**Use big data analytics for management of production**

by

**Mawande Skibi**

Submitted in partial fulfilment of the

requirements for the degree

**Master of Engineering**

in

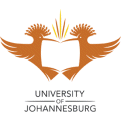
**Industrial Engineering**

**Faculty of Engineering and the**

**Built Environment**

at the

**UNIVERSITY OF JOHANNESBURG**



**Supervisor: Mrs. Steenkamp**

**December**

**2020**

**Plagiarism declaration**

* I, Mawande Skibi, hereby declare that this dissertation is wholly my work and has not been submitted anywhere else for academic credit either by myself or another person.
* I understand what plagiarism implies and declare that this dissertation is my ideas, words, phrases, arguments, graphics, figures, results, and m organisation except where reference is explicitly made to another’s work.
* I understand further that any unethical academic behaviour, which includes plagiarism, is seen in a serious light by the University of Johannesburg and is punishable by disciplinary action.

**Signed**: …………………………………….

**Date**: ……………………………………….

**© 2020 University of Johannesburg**

**Publications**

1. Mawande Sikibi

Use green initiatives for management of logistics industry

SAIIE31 Proceedings,

5th – 7th October 2020,

Muldersdrift, South Africa

© 2020 SAIIE

**\* SAIEE31 Award for best Panel Session**

[**https://conferences.sun.ac.za/index.php/saiie31/saiie31/author/submissionReview/4487**](https://conferences.sun.ac.za/index.php/saiie31/saiie31/author/submissionReview/4487)

**Acknowledgment**

First, I would like to express my greatest thanks to God for His blessing in completing this research study. My highest gratitude goes to my supervisor for all her assistance, guidance, advice, and patience throughout this research study. She had an active interest in this study, and her insight helped me formulate, redefine, and implement my approach to this study. Her patience in guiding me in doing this research is something I admire.

I express my gratitude to the company who co-operated in this study, for without management of production co-operation in this research study, could not possibly be completed. One of the pleasant aspects of preparing for the research study is the opportunity to thank those who have contributed to making this study possible.

I am also thankful to other seen and unseen hands that have given me direct and indirect assistance in completing this study. From the bottom of my heart, thank you very much. Finally, thanks most profoundly to the production managers for their understanding and permission to research in their company.

**Table of contents**

Plagiarism declaration……………......…………….........….........……………………...…………i

[Publications ii](#_Toc57135179)

[Acknowledgment iii](#_Toc57135180)

[List of figures viii](#_Toc57135181)

[List of tables ix](#_Toc57135182)

[List of appendices x](#_Toc57135183)

[List of acronyms xi](#_Toc57135184)

[Abstract xii](#_Toc57135185)

[Chapter 1 Introduction 1](#_Toc57135186)

[1. Introduction 1](#_Toc57135187)

[1.1. Background of the study 2](#_Toc57135188)

[1.2. Problem Statement 2](#_Toc57135189)

[1.3. Research Questions 3](#_Toc57135190)

[1.4. Aim 3](#_Toc57135191)

[1.5. Objectives 3](#_Toc57135192)

[1.6. Significant of the study 4](#_Toc57135193)

[1.7. Definitions of terminology 7](#_Toc57135194)

[Chapter 2 Literature review 10](#_Toc57135195)

[2.1. Big data 10](#_Toc57135196)

[2.2. Data storage 13](#_Toc57135197)

[2.2.1. Data file 14](#_Toc57135198)

[2.2.2. Database 15](#_Toc57135199)

[2.2.3. Sensor data 15](#_Toc57135200)

[2.3. Data generated 17](#_Toc57135201)

[2.3.1. Machine-generated 17](#_Toc57135202)

[2.3.2. Organisation-generated 17](#_Toc57135203)

[2.4. Data analytics 18](#_Toc57135204)

[2.4.1. Predictive analytics 20](#_Toc57135205)

[2.5. Research Gap 21](#_Toc57135206)

[Chapter 3 Theoretical framework 23](#_Toc57135207)

[3.1. Big data characteristics 23](#_Toc57135208)

[3.1.1. Classification of big data 25](#_Toc57135209)

[3.1.2. Big data analytics 33](#_Toc57135210)

[3.2. Big data strategies 38](#_Toc57135211)

[3.2.1. Select correct data 39](#_Toc57135212)

[3.2.2. Build models that predict and optimise company outcomes 40](#_Toc57135213)

[3.2.3. Transform capabilities of the company 40](#_Toc57135214)

[3.3. Produce a positive outcome through big data 41](#_Toc57135215)

[3.3.1. Cost reduction 41](#_Toc57135216)

[3.3.2. Quick decision making 42](#_Toc57135217)

[Chapter 4 Research Methodology 43](#_Toc57135218)

[4.1. Research design 43](#_Toc57135219)

[4.2. Research methods 44](#_Toc57135220)

[4.2.1. Research Approach 45](#_Toc57135221)

[4.2.2. Research study approach 46](#_Toc57135222)

[4.2.3. Participant 46](#_Toc57135223)

[4.2.4. Research strategy of this study 47](#_Toc57135224)

[4.2.5. Sources of data collection 47](#_Toc57135225)

[4.2.6. Data collection 48](#_Toc57135226)

[4.2.7. Data analysis method 49](#_Toc57135227)

[4.3. Research credibility 54](#_Toc57135228)

[4.4. Research validity 54](#_Toc57135229)

[4.4.1. Statistical validity 55](#_Toc57135230)

[4.5. Justification 55](#_Toc57135231)

[4.6. Ethical considerations 55](#_Toc57135232)

[4.7. Summary 56](#_Toc57135233)

[Chapter 5: Findings and Results 57](#_Toc57135234)

[5.1. Findings of the study 57](#_Toc57135235)

[5.1.1. Machine downtime 57](#_Toc57135236)

[5.1.2. Equipment downtime 57](#_Toc57135237)

[5.1.3. Machine error log messages 58](#_Toc57135238)

[5.1.4. Product Order data 58](#_Toc57135239)

[5.1.5. Worker absenteeism data 58](#_Toc57135240)

[5.1.6. Alarm management data 59](#_Toc57135241)

[5.2. Results 59](#_Toc57135242)

[5.2.1. Machine downtime results 59](#_Toc57135243)

[5.2.2. Equipment downtime results 60](#_Toc57135244)

[5.2.3. Machine error log message results 60](#_Toc57135245)

[5.2.4. Product order data results 61](#_Toc57135246)

[5.2.5. Worker absenteeism data 61](#_Toc57135247)

[5.2.6. Alarm management data results 61](#_Toc57135248)

[Chapter 6 Recommendations and Conclusions 63](#_Toc57135249)

[6.1. Research process summary 63](#_Toc57135250)

[6.2. Summary findings 64](#_Toc57135251)

[6.2.1. Data set theme 1 64](#_Toc57135252)

[6.2.2. Data set theme 2 64](#_Toc57135253)

[6.2.3. Data set theme 3 65](#_Toc57135254)

[6.2.4. Data set theme 4 65](#_Toc57135255)

[6.2.5. Data set theme 5 66](#_Toc57135256)

[6.2.6. Data set theme 6 66](#_Toc57135257)

[6.3. Implications for research 66](#_Toc57135258)

[6.3.1. Data quality 67](#_Toc57135259)

[6.3.2. Data relevance 67](#_Toc57135260)

[6.3.3. Data cohesion 68](#_Toc57135261)

[6.4. Implication for practice 68](#_Toc57135262)

[6.5. Contribution to the study 69](#_Toc57135263)

[6.6. Limitations and delimitations of the study 69](#_Toc57135264)

[6.7. Recommendation for future research 70](#_Toc57135265)

[6.8. Conclusion 71](#_Toc57135266)

[7. References 72](#_Toc57135267)

[8. Appendices 86](#_Toc57135268)

**List of figures**

[Figure 1: Classification of big data 24](#_Toc57138583)

[Figure 2: Data analytic technique 35](#_Toc57138584)

[Figure 3: The relation between decision sciences and big data 38](#_Toc57138585)

[Figure 4: Types of research designs sources 44](#_Toc57138586)

[Figure 5: Data quality measures 51](#_Toc57138587)

**List of tables**

[Table 1: Characteristics of data 25](#_Toc57138606)

[Table 2: The definitions of big data analytics 33](#_Toc57138607)

**List of appendices**

[Appendix A: Authorization letter 86](#_Toc57138583)

**List of acronyms**

BI - Business intelligence

DBMS - Database Management System

PM - Preventative Maintenance

**Abstract**

Nowadays, data is indispensable in the production industry sector. Big data analytics and initiatives are used to gain insights that could assist to make strategic decisions to evaluate this data. The concept connects to Industry 4.0, which emerged at the commerce of the 21st century. Technology giants can use big data analytics technologies to forecast future outcomes. For the large volume of data and voluminous databases that can be structured, semi-structured, or unstructured, the term big data is widely used for a plethora of advanced digital techniques designed to reveal trends. This study aims to use big data analytics for management to predict future outcomes. However, these analytics examines the large data sets of production processes to underline insights and patterns. The study uses an intrinsic case study to better understand the methods used to extract the database's data to achieve superior results. The researcher participated in the process in an active participatory manner and demonstrated to the company the use of predictive analytics to predict and analyse big data from various big data sets to gain insights and patterns. These methods managed to use historical data and turn it into actionable information about what may happen. The results obtained from historical data assisted production management in making strategic decisions in the selected data sets. This study proposes predictive analytic as the method for big data to predict the future conditions that drive the production industry.

