (5) Value. | Saranya and Asha [22] |
| 6 | Telecommunication | 3 V's | (1) Volume, (2) Variety, and (3) Veracity. | Wang and Jones [23] |
| | | 3 V's | (1) Volume, (2) Velocity, and (3) Variety. | Zahid et.al. [24] |
| 7 | Law enforcement | 3 V's | (1) Volume, (2) Velocity, and (3) Variety. | Hannah-Moffat [25] |
| | | 5 V's | (1) Volume, (2) Velocity, (3) Variety, (4) Veracity, and (5) Value. | Capgemini UK [26] |
| 8 | Waste Management | 3 V's | (1) Volume, (2) Velocity, and (3) Variety. | Xu et.al. [27] |
| | | 4 V's | (1) Volume, (2) Velocity, (3) Variety and (4) Value. | Lu et.al. [28] |
| 9 | Education | 3 V's | (1) Volume, (2) Velocity, and (3) Veracity | Jules and Salajan [29] |
| | | 5 V's | (1) Volume, (2) Velocity, (3) Variety, (4) Veracity, and (5) Value. | Miah et.al. [30] |

Table (2). Sectors Ranking according to their engagement with the growth of the V's.

| Rank | Sector/s | No. of V's |
|--------|--|------------|
| First | Retail & Oil and Gas | 7 V's |
| Second | Agriculture, Manufacturing, Health Care, Law enforcement & Education | 5 V's |
| Third | Waste Management | 4 V's |
| Fourth | Telecommunication | 3 V's |

5. Conclusions

As the word self explains, it is very difficult to describe Big Data given its huge size and the categorizations are there to differentiate its characteristics. The Big Data types can be even more complicated. This paper took into account many articles and research which had different views on the number of Big Data Vs. Many researchers had some valuable inputs on the different characteristics and most of the authors agreed that there are more and more categorizations that can be considered since big data is massive. Due to the complexity and scale of Big Data, all the other characterisations have been added over the year and play a major role in analytics and their sectors' needs. In this paper, based on related work over the past years, the authors have introduced the growth in BD V's from three, four, five, seven then ten data with clear definition for each characteristic. The paper also presented clear examples of the recent use of V's by the industry sectors. A Final conclusion, it is worth mentioning that most of the characteristics have recently been introduced, which indicates that the use of those by the industry sectors are still in development and not enough related studies have been found to strongly support such characteristics. Also, it was noted that some sectors are more advanced in considering the new V's than others and they select their V's based on the data treatments and demands.

References

- [1] S. Sagiroglu and D. Sinanc, "Big data: A review", 2013 International Conference on Collaboration Technologies and Systems (CTS), pp. 42-47, San Diego, CA, US.2013 doi: 10.1109/CTS.2013.6567202.
- [2] J. Johnson, S. Friend and H. Lee, "Big Data Facilitation, Utilization, and Monetization: Exploring the 3Vs in a New Product Development Process", Journal of Product Innovation Management, vol. 34, no. 5, pp. 640-658, 2017. Available: 10.1111/jpim.12397 [Accessed 21 May 2021].
- [3] M. Ghasemaghahi, "Understanding the impact of big data on firm performance: The necessity of conceptually differentiating among big data characteristics", International Journal of Information Management, vol. 57, p. 102055, 2021. Available: 10.1016/j.ijinfomgt.2019.102055.
- [4] W. Fan and A. Bifet, "Mining big data", ACM SIGKDD Explorations Newsletter, vol. 14, no. 2, pp. 1-5, 2013. Available: 10.1145/2481244.2481246.
- [5] X. Dong and D. Srivastava, "Big data integration", 2013 IEEE 29th International Conference on Data Engineering (ICDE), 2013. Available: 10.1109/icde.2013.6544914 [Accessed 21 May 2021].
- [6] L. Berti-Equille and M. Ba, "Veracity of Big Data", Journal of Data and Information Quality, vol. 7, no. 3, pp. 1-3, 2016. Available: 10.1145/2935753.
- [7] J. Kepner, V. Gadepally, P. Michaleas, N. Schear, M. Varia, A. Yerukhimovich and R. K. Cunningham, "Computing on masked data: a high performance method for improving big data veracity", 2014 IEEE High