**Keywords**

Analytics, Big data, Database management system, Data sets, Effective, Efficient, Insight, Organisation, Semantics.

**Chapter 1 Introduction**

**1. Introduction**

The world is becoming more and more digitised, and most data is in digital format. This is not the data era; this is the big data era. Big data has gained momentum now, more than ever before. Data is the key to optimise industries. Having an obvious strategy is useful when considering massive amounts of data. A better approach is about what the company wants to obtain out of the data. At this moment, it does not matter what the data industry has or has collected. Many processes can be achieved on a single platform using big data. With big data software assistance, companies can store vast amounts of data, analyse the data, and visualise it. The use of big data allows production management to make informed decisions. Companies that rely, as guides, on big data are more advanced than others. Big data can handle a range of resources with the correct tools and methods. Production companies that have managers who can support significant data initiatives could be more proactive.

Using big data requires several changes in the production industries. Therefore, the correct attitude of managers towards significant data initiatives depends on the management characteristics. The support of management is important in advancing the use of substantial data initiatives. Not all generated data are useful because generally only 33% of all data is used. However, the average organisation processes 0.5% of data collected from the storage platform. This information is considered the most valuable element, yet 32.5% of the data is an essential resource for decision-making. The organisations do not use the remaining 32.5% to provide insight. Big data for management production is necessary; the importance of understanding how to use big data to gain value from the raw data produced.

Big data, which is possible because of the extensive data produced from the production processes, significantly affects the production industry. The behaviour of managers is an important message in production. Big data does not revolve around the amount of data available but how the management of production handles it.

**1.1. Background of the study**

The more information is processed, the more data is created. This results in more data that should be organised for decision-making, as Larson (2013) stipulated. Shin et al. (2015) explore how the effect of overloaded information on big data can be managed. The information obtained from multiple sources poses another obstacle. Wagner-Muns et al. (2017) noted that the challenge and the pace and variety of information is data volume. In addition, Jagadish et al. (2014) indicated that big data consists of a multi-step value generation process, such as data extraction, data cleansing, and data acquisition.

In the early 2000s, the widespread use or use of big data in production gained momentum, producing vast data volumes. As Gartner (2015) claims, the number of technology applications would reach 20 billion by 2020. Database management technology is one of the technology applications used by big data. This is used for the analysis of large volumes of production-extracted data. Chaurhuri et al. (2011) also stressed that this application could be used by management to make sense of large data sets.

Consequently, this has become increasingly important for the accumulation of more extensive data sets. Many of the different sources, as advocated by Chen et al. (2012), barely existed 20 years ago. Bumblauskas et al. (2017) extended the importance of using big data sets in manufacturing to generalize the argument 'We can generate anything.' Bumblauskas et al. (2017) Dubey et al. (2020) observe the effect on maintaining big data productivity. In the manufacturing sector, not using big data could improve because it assumes that the raw data is worth retaining.

**1.2. Problem Statement**

The company faces the challenges of data that are growing at a faster rate. It influences their storage platforms, which affects the storage capacity. Data collection developments have widened the field of business intelligence (BI). To enforce data to allow access to the reports, the organisation uses traditional storage platforms. The modern method, on the other hand, views data governance. These storage platforms receive information on each interaction daily, allowing managers to have thousands of interlocking data sets. This data has become an obstacle in the storage platforms, and the company does not know how to use this data to advance company strategic goals. There is a widespread perception in the company that data are purely for analyst’s use.

**1.3. Research Questions**

Develop a general overview to describe the industrial sector in big data analytics technologies. Big data is one of the leading technologies that can provide a research roadmap and solve industrial challenges. The research question is critical for companies that have deficiencies in data management and do not use the latest big data analytics technologies. As far as the researchers know, while big data analytics have laid the foundations for technology and science growth, no comparable research has been found. The following research questions require to be addressed to provide solutions to the company's overall data growth rate:

* Does the company understand how to manage big data analytics?
* What is the analytical technique that would be suitable for managing big data?
* How would the company govern the data and change the future data outcome?

**1.4. Aim**

This research study aims to use the modernized approach to present the advantages of big data analytics. This strategy will handle the production environment, which is crucial for solving problems and uncovering new opportunities. This would hold a potential competitive advantage among other companies, provided the data is used effectively.

**1.5. Objectives**

The following approaches for the study are required to achieve the aim mentioned above.

* To illustrate to the company the advantage of using big data analytics to manage production data.
* To use the analytical technique for predicting the data that the company stored.
* To guide the company on how to govern the data for the future outcome behaviour of data.

**1.6. Significance of the study**

Technology is indispensable nowadays. To gain insights and to assist with strategic decision making, big data technologies are emerging to analyse this data (Bibri et al., 2020). Big data analytics, in data science, is a revolution. The company's use of data mining increases data monthly. At the commencing of the 21st century, these principles were developed. The technology giants have recently been pursuing the use of big data technologies. Big data refers to large, structured, or unstructured data. This large amount of data is generated by the organization daily. Big data analytics’ perspective explores massive data sets (Mergel et al., 2016).

Analytics requires advanced techniques and data processing that can be derived from multiple sources of varying sizes. Big data is characterized by a wide variety, volume, and velocity. The data sets come from different sources, including audio, video, logs, and many others. In big data analytics, analytic methods such as time series and data mining are used. The data provides an analysis for the company to be collected, prepared, and blended. In multiple ways, this data can be widely used. The unique analytics are used to evaluate the origins of data for new insights using these techniques. Data is typically data generated on a wide scale in real-time.

Many researchers have tried to explore big data, analyse the phenomenon, and understand how company data resources can generate and gain value. Studies are still expanding in this area, and more researchers are focusing their attention on the fields of big data and analytics. Several research studies in this area have continued to increase, offering a strong incentive to synthesize the main research streams. These critically examine the general area to identify streams that have not attracted attention despite the obvious practical requirements. Big data has been a crucial foundation for growth in productivity and innovation (Sun, 2020). Big data has gained the most attention in the sector and synthesizes the accumulated information. This sector has existed, the production company deficiency knowledge of data, and innovative changes.

Big data offers new possibilities for those who use it, paying attention to big data analytics to enable growth and management and support society (Guo et al., 2017). A modern understanding of the environment is made possible using big data technologies to create inferential tools to analyse data findings and develop algorithms for advanced correlations. Overall, society is benefiting from the emergence of big data. The extraordinary amount of knowledge and the profit that the business can derive from it is the subject of big data in society. Data science, known as predictive analytics, may explain data mining technologies to handle vast data volumes. Predictive analytics succinctly learn from data to predict how the data will behave in the future. To consider the probability effects on data mining detects trends in knowledge sets.

Big data analytic techniques are rising in demand as is the interest in big data. The company can discover new opportunities and gain new insights to operate effectively. Using big data is necessary to obtain ideal results from analytics. However, using the standard technique cannot handle such a large amount of data. However, it may take many days for companies to use a simple technique in a certain way to obtain accurate results. The company's ability to use data in various ways to generate useful insights can be done on a large scale, but not on a smaller scale. Therefore, to extract new insights from the information, the company that can use big data usually hire data scientists to process it. Using big data allows the company to move faster and manage resources more effectively (Mazzei et al., 2017). To make better company decisions to provide meaningful details to manage resources. Retaining in mind the critical aspect of using big data analytics, the company will develop strategies. Big data analytics assists operations to become more effective (Kibria et al., 2018). This further improves the company's effectiveness. The new analytics methodology makes data analytics simpler and quicker (Sun et al., 2018). These analytics speed up decision-making and saves time and energy. Data analysis has become an essential part of the company more than ever. This research is of great significance because it has drawn attention to production management and data knowledge. It plays a vital role in the company's effective response to changes, increasing productivity, and paving the way for development and innovation. Many scientific studies have resolved the importance of data knowledge initiatives to the organization. This research is of great importance because it has drawn attention to the management of production and data knowledge. To make informed decisions and assist strategic plans, use big data.

Using big data can assist the company in understanding the privilege of owning data. A company that uses big data may benefit more than other companies. Big data initiative plays a crucial role in growth and science; it can support the theoretical development of research and benefit society's production (Duan et al., 2019). Big data is transforming the way the manufacturing sector operates. Big data creates job opportunities for personnel who have data skills.

The professional personnel are on the rise with the skill set, interest, and investment in big data technology. It creates great opportunities for people in this industry. Recently, it is challenging to find any company that does not have a presence on the storage platform; the organisation will require experts in data analytics. This makes it a smart career change that has a corporate future. Big data presents many promising opportunities to differentiate the production industry (Dekhtiar et al., 2018). Big data provides hope to the company that uses data of any size and volume, and it has become an advantage to the company. The company can use big data effectively to enhance production. For this reason, for analysts, the career opportunities in this field are unlimited. A company with big data technologies interests can bring job opportunities for big data analytics skills (De Mauro et al., 2018).

Big data analytics significance contributes to intense competition and increased demand for big data personnel (Ghani et al., 2019). For a company to use the information for transformation, data analytics is essential. The use of analytics will increase analysts' knowledge of the industry. Data analytics offers companies an opportunity to learn about the company's prospects. Big data analytics in various fields and industries, criteria and importance exist. Therefore, remaining up to date with this method is important for a business. In the meantime, using this analytical method correctly, the company will achieve some advantages. The data is as complex as recent without big data analytics; the data set is large. Big data analytics tools allow the company to identify incorrectly and introduce new strategies (Grover et al., 2018). However, big data is so much deeper and broader. Currently, big data can excel as an advantage in practice for almost any purpose.

Big data and advanced analytics have advanced the world, not because the data is big, but because it is potentially important. There are few fields where more creativity and investment encounters than big data analytics. New and improved data analysis methods provide ways of coping with the challenges of achieving size. Technologies and strategies for data analytics offer a tool to analyse data sets and take away new information. There is a tremendous opportunity to invest and incorporate data analysis tools and techniques, and the company will require to adapt, innovate, and use decision-making methods. Big data analytics in the scientific world is a revolution (Bragazzi et al., 2016). The company uses of big data analytics is growing every month.

Big data companies may continue to drive the technology revolution. The value of using big data analytics is illustrated in this dissertation review, contributing to more competition. In the production sector, there are massive demands for and significant importance of big data analytics. To remain conscious of these strategies are necessary. Hence, this study is essential to remain aware of these techniques. Using these analytics tools correctly, the company will gain much in their area. Exploiting this data hopes to demonstrate this potential. Using analytics, companies can learn how to use big data and plan to lead companies more effectively. This can be view to function in real-world problem-solving.

**1.7. Definitions of terminology**

The research terminologies are essential to understand the critical concept, mainly the following chosen terms that are unusually understood. These specific terms below are defined:

**Organisation**- is an organised group of businesses or managers to meet to pursue collective goals.

**Big data analytics** - The method of analysing big data to reveal knowledge is a dynamic one. This applies to the way a large amount of data or big data is processed. George et al. (2016) posit that big data refers to collected and handled large and varied data. By comparison, data science creates models that capture, visualize, and analyse the data's underlying patterns.

**Digitalisation** - Conversion of information into a digital format is the operation. Ross (2017) suggests that digitisation means switching from analogy-to-digital data to streaming current processes.

**Structured data -** refers to any data within a record or file in a fixed field. It follows a standard order in which data conforms to a data model and easily accessed (Brinkmann, 2016).

**Semi-structured data -** is the type of structured data that does not obey the tabular structure of relational database related data models.

**Database management system (DBMS)** – The software uses the information storing, retrieval, and updating in a database.

**Source** - is where data originated, from the perspective of systems that consume this information, the source produces digital information.

**Heterogeneous**- Consists of different parts from each other (Richard Rennie, 2016).

**Generate**-is to produce or bring into existence to create vital information and distribute (Ludlow et al., 2017).

**Semantic** - In natural and artificial language, the study of meaning is formed, perceived, explained, blurred, and illustrated. This term may refer to subfields, including philosophy and computers, of many distinct disciplines.

**1.7. Chapters Outline**

This chapter outline is a useful summary of the content found in the mentioned chapters. It assists the research study to assist in organising the process and make it easy to comprehend and find the main points of the chapters:

**Chapter 1: Introduction**

The chapter commences by introducing the context and the background of this study. This accompanies by the definition of the terminologies used in this study. The study demonstrates how it was conducted.

**Chapter 2: A literature review**

In this field of research, this chapter discusses what has been written. The cited literature supports the theoretical claim being made and indicates that the authors understand the key ideas.

**Chapter 3: Theoretical Framework**

This chapter deals with theoretical frameworks and a theoretical framework for exploring big data knowledge of analytics.

**Chapter 4: Research Methodology**

This section provides an outline of the research methods that were followed in the study. It includes information on the company and the parameters used in the analysis. The researcher outlines the research design selected for this research and the reasons for this selection. It also explains the method used for data collection and provides the procedures that have been followed to carry out this study. The researcher discusses the techniques to examine the data. Ethical consideration discusses issues undertaken by the participant and the company.

**Chapter 5: Findings and Results**

The findings of the study and the results obtained from data analysis are discussed in this chapter. It then listed the challenges identified by selected data sets and followed by the results section based on the structure observation results. The findings of the research are focused on methodologies used to collect knowledge.

**Chapter 6: Conclusion and Recommendation**

The study concludes this chapter by offering a review of the analysis, stating the contribution, implications, research limitations, and providing potential research suggestions.

**Chapter 2 Literature review**

Interest in production companies is rising, and data research is continually emerging. These insights can make big data meaningful. Theories have been proposed to explain the motivation behind big data. However, studies believe that big data is beneficial to use in various sectors. The literature covers big data and analytics techniques, such as predictive analytics, predicting a company's future outcomes, and various big data analytics tools and strategies, allowing big data application to the production companies. This literature review lays out the state of existing research using big data, which seeks to enrich and expand how big data contributes to a conceptualisation of data. This research study is structured as follows; it summarises the literature on big data analytics and a summary of the current state of knowledge and highlights the gaps that exist in this study for research.

**2.1. Big data**

The growth of data has brought new changes to production management. Big data has been proven as a game-changer in the world. This has brought prevalent new opportunities for companies to gain insight from data. Big data is defined by data sets that grow exponentially. The data set's growth makes it difficult to manage and more challenging to gain insight (Ohlhorst, 2012). Gardomi and Haider (2015) described big data as the amount of data required to be processed. Although some practitioners argue that massive data sets are not frequently complex to overcome these challenges, organisations should still implement data governance. Tole (2013) asserts that using a viable solution for vast and multifaceted data is a challenge that companies need to continuously learn and then implementing new approaches to address. Dean (2014) emphasises the importance of inheriting big data. They believe in overcoming these challenges in the future. Inhering data is to carry on where the big data concept ended. The modern organisation should have intelligent analytics that allows company insight from big data (Gunther et al., 2017). Although many organisations are embracing analytics, only a handful have achieved this level. Therefore, to understand data analytics better, industries realise the importance of producing a massive amount of data to support strategies and make the right decision.

The term big data coined this massive amount of produced data and storage. Kitchin and Mc Ardle (2016) claim that volume and variety are not critical characteristics of big data; but velocity is crucial. Using traditional processing applications to process this data has become difficult. The data's size varies, from gigabytes to petabytes (Villars, Olofson, and Eastwood, 2011). A small amount of data may become big data, as specified by Feldman et al. (2020). However, because traditional process applications cannot accommodate much data, raw data becomes more critical when dealing with large amounts of big data, processing raw data, and generating insights. However, this requires the correct technology to handle it. Therefore, this data comes in numerous forms, with source information and requires the correct technique to handle it.

Big data is messy in the form of data generated. Data has moved from static to dynamic. Davenport and Dycle (2013) emphasised that large organisations regularly gather big data and use analytics to provide decision support. They further emphasised the technological requirements behind the processing of a large amount of data. For the digitization of data, the exploration of massive amounts of data and types has progressed exponentially. The sources include videos, texts, audios, and images that the organizations contribute (Eberendu, 2016). These data types are necessary to analyse daily operations to determine customer reactions, product personalisation, and consumers' product preference. Deb (2014) pointed out that the information is innovation but continues arguing that it has become a daily process. Therefore, with the continuous growth of massive amounts of big data, the importance of using massive data becomes more important. Regardless of the data, relevant data will become the key to industries that require to gain an advantage.

Boyd and Crawd (2012) claimed that big data is a technology that has been believed to maximize computation capability and algorithm precision to harvest, probe, and manipulate sizeable data sets to solve challenges. The epistemic assumption drives the pursuit of big data that expansive data sets offer superior forms of intelligence and erudition. This view has been challenged in the existing literature. Digital data are becoming computation-intensive and data-intensive, and manipulation requires significant logistical challenges (Borgman, 2015). Borgman argues that big data is not necessarily a new concept but observes the current distinction between big data and small data as somewhat analogous to the distinction made in the 1960s about big science and small science. While currently witnessing a ‘Big Data Movement’ (Parks, 2014), historically, there have existed much larger data sets than what some currently regard as big data. Myer-Schonberger and Cukier (2013) warned that underlying data believed to generate knowledge might be extensive but could be improperly used. However, compared with using small data, the consequences of misuse may be more significant. In contrast, big data has the potential for optimising and advancing the interest of big data.

Data sets generated through large amounts of historical data have a long history. However, there are new ethical requirements for data production and access (Bancroft et al., 2014). A key feature of big data is that researchers to modern data analytics do not intentionally engender it and aim to gain insights from existing data. The analysis typically occurs after data is collected and stored (Chandler, 2015). There has been an assumption that big data can afford easy access to large amounts of useful data. The digital data generated daily are potentially useful, particularly the big data advantage can be easily accessed (Boyd and Crawford, 2011). However, gaining access to big data requires identifying patterns within relevant data, and organisations typically own the data. Access may depend on the convergence interests of different data sources.