- Performance Extreme Computing Conference (HPEC), 2014. Available: 10.1109/hpec.2014.7040946 [Accessed 21 May 2021].
- [8] T. Nguyen, "A Framework for Five Big V's of Big Data and Organizational Culture in Firms", 2018 IEEE International Conference on Big Data (Big Data), 2018. Available: 10.1109/bigdata.2018.8622377 [Accessed 21 May 2021].
- [9] J. Debattista, C. Lange, S. Scerri and S. Auer, "Linked 'Big' Data: Towards a Manifold Increase in Big Data Value and Veracity", 2015 IEEE/ACM 2nd International Symposium on Big Data Computing (BDC), 2015. Available: 10.1109/bdc.2015.34 [Accessed 21 May 2021].
- [10] M. Khan, M. Uddin and N. Gupta, "Seven V's of Big Data understanding Big Data to extract value", Proceedings of the 2014 Zone 1 Conference of the American Society for Engineering Education, 2014. Available: 10.1109/aseezone1.2014.6820689 [Accessed 21 May 2021].
- [11] J. Ranjan, "The 10 Vs of Big Data framework in the Context of 5 Industry Verticals", PRODUCTIVITY, vol. 59, no. 4, pp. 324-342, 2019. Available: 10.32381/prod.2019.59.04.2.
- [12] L. Husamaldin and N. Saeed, "Big Data Analytics Correlation Taxonomy", Information, vol. 11, no. 1, p. 17, 2019. Available: 10.3390/info11010017.
- [13] M. Anshari, M. Almunawar, S. Lim and A. Al-Mudimigh, "Customer relationship management and big data enabled: Personalization & customization of services", Applied Computing and Informatics, vol. 15, no. 2, pp. 94-101, 2019. Available: 10.1016/j.aci.2018.05.004.
- [14] M. Holmlund, Y. Van Vaerenbergh, R. Ciuchita, A. Ravald, P. Sarantopoulos, F. Ordenes and M. Zaki, "Customer experience management in the age of big data analytics: A strategic framework", Journal of Business Research, vol. 116, pp. 356-365, 2020. Available: 10.1016/j.jbusres.2020.01.022.
- [15] E. Lioutas and C. Charatsari, "Big data in agriculture: Does the new oil lead to sustainability?", Geoforum, vol. 109, pp. 1-3, 2020. Available: 10.1016/j.geoforum.2019.12.019.
- [16] F. Tseng, H. Cho and H. Wu, "Applying Big Data for Intelligent Agriculture-Based Crop Selection Analysis", IEEE Access, vol. 7, pp. 116965-116974, 2019. Available: 10.1109/access.2019.2935564.
- [17] M. Moktadir, S. Ali, S. Paul and N. Shukla, "Barriers to big data analytics in manufacturing supply chains: A case study from Bangladesh", Computers & Industrial Engineering, vol. 128, pp. 1063-1075, 2019. Available: 10.1016/j.cie.2018.04.013.
- [18] N. Shukla, M. Tiwari and G. Beydoun, "Next generation smart manufacturing and service systems using big data analytics", Computers & Industrial Engineering, vol. 128, pp. 905-910, 2019. Available: 10.1016/j.cie.2018.12.026.
- [19] S. Tewari and U. Dwivedi, "Ensemble-based big data analytics of lithofacies for automatic development of petroleum reservoirs", Computers & Industrial Engineering, vol. 128, pp. 937-947, 2019. Available: 10.1016/j.cie.2018.08.018.
- [20] C. Xu, S. Misra, P. Srinivasan and S. Ma, "When Petrophysics Meets Big Data: What can Machine Do?", Society of Petroleum Engineers SPE Middle East Oil and Gas Show and Conference, 18-21 March 2019, Day 4 Thu, Manama, Bahrain. Available: 10.2118/195068-ms.
- [21] S. Dash, S. Shakyawar, M. Sharma and S. Kaushik, "Big data in healthcare: management, analysis and future prospects", Journal of Big Data, vol. 6, no. 1, 2019. Available: 10.1186/s40537-019-0217-0.
- [22] P. Saranya and P. Asha, "Survey on Big Data Analytics in Health Care", 2019 International Conference on Smart Systems and Inventive Technology (ICSSIT), 2019. Available: 10.1109/icssit46314.2019.8987882 [Accessed 22 May 2021].
- [23] L. Wang and R. Jones, "Big Data Analytics in Cyber Security: Network Traffic and Attacks", Journal of Computer Information Systems, pp. 1-8, 2020. Available: 10.1080/08874417.2019.1688731.
- [24] H. Zahid, T. Mahmood, A. Morshed and T. Sellis, "Big data analytics in telecommunications: literature review and architecture recommendations", IEEE/CAA Journal of Automatica Sinica, pp. 1-22, 2019. Available: 10.1109/jas.2019.1911795.

- [25] K. Hannah-Moffat, "Algorithmic risk governance: Big data analytics, race and information activism in criminal justice debates", *Theoretical Criminology*, vol. 23, no. 4, pp. 453-470, 2018. Available: 10.1177/1362480618763582.
- [26] Capgemini UK. *The Future of Policing In A Data-Driven World*, 2019. [online] Available at: <<https://www.capgemini.com/gb-en/2019/11/the-future-of-policing-in-a-data-driven-world/>> [Accessed on 22 May 2021].
- [27] J. Xu, W. Lu, M. Ye, F. Xue, X. Zhang and B. Lee, "Is the private sector more efficient? Big data analytics of construction waste management sectoral efficiency", *Resources, Conservation and Recycling*, vol. 155, p. 104674, 2020. Available: 10.1016/j.resconrec.2019.104674.
- [28] W. Lu, C. Webster, Y. Peng, X. Chen and K. Chen, "BIG DATA IN CONSTRUCTION WASTE MANAGEMENT: PROSPECTS AND CHALLENGES", *Detritus*, vol., no. 0, p. 1, 2018. Available: 10.31025/2611-4135/2018.13737.
- [29] T. Jules and F. Salajan, *The educational intelligent economy*. Bingley, U.K.: Emerald Publishing Limited, 2019.
- [30] S. Miah, M. Miah and J. Shen, "Editorial note: Learning management systems and big data technologies for higher education", *Education and Information Technologies*, vol. 25, no. 2, pp. 725-730, 2020. Available: 10.1007/s10639-020-10129-z.