Access to data has become contentious. Despite the claims about big data and calls to make data accessible to users, production areas are data-poor fields where better data are hard-won and precious (Mills, 2019). Borgman (2015) argues that companies are typically rewarded for generating original data, and few would disagree that competitive companies usually gain an advantage by collecting data. However, the most competitive companies tackle solutions for the challenges of collecting data in new ways. There is a trend toward large repositories of open access to data such as indexes potentially benefiting from the development of metrics associated with these data repositories.

A large proportion of big data may still be proprietary data. Big data require vast amounts of attention to maintaining and organising metadata to use the data. Big data rendered in digital form are potentially more than cultural artifacts, for the rapid change of technologies and the software believed to store and analyse. In short, the source of the data is extracted and applied; further data will be for use.

**2.2. Data storage**

In the past ten years, the requirement to handle the data explosion (Turner, 2016) and the shift in hardware from vertical expansion to horizontal expansion led to a new explosion of big data storage brought about by the traditional database model. Big data storage technology refers to volume, velocity, or variety challenges, but it does not belong to the category of relational database systems. This does not mean that relational databases cannot address these challenges, but alternative storage technologies such as columnar stores and innovative combinations of different storage systems are more efficient (Marz and Warren, 2014). This results from the future technologies of data storage to gain insight concerning data storage. It became apparent that big data technology storage has become a commodity for a company. These storage technologies are the key driving force of advanced analytics and can change production society, and companies can make critical decisions. This assesses the current state of data storage technology, which can process large amounts of data and identify data storage trends. These approaches usually sacrifice data consistency to maintain rapid response when the volume of data continues to increase. However, several factors have driven the rise of big data. Companies can store and preserve more information than ever before. Although the company will receive and store the data, it must be analysed before it can be used. Storage solutions require to handle data volume, velocity, and variety of data.

Big data storage involves storing and managing data in a scalable way to meet applications that require accessing the data. However, an ideal big data storage system will allow the storage of an almost unlimited amount of data, flexibly and efficiently handle various data models, and support structured and unstructured data. Big data storage mainly supports storage with many data files and input or output operations on the storage. A typical big data storage infrastructure consists of a redundant and scalable supply of directly attached storage formats. The storage infrastructure is connected to computing server nodes that can quickly process and retrieve large amounts of data. Most large storage infrastructures support big data analytics solutions.

Big data storage is an infrastructure specifically designed to store, manage, and retrieve large amounts of big data. This storage can store and sort big data that the application can readily be accessed, used, and processed. There are ways to collect voluminous data in various formats. This process describes the three Vs.: the volume, variety, and velocity of data. Although, the storage encompasses a fourth V: Veracity. This is the most important element and most challenging issue to solve through data cleansing of databases. Although big data deficiency structure comes from various sources, it paves the way for big data analytics to emerge as a strategic analytic arm.

In summary, big data is solely the size of the data set. Considering this, the science behind big data is still more focused, intending to mine-specific data subsets from multiple large storage volumes. Companies use big data analytics to gain greater intelligence from metadata. Although there is no formal definition of specific volume size, big data storage refers to the volume that grows exponentially to a terabyte or petabyte size. Therefore, big data storage can collect large amounts of data generated at variable speeds. The following are different data storage files: data files and text files.

**2.2.1. Data file**

Data files store computer data used by computers to distribute files, includes input and output data. The file is optimised for large files, in which a single file is divided into multiple blocks and spread to improve access. However, it rarely contains the code and instructions to be executed. It defines an application but is used only for information used as input and written output by specific software applications. Therefore, data files handle large data sets displayed in different formats.

**2.2.1.1. Text file**

Text files are the most basic and human-readable files. It can be read or written in any programming language and is mostly separated the tabs or commas. When numeric values require to be stored as strings, the text file format takes up more space. The following data types are challenging to represent numbers, text, binary, date, and time. Aggarval (2012) emphasizes information obtained from data types in the form of text sources. The text file contains human-readable characters that are structured as a series of text lines. For instance, this file stores in the computer file system in a text file the text and serial number or part number. The modern text uses statistics to capture data and patterns in a human language for analysis and infer machines to understand the meaning of texts and perform various textual tasks. This can assist companies in using big data to discover ideas.

**2.2.2. Database**

Big data addresses data management and analysis issues in production. Traditional databases segregate historical data for analysis. These databases are mostly structured; however, big databases address data analytics over an integrated scale-out computer and data platform for unstructured data in near real-time. Big data technologies address decision supporting tools for searching large data volumes with several analytical methods. These tools would address several machine learning techniques and statistical modeling tools.

A database is organised data collected to store and be accessed electronically from a computer system. This refers to a set of data and how it is organised. To retrieve all data sets from computers, database management systems (DBMS) assist accesses the data. The DBMS provides various functions that allow entry, storage, and retrieval of large quantities of information and manage how it is organised.

The term database refers to collecting spreadsheets as size and requirements typically necessitate using a DBMS (Bestavros et al., 2012). The database is physically dedicated to computers that hold multiprocessors with a great memory, which has administrative specialised data information, such as computerised parts, production bills, handle fault data, and inventory systems. This involves stored data history for a part organised by serial number for easy retrieval. The monitoring part generates all types, such as assembly, test, and repair operation data.

**2.2.3. Sensor data**

Sensor data has become a rich source of information that is used to understand patterns' behaviour better. This data enables a company to create a prediction, allowing identification of trends and improving experiences. Intelligent companies enhance the advancement of sensor data technologies. The approaches are focusing on enhancing sensor performance for the smart meter. However, these approaches cannot address the ever-growing sensor data management and storage challenges. This is because the sensor data behavioural patterns demand more time (Fournier-Viger et al., 2013). Sensors are exponentially increasing the volume of data collected, and companies should have a platform to organise this data to understand and act on it. Sensor data signals perceive the state of the physical object movement records of the sensor devices required to analyse smart meters' functionality to optimise security and performance. This assists managers with decision-making and can be utilised to monitor and predict outcomes and facilitate better management decisions.

Sensors connected to production machines, robots, and production lines within a specified time frame can track variables such as temperature, vibration, and machine timing and then provide information on the analysis platform. Sensor data can prove what is happening on any machine. It can monitor inputs and outputs to decide when necessary. When the sensor data respond to certain physical environment inputs in response to the detected device's production, the managers can analyse the data, and ensure that maintenance of the the device takes place before the risk occurs. However, the output can guide the process. Therefore, effective sensor data use is critical for analysing event sequences' pattern (Elragal and Gendy, 2013).

The sensor devices used for object behaviour record the sequential arrangement of sensor data. According to Kim et al. (2011), various data analysis tools are necessary to analyse each subject's behaviour data movement. However, sensor data requires advanced analysis to understand behaviour patterns and send signals. These challenges are important issues in developing sequential patterns of sensor data analysis to identify new patterns. In addition, correctly modeled data is essential to describe the behaviour of event records. Sensors monitor changes in the machine, and for advances in sensor technology, large amounts of sensor data can be easily collected. Therefore, historical data about the patterns will be available. The predictive strategy uses a variety of techniques to predict the data to cope with such high-dimensional challenges. Among these techniques, data mining and time series methods are usually the most suitable for processing high-dimensional and unstructured data, as described by Kaisler et al. (2013) and Sicular (2013). Models are increasingly using these techniques to combine data from different sources to create insights. These technologies may provide better data analysis tools.

The sensor data components can be converted back to performance, assisting the company to find potential breakdown challenges. The data generated from these sensors can predict future results more accurately. Data has quickly become one asset companies can own. Data analysis technology has developed rapidly. These sensors can facilitate company usage analysis, and these analyses rely on the collected data.

**2.3. Data generated**

Big data is not a newcomer; it has existed long before the phrase, and more storage platforms will generate more data. A human, a machine, or a human-machine combined can generate big data. It covers organisations and machines. These are generated and stored in structured, unstructured, and semi-structured formats. This is broadly based on the sources of information. Therefore, data is generated from a piece of equipment by the organisation.

**2.3.1. Machine-generated**

This data is generated directly from machines without the intervention of humans. A Yale University professor emphasised machine data could generate this data through independent computation rather than human actions (Abouzied et al., 2010). Regardless of the differences of definitions, machine-generated data is unstructured and derived into a common structure. The utilisation of this machine data has been largely neglected. However, this machine data has great potential for obtaining valuable information. It contains important insight that can be found in machine data. This information generates from industrial equipment, network, and sensors installed in machinery logs that track data behaviour. Therefore, the company can merge these data sets to improve competitiveness and provide innovative new products. A machine equipped with sensors generates a large amount of sensor data. These companies characterise sensor records, such as volume, speed, and variety, as independent sources. The extraction of valuable information is important to optimise computer performance.

**2.3.2. Organisation-generated**

The organisation stores this data for current and future use and analysis of the past. Organisations can use this data for pattern recognition to estimate product demand. Big data has become necessary because organisations collect large amounts of data that contain useful information about the production process. When the amount of data recorded by these processes applies to produce products, organisation collect structured data, including metadata context—using these process records of the amount of data, such as manufacturing a product. However, data in any form of record is in the file structured data. Structured organisational data is a useful and valuable source of information. Organisations require to pay attention to information to hold the full advantage of the potential. The combination of structured, semi-structured, and unstructured data collected by the organisation can be collected to obtain information and use it in predictive modeling and advanced analytics applications.

The system for processing and storing big data has become a common component of the organisation's data management architecture. However, big data is not equal to any specific amount of data. Therefore, before using it for big data analytics, the organisation's data that is not clear, should be considered. Organisations should apply sufficient processing capacity to big data tasks to achieve the required velocity.

**2.4. Data analytics**

Big data can provide richer insights and discover hidden patterns by using accurate analytics. Minelli et al. (2013) pointed out that when they discover the meaning of more data, this can bring more possibilities to the company. Ten years ago, companies used fundamental big data analysis to discover hidden patterns and trends. At present, these theories have been challenged. The new advantages of modern data analytics are speed and efficiency—the ability to work faster and provide the company with an unprecedented competitive advantage. The rise of self-service analytics is not for analysts but also companies of all sizes (Tiwari et al., 2012). With the assistance of a new analytics technique, companies can leverage big data analytics to drive company objectives from operation to improve using big data (Henke et al., 2016). Big data analytics may transform and provide benefits, including optimising production (Bughin, 2016).

A new study published in 2015 showed that the focus has shifted from reporting to self-service analytics (Kouladouros et al., 2015). Traditional reporting platforms are not designed to manage the exponential growth of sources, data volumes, and data complexity. Modern platforms support the requirement for companies to gain greater accessibility and analytical insights from various data sources. Companies have changed the traditional model to meet the requirement of insight time in a competitive business environment. Modern platforms aim to democratise analytics through self-service capabilities. The platforms are characterised by agility, flexibility, and ease of use (Kouladouros et al., 2015).

Big data analytics is analysing data using statistical models and computational techniques. It combines traditional analysis techniques and mathematical models to obtain information. These can perform the same function to extract information—however, by applying science in analytics. Big data analytics refers to a set of processes and statistical models that extract information from data sets. Big data analysis uses statistical techniques such as time series analysis to take root in data, including quickly mining data, to obtain real-time data (Chen et al., 2012).

Despite the claims that big data analytics can add value to companies, the knowledge is still crucial for organisations, and the challenges are still limited (Wamba et al., 2017). Sharma et al. (2014) emphasized that although there is evidence that big data analytics can create company value, it requires more in-depth analysis. Recent research believes that obtaining value from big data analytics results from these technologies' centralized organisation into operations. Therefore, it requires developing company-wide big data analytics functions (Mikalef et al., 2019; Gupta & George, 2016). It uses data science, computer science, mathematical models, and statistical knowledge to explain phenomena in descriptive analytics and develops prediction-based techniques to predict future predictive analytics results.

The study was conducted on big data analytics to gain insights from the data, as stated by LaValle et al. (2010). Although reportedly a better way to analyse data was found, companies are not concerned with what has happened or why. These are commonly known as descriptive analytics. However, the main concern is to discover what is happening in the present and what may happen in the future, widely known as predictive analytics. Furthermore, companies afterward are concerned about what measures should be taken to find the optimal results, commonly known as prescriptive analytics. For effective analytics results, it is necessary to address big data aspects, such as heterogeneity (Fan et al., 2014).

Moreover, analytics techniques can process large amounts of data (Russom, 2011). Therefore, organisation analytics can be classified into these types of analytics, as explained. Predictive analytics is significant to the organisation.

**2.4.1. Predictive analytics**

Analytics has transformed over the years and are advancing rapidly. The company continues to develop, moving from focusing on historical questions to more forward-looking predictive analytics. However, recent advances in the technologies that underpin it, include machine learning, have made it more accessible. Predictive analytics is used to predict future outcomes. However, it is essential to note that it cannot predict future events. It predicts the probability of an event. Analytics solutions offer a convenient way to leverage production data, enabling companies to make faster and smarter predictions. Although it has been around for decades, predictive analytics is a technology whose time has finally come. Predictive analytics analyse current and historical data to provide insights into what may happen and has acceptable reliability. Abbott (2014) defines predictive analytics as discovering patterns using recognition techniques. It means these techniques can solve challenges (Tofail et al., 2014). The system has gone through several iterative stages and uses advanced information to find the challenge's outcome.

Predictive analytics can only forecast what may happen because predictive analytics are probabilistic. However, this extension of predictive analytics may perform with an acceptable level of reliability. Predictive analytics can assist companies with a range of challenges. Companies can use predictive analytics to analyse historical data and facts to identify potential risks and opportunities (Lebied, 2016). The essence of predictive analytics is to devise a modeling technique such that the existing data is understood to extrapolate the future occurrence or simply predict future data. Hence, predictive analytics include the validation of the model technique that provides accurate predictions. Predictive analytics relies on data mining and time-series algorithms such as testing the data. Therefore, developing a response plan for a sustainable future. Predictive analytics should be aligned with the correct data.

Predictive analytic solutions transform raw data to easy to understand and actionable insights. It can assist companies to leverage massive amounts of data and make real-time decisions that significantly impact production data. Predictive analytics solutions can identify challenges; days, weeks, or months before they occur, companies can become more proactive. Companies can use predictive asset analysis solutions to reduce the time spent looking for potential problems and spend more time taking actions to make the most of each asset. Companies that incorporate predictive analytics into company strategies can realise significant benefits by identifying faster and identifying emerging opportunities and correcting challenges and problems faster (Siegel, 2013). The prediction of future data relies on existing data. Therefore, any company that requires to increase productivity can immediately take corrective actions rather than merely reactive.

Prescriptive analytics is still at the newborn stage, and only a few companies have used its power. However, the advancements in predictive analytics pave the way for development. Predictive analytics in production can use historical data to make predictions (Pandya et al., 2018). Software tools represent the latest wave of technology to advance the field. Predictive analytics has the power to transform the production industry (Lechevalier et al., 2014). These analytics benefits encompass the quality and efficiency of data and the effectiveness of the organisation. However, the opportunity for predictive analytics is not only what can be predicted, but the company can be better with it. The historical data currently analysed can probably become a prediction, but companies can turn data into predictions (Gandomi and Haider, 2015). A massive amount of data is created and can be evaluated through predictive analytics tools to provide companies a better understanding of dynamics and trends. Predictive models exploit patterns that are found in the data sets for opportunities to be used. Solutions became increasingly important as companies realised the importance of using collected data to generate insights and detect patterns that can determine the course of action to reach company goals.

**2.5. Research Gap**

Studies have focused on storing big data. However, big data requires developing new analytics approaches to handle big data challenges, such as processing large amounts of data. Manufacturing managers can obtain insights into structured, semi-structured, and unstructured data from big data analytics (Zhong et al., 2015; Jeble et al., 2015). However, few studies attempt to manage large-scale production data's heterogeneity and increase insights into the production industry. Therefore, this research aims to resolve this gap and address the challenges in this research through the literature. This gap can also assist in promoting the future direction of the field.

This study explores big data innovation, which has recently attracted great interest and brought organisations' benefits for unprecedented opportunities (Elgendy and Elragal, 2014). At present, the industry continues to develop, and a large amount of high-velocity data is generated daily. These data layouts have inherent details and tacit knowledge patterns that should be extracted and utilized. Therefore, big data analytics can take advantage of organisational changes and enhance decision-making capabilities by applying advanced analytic techniques on big data and revealing hidden insights and valuable information. This review views large amounts of data to manage production data. This field is important because the data has increased. To better understand how to use big data for production management. Therefore, conducting more research in this field is important.

Finally, any new technology applied correctly can provide several potential benefits and innovations (Elgendy and Elragal, 2014). Big data is a remarkable field with a bright future when the approach is correct (Chen and Zhang, 2014). However, it requires proper storage, management, cleaning, handling, and analysis. The exponential growth of data has brought difficulties for the volume, velocity, variety, and source of additional data to be processed. Therefore, future research can provide a roadmap or framework for big data management, which can solve the previously mentioned difficulties.

Big data analytics is of great significance in the data era and can provide unforeseen insights and benefits in various fields. The correct application of big data analysis may exponentially impact the organisation, affecting the production industry. Big data techniques and technologies can also manage the exponential growth of production data and reduce database performance challenges by enhancing the ability to expand and capture the required data. Vidgen et al. (2017) pointed out that organisations face several challenges when gaining big data analytics insights. These challenges are related to how to orchestrate and use big data analytics. There is limited knowledge of how the context influences such capabilities critical in realising performance gains (Guenter et al., 2017).

**Chapter 3 Theoretical framework**

This chapter presents the concepts and theories that relate to big data. It determines the theoretical basis necessary for the emergence of big data in strategic decision-making through resource-based theoretical research. Furthermore, this section also conducts a comprehensive literature review to enrich knowledge about the fields mentioned above. The framework focuses on different applications of big data based on the development of organisational capabilities. This study's primary objective is to stimulate the research aim strategy and engage a wide variety of research contributions to support development between organisational analytics methodologies and strategy methods. The theoretical framework is one of the most important aspects of the research process (Osanloo and Grant, 2016). It serves as a guide for structuring and supporting research. This study introduces components discussed hereto in the literature.

Big data nowadays has emerged as a research topic. However, the term is seen as an umbrella term. The interest in the current big data and analytics has widely been believed to solve challenges without traditional methods. The importance of big data is rapidly growing as organisations gear up to leverage information to stay ahead of the competitors. This study has made three contributions to the discussion of big data in the scientific community. First, it elucidates a big data framework. Second, it addresses the ability to use big data through strategies to take advantage of increased data. Third, the evidence to produce positive outcomes through big data.

**3.1. Big data characteristics**

Big data has grown in prominence. The growth is used to describe the phenomenon in data volume, complexity, and disparity. The definition of big data is not consensual in literature, and there may be some confusion around what it means. Big data is not only an environment where accumulated data has reached large proportions. The word ‘big’ does not refer to size. This is only a capacity issue; the solution would be relatively simple. Instead, big data refers to an environment in which data sets have grown large to be processed, managed, stored, and retrieved within an acceptable time frame (Slack, 2012).

Big data is a large-scale structured, semi-structured, and unstructured data sets challenging to compute using traditional DBMS (Bhogal and Choks, 2015). An increasing number of organisations are producing large data sets, which commerces at a few terabytes. In the following section, the study discusses the characteristics and lifecycle of big data. Gartner and Hiebl (2018) proposed a 3V model for big data. The big data classifications of big data are described using a multi-V model in Figure 1. This is important for the reliability and accuracy of the data (Assuncao et al., 2014).

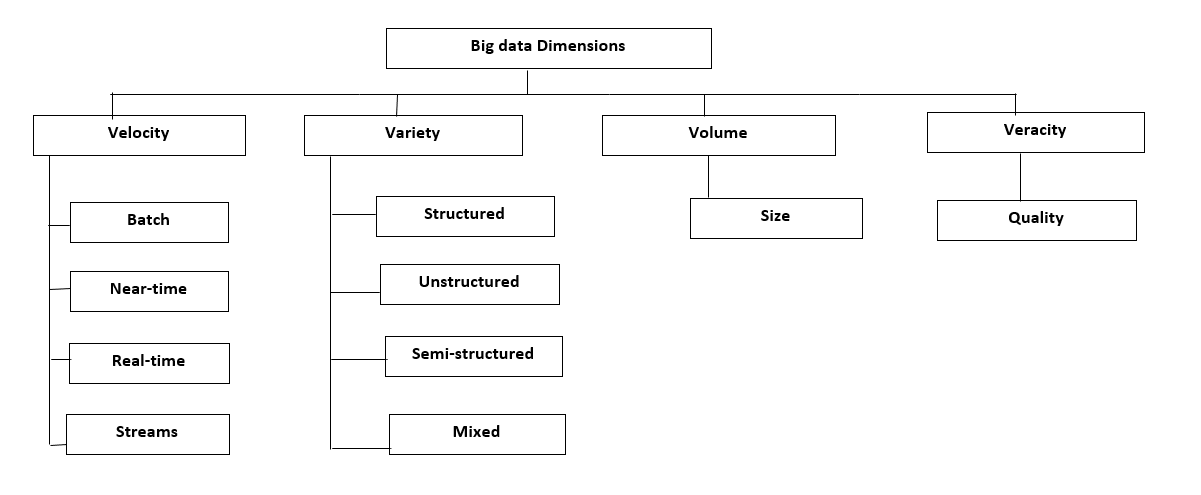


Figure 1: Classification of big data, adapted from (Assuncao, 2014)

Big data can be characterised by three fundamental factors: velocity, volume, and variety. According to Finkelstein (2015), structured information can be easily stored in relational databases of spreadsheets, with regular columns and rows. However, unstructured information, such as images, sensors, and videos, makes up about 85% of data generated today and presents challenges in deriving meaning with conventional intelligence tools. Information-producing devices, such as sensors, continue to multiply. This information sharing option represents a fundamental shift in the way organisations interact with each other. The big data characteristics shape how production organisations manage, store, ingest, analyse, and distribute data across the sector. Table 1 illustrates the big data characteristics and highlights the difference between big data from the historical perspective of standard data.

Table 1: Characteristics of data (Finkelstein, 2015)

|  |  |  |  |
| --- | --- | --- | --- |
| **Characteristic** | **Description** | **Attribute** | **Driver** |
| Volume | The sheer amount of data produced, or data intensity should be consumed, analysed, and decisions taken based on complete data analysis. | A high volume of data is being produced by the digital universe, which is expected to increase exponentially. | Increase in data sources, higher resolution sensors. |
| Velocity | How quickly data is generated, changed, and how quickly data should be obtained, interpreted, and processed. | Metrics used can be defined in the usability, applicability, and time value segments. | Increased data of sources, connectivity throughput, and improved data-generating computer processing ability. |
| Variety | Integration, management, governance, and architectural pressures in IT are generated by the increase of information coming from new sources both within and outside the company or organizational walls. | The data can be broken down into the following parts: structured, unstructured, semi-structured, and complexity. | Audio, images, videos, and conventional data sensors. |
| Veracity | The accuracy and provenance of the data obtained. | For data inconsistency, incompleteness, uncertainty, latency, deception, and model, big data's quality may be better, harmful, or undefined. |  |

**3.1.1. Classification of big data**

Big data are classified into different categories to understand characteristics better. Table 1 illustrates the numerous characteristics of big data. The classification is important because of the large-scale storage format. Classification is based on velocity, veracity, volume, and variety (Kitchin, 2017). Each of these categories has characteristics and complexities, as described in Figure 1.

According to these theories, big data states a large volume of various multi-source and real-time data created during product design for the product life cycle. The organisation, big or small, manages a considerable amount of data generated through various data points and production processes. Organisations can handle the data using excel worksheets, access databases, or other similar tools. However, when data cannot fit into such tools, and human errors occur to exceed acceptable limits due to intensive manual processing, big data and analytics come in to assist the process.

**3.1.1.1. Volume**

This volume presents the most immediate challenge to conventional structures (Anuradha, 2015). Companies have a large amount of archived data in logs but cannot process the data. The benefit gained from processing large amounts of information is the main attraction of big data analytics. Having more data leads to better models: simple math bits can be unreasonably useful, given large amounts of data (Mayer-Schönberger and Cukier, 2013). Hence, the volumes of data are greater than those conventional relational database infrastructures can cope with, processing options breakdown broadly into the choice between massively parallel processing architectures.

This data is about the volume that can reach unprecedented heights. As a result, it is not uncommon for large companies to have terabytes or even petabytes of data in storage devices and servers. These data assist in shaping the future of a company and the actions. The size of the data determines the number of categories in the data set.

Data volume refers to the data sets' size required to be analysed and processed, which is frequently larger than terabytes and petabytes. The sheer volume of the data requires distinct and different processing technologies than traditional storage and processing capabilities.

**3.1.1.2. Velocity**

Velocity refers to a measure of how fast the data in a large data set is generated. Currently, most organisations can easily fixate on the rate at which data is pouring in. However, velocity underscores the requirement to process the data quickly and, most importantly, use it at a faster rate than ever before. These velocity-related challenges are typically viewed as technical ones, but there is more to it than technology. Process and culture can hold a company back from a speed and agility perspective-no matter how fast data is collected and processed (Yang, 2012). Velocity can be more important than volume because it can bring more significant competitive advantages. In most cases, it is better to obtain limited data in real-time than to obtain large amounts of data at low speed.

The importance of velocity is the increasing rate of data flowing into the organisation. Velocity is a 3Vs framework component that is believed to define the speed of an increase in data volume and relative accessibility. Velocity assists organisations understand the relative growth of big data and how quickly that data reaches sourcing, applications, and systems. Velocity is a combined data infrastructure and architecture to manage data and deliver it to recipients as quickly as possible. Velocity can be categorised as:

**3.1.1.2.1. Batch**

Batch processing is the number of data arriving as the set group within a certain interval of time. Big data applications process data in batches and have batch processing velocity. The batch processing layer assumes that the input data process over time. This has the advantage of ensuring accuracy. The processing changes can be applied to the data each time, but it may impose an enormous batch processing burden on the system. Batch data processing is computer processing to complete a batch of large volumes of data at once. Data can easily contain millions of records a day and stored (Watson, 2014). This is an effective method for processing large amounts of data collected for some time. Batch processing is most used when processing large amounts of data, and the data source is an old system. Batch data is an effective way to process large amounts of data collected, input, processed, and produce batch results. Batch processing requires separate programs for input, processing, and output.

The data generated on the mainframe is suitable for batch processing. Accessing mainframe data and integrating it into modern analytical environments takes time, making streaming infeasible in most cases. Batch processing is effective when real-time analytics results are required and processing large amounts of information is more important than obtaining fast analysis results. However, the data flow may involve big data to process large amounts of data.

**3.1.1.2.2. Near real-time**

The time between data arrival and processing is short, close to real-time. Near real-time refers to data processing and communication that quickly respond to the event after it occurs. As the name implies, this form of data processing is fast but not instantaneous, with results ranging from a few seconds to a few minutes. With the advent of the big data era, more and more organisations are participating in the data processing.

**3.1.1.2.3. Real-time**

The data arrives and is processed continuously, allowing real-time analysis. The growth of data and the resulting importance has changed the way data is viewed. Velocity is essentially a measure of the speed of data input. Some data will be entered in real-time, while other data may be launched in time, sent in batches, and since the platform will not experience the entered data at the same speed, it is important not to generalise.

**3.1.1.2.4. Streaming**

Real-time data arrives and processes according to the incoming data stream. Stream processing is the key to real-time analytics of results by constructing data streams, feeding them to analysis tools when generated, and obtaining near-instant analytics results using these platforms. As mentioned earlier, the data source's nature plays an important role in defining whether the data is suitable for stream processing (Hashem et al., 2015).

**3.1.1.3. Veracity**

The veracity refers to how accurate the data which may be equivalent to the quality. It generates a deeper understanding of data and how to contextualise it to act. More specifically, when it comes to the accuracy of big data, it is the quality of the data itself and the credibility of the data source, type, and processing. Deviations such as bias, duplication, inconsistencies, or abnormalities, should be eliminated to improve big data accuracy. In several big data applications, controlling data quality and accuracy has proven to be a big challenge. The quality of captured data can vary widely. The accuracy of the analysed data depends on the accuracy of the source data.

In this modern digital world of data, the proper use of data can play a critical role in a company's success. Data sets are exploding at an ever-accelerating rate, to collect and analyse data to maximum effect is crucial. The big data to be analysed will be reliable and accurate. The data quality depends on the size of the data set (Wang et al., 2018).

**3.1.1.2.1. Quality**

Veracity refers to the quality of the data to be analysed. High accuracy data has records worth analysing and contributes to the overall result in a meaningful way (Grover et al., 2018). At the same time, low accuracy data contains a high percentage of meaningless data. The non-valuable in these data sets referred to as noise.

**3.1.1.3. Variety**

The subsequent aspect of big data is the variety, not structured data. The relational databases are challenging to put big data. This means that the category of big data belongs to an essential fact that requires to be known by the data analysts handling a variety of structured and unstructured data. This gr increases the complexity of storing and analysing big data. Almost 90% of the data generated is data in unstructured form. Variety is one of the most important characteristics of big data. Sources of big data generate different forms of data. Consequently, as big data companies grow, designing algorithms for big data mining and analysis becomes more challenging.

Data was once collected from one place and delivered in one format. Once adopted in the form of database files such as Excel, CSV, and Microsoft Access, presented in traditional forms, such as video, text, pdf, and graphics. Although this data is useful, it generates more work and requires more analytical skills to decipher this incoming data, making it easier to manage and work. Big data is more than large amounts of data: the way of providing opportunities to take advantage of new and existing data and to discover new ways to capture data to make a difference for company operators. Big data can be categorised into the following structured data, unstructured, semi-structured, and mixed data (Bathla et al., 2018).

**3.1.1.3.1. Structured data**

Structured data generally refers to data whose length and format are defined for big data. The development of technology usually provides structured data sources in real-time. Data sources are divided into two categories: The generated Computer or machine refers to the machine's data without human intervention. Human-generated refer to data that humans interact with computers. It has been organised into a formatted repository that is typically a database. It involves all data that can be stored in a database with rows and columns of tables. These have relational keys and can be easily mapped to pre-designed fields. Data is recently most processed in the development of technology and the simplest way to manage information. The structure suggests this is organised and neat data. Structured data is organised in tabular formats such as rows and columns, and there is a relationship between different rows and columns (Kandel, 2011). Therefore, it is organised and formatted, easy to store, process, and access. It can easily cooperate with most standard analytics models.

Structured data uses predefined and expected format. This may come from different sources, but the common factor is that the fields are fixed, as the way that is stored, hence it is structured. This predetermined data model enables easy entry, querying, and analysis. Structured data refers to data that enters a relational database with a row and column-oriented structure and exists in operations or algorithms. These data are easy to enter, save, find, and analyse. However, it should be well-defined regarding field name and character types such as alpha, numeric date, and currency. Structured data is restricted in usage because of inflexibility. Although unstructured data may have a native, internal structure, it will not be structured in a predefined way.

In this data, each record will have a time stamp. Hence, each field has a defined purpose. It makes it easy to manually query and easy for data mining algorithms to identify patterns and, in several cases, identify anomalies outside of these patterns. Those two fields are predefined, and data mining algorithms could easily identify patterns and anomalies with a few minutes’ worth of records. Text files are displayed entitled columns and rows, which can easily be ordered and processed by data mining tools. This could be visualised as a perfectly organised filing cabinet where everything is identified, labeled, and easy to access.

Despite the vast difference in technical complexity between these examples, it is illustrated that structured data drills down to use established and expected elements. Timestamps will arrive in a defined format; it cannot transmit a timestamp described in words outside of the structure. A predefined format allows for easy scalability and processing, even handled on a manual level. Structured data can be used for long source defines the structure.

**3.1.1.3.2. Semi-structured data**

Semi-structured data does not reside in a relational database, but that has some organisational properties that make it easier to analyse (Cumbley and Church, 2013). The semi-structured data is designed to evolve the relational data that allows data representation with a flexible structure. The term semi-structured data is a form of structured data that does not conform with the formal structure of data models associated with relational databases or other forms of data tables. Nonetheless, it contains tags or other markers to separate semantic elements and enforce hierarchies of records and fields within the data (Sukanya and Biruntha, 2012). Semi-structured data is increasingly occurring since full-text documents and databases are not the only forms of data on the internet. Different applications require a medium for exchanging information.

**3.1.1.3.2. Unstructured data**

Unstructured data is information that is not stored in a specific format. This data is large data sets that are difficult to analyse with traditional tools (Zhang and Huang, 2013). It is not a better fit for a mainstream relational database. Unstructured data typically consists of files for word processing documents, spreadsheets, and PDFs. It can be types of data such as numeric, text, binary, dates, and time. Irregularities and disorganisation within unstructured data make it challenging to handle and understand. Ideally, this information would be converted into structured data. However, this would be costly and time-consuming. Das and Kumar (2013) argued that unstructured data would account for 90% of data in the coming decade. They stated that analysing this massive amount of data would expose new improvements that were impossible previously. One argument is that although some form of structure is not formally identified, it can still be implied and should not be labeled “unstructured.” The counterpoint states that data has some form of structure but is not helpful to the processing task at hand; it may still be characterised as unstructured. Furthermore, not all types of unstructured data can easily be converted into a structured model. Some authors might believe that unstructured data replaces structured data. Companies have not fully exploited the potential of structured data. Better predictive analytical capabilities continue to be used. Unstructured data provides access to new insight not otherwise available in structured data.

It has some organisational framework but does not have the complete structure required to fit in a relational database. Semi-structured data has a self-describing structure that contains tags or attributes to separate various entities within data. Structured data comes with a definition. Unstructured data is the opposite of that. Rather than predefined fields in a purposeful format, unstructured data can come in all shapes and sizes. Though typically text, unstructured data can come in forms to be stored objects: images, audio, video, document files, and other file formats. The common point with all types of unstructured data comes back to the idea of deficiency definition. Unstructured data is more commonly available, and fields may not have the same character or space limits structured data. Given the wide range of formats comprising unstructured data, unsurprisingly, this type typically accounts for about 80% of organisation data. Unstructured data is not suitable for spreadsheets or data storage.

**3.1.1.3.4. Mixed data**

Mixed data are structured, and unstructured textual data generated from various sources. Some data generated from these sources are structured, while the other is unstructured. Analysis and storage of structured data have been ongoing for a long time; unstructured data has appeared recently on a massive scale. Databases for storing unstructured data and analysis techniques to obtain results have been recently developed. This category constitutes both structured and unstructured data (Jacobson et al., 2014). To obtain reliable results in big data analytics may be combined by converting unstructured data and analyses.

Structured and unstructured data are best understood when data are in a structured data form. Structured data is data that is represented by numbers, tables, rows, columns, and attributes. Dave and Gianey (2016) argue that it is time to stop criticizing unstructured data. Text is easily one of the most structured data types. Few authors would argue with this; therefore, it should be no surprise that the digital data produced in this new world does not conform to traditional tabulated data categories easily assimilated by humans and processed by traditional computational methods. Perhaps it would be more accurate to say that nobody knows how much data has been produced but agree that there is an awful lot, and the amount is growing at an increasing rate.

Data with high volume, high velocity, and wide variety should be processed with advanced analytics and algorithms to reveal meaningful information. The knowledge domain that handles the storage, processing and analysis of these data sets is marked with big data due to these characteristics.

**3.1.2. Analytics of big data**

The growth of big data continues to be rapid, with organisations involved in data managing and analysis. Big data analytics are used to facilitate quicker and better decision-making by companies seeking to benefit from big data because it is challenging to analyse data sets using methods and infrastructure focused on traditional data management (Constantiou et al., 2015). Therefore the demand for new resources and methods dedicated to big data analytics, is growing. The rise of big data affects the data itself to process and collect and process to extract final decisions. Providing big data resources and technologies can assist control the growth of exponential network-produced data and increase the organisation’s ability to scale and collect the necessary data to minimise challenges with database efficiency (Elgendy et al., 2014). Table 2 clarifies more concepts of big data analytics.

Opening every science today, whether online or in the real world, include data science, analytics, big data, or any variation of these terms (Agarwal and Dhar, 2014). Some researchers concentrated on the concept of big data meanings (Akter et al., 2016; Mikalef et al., 2018), while others examine the methods, techniques, and procedures required for analysis (Russom, 2011) and others aim to clarify the effect on a company value of big data analytics (Mikalef et al., 2018).

Table 2: Big data analytics concepts, adapted from (Mikalef et al., 2018)

|  |  |
| --- | --- |
| **Authors and date** | **Definition** |
| Loebbecke and Picot (2015) | Big data analytics: a means to analyse and interpret any digital information. These are essential for the development of advanced artificial intelligence and cognitive computing capabilities. |
| Kwon et al. (2014) | Big data analytics: The technologies (data mining and database) and techniques (Analytical methods) that a company can use to conduct large-scale analysis. |
| Ghasemaghaei et al. (2015) | Big data analytics refers to tools and processes usually applied to large and dispersed data sets to obtain meaningful insights to improve organisational performance. |
| Lamba and Dubey (2015) | Big data analytics has defined the application of multiple analytic methods, which can diversify big data to provide actionable prediction results. |
| Muller et al. (2016) | Big data analytics: statistical modeling of large, diverse, and dynamic data sets. |

Companies may attempt to gather data to consider the nature and significance of decision-making. The amount of information to be evaluated is large and includes different forms. Massive high-dimensional, heterogeneous, complicated, unstructured, erroneous, incomplete, and noisy (Ma et al., 2014) are aspects of big data and involve improvements in approaches to statistics and data analysis.

Applying algorithms to analyse big data's content is an essential part of the data analytics used for 1) analysing data information and relationship sets, 2) extracting previously unknown valid patterns, and 3) assuming that meaningful relationships between stored variables are identified. This section covers different types of big data analytics, commercing with the available data analytics techniques and some common big data analytics suites, and finally covering some frameworks and resources for big data. Techniques of data analysis can be characterized into three groups, as illustrated in Figure 2:



Figure 2: Data analytic technique

**3.1.2.1.** **Supervised**

Supervised methodology refers to where the training data can be trained, checked, and marked. The mark represents the full history of the known data, and the history of the data variables is known. Supervised learning includes training the system based on labeled data, requiring the supervisor to expect the output from each input that can be trained according to the system's expectations. Predictions can be made in "Classification and fault detection applications and channel encoding and decoding" when the device is trained (Sharpe et al., 2019; Cui et al., 2019).

The function between the input and output is assumed to be approximated by this technique. The principle is to allow the system to learn the training data set classifier and then apply this classification to the unlabelled documents of the unknown data sets. Therefore, learning from examples is included in this learning technology (Boyd-Graber et al., 2014; Müller et al., 2016; Breed and Verster, 2019).

**3.1.2.2. Un-supervised**

The data for training is unlabelled. Unlabelled data means the data's history is missing; for data variables, no history is available, and the data has not been trained and tested. Separate training data is required for unsupervised techniques (Boyd-Graber et al., 2014; Müller et al., 2016; Breed and Verster, 2019).

Unsupervised learning involves deducting functions from unlabelled data to present unknown structures. This technique does not require a supervisor, which ensures that testing based on unlabelled data feedback can be performed independently of the system (Cui et al., 2019).

**3.1.2.3. Semi-supervised**

Some of the information is labeled, and some of the information is unlabelled, combining supervised and unsupervised processes. For both labeled and unlabelled data, this algorithm is suitable. Even if the data is incomplete or deficiency a training set, classifiers for such data sets may also be learned. Both supervised and unsupervised approaches concentrate on one aspect of the target separation or independent variable distribution, respectively. Using them together can produce better outcomes (Breed. and Verster, 2019).

LaValle et al. (2011) studied big data analytics’ potential and described the ability to use big data in decision-making. Similarly, Wixom et al. (2013) studied big data analytics to generate value for the business. It recognized the importance of strategy and data management by conceptualizing the dimensions of big data analytics. The research demonstrates that by growing decision-making speed and allowing big data to spread more broadly around the organization, establishing big data analytics will optimize business value.

Chen et al. (2012) demonstrated that related technologies and business analytics could better understand the company, and LaValle et al. (2011) demonstrated that the top-performing organisations could decide based on rigorous analysis, while the lower-performing organizations decide more than twice quickly (Sharma et al., 2014). Similarly, "the ability to use data management, infrastructure (technology) and talent (personnel) capabilities to provide business" according to Kiron et al. (2014).

A study by Akter et al. (2016) developed a strategy based on previous studies that showed the significance of management and technology in the world of big data. This thesis proposes and studies the effect on a systematic methodology. Elgendy (2013) further suggested big data, analytics, and decision-making system in which big data analytics techniques are integrated into the decision-making process.

The decision-making process and how to incorporate big data analytics into the decision-making process was explained by Elgendy and Elragal (2016). It can be believed that design science methodology maps big data resources and analytics to different decision-making processes. As a result, it is possible to determine the added value gained by incorporating big data analytics into the decision-making process (Elgendy and Elragal, 2014; Elgendy and Elragal, 2016).

Despite such obstacles, at every processing point and implementing big data, advanced technologies and tools assist decision-making. In decision-making and forecasting sectors such as healthcare, retail, tourism, marketing, the financial sector, and transport, big data plays an important role (Elgendy and Elragal, 2014).

The use of big data requires support for decisions. Decision-makers, however, must decide the value required and concentrate on identifying the appropriate methods, strategies, and resources to select the necessary ones. An ideal decision: this method also relies on decision-makers' wise and rational decisions (Wang et al., 2016).

At each point of the big data process, decision-making takes place, including data collection, data cleaning, data processing, data visualisation, and prediction. However, it is often difficult to incorporate an effective solution for each method, and strategies can be used in big data work for decision-making. Some decisions include disciplinary feedback, including data mining, analytics, machine learning, visualization, and social network analysis. There are three classifications of specific big data technologies: batch processing, stream processing, and hybrid processing tools (Wang et al., 2016). The relationship between decision science and big data is illustrated in Figure 3.

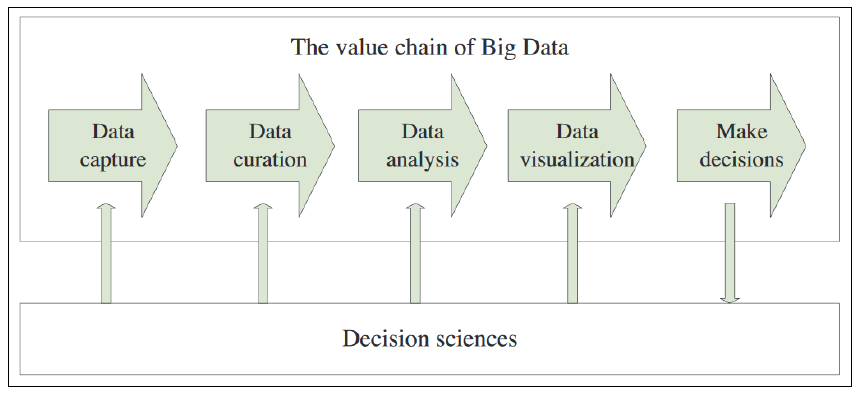


Figure 3: The relationship between and big data and decision sciences, adapted from (Wang et al., 2016)

**3.2. Big data strategies**

The increased technical infrastructure is taking advantage of more and more businesses; the industry is rapidly expanding. Businesses continue to evolve and find new ways of using the ever-increasing amount of data. Management continues to grasp the primary role of using big data in digitising organisational technology and how big data initiatives can influence company decisions. Big data is expected to provide organisational science with new theories and strategies in recent studies and can play a central role in creating new management strategy methods. This study adds vitality to this promising literature through an integrated organizational theory and stimulates a broader debate about management.

Big data and analytics have risen to the top of the agenda for the manufacturing industry. This changed the way companies conduct business and achieved very substantial changes in efficiency seen as enterprises revamped the core processes in the 1990s. Data-driven methods take root; in competitive differentiation, this becomes an increasingly significant point. Among thousands of businesses in six data-rich industries to make the most of the data and analysis, this exploration requires three mutually supportive functions. First, multiple data sources should be able to define, integrate, and manage the organisation. Second, to predict and maximize performance, businesses require to be able to develop advanced analytics models. The third and most important argument is that the management department should have the power to alter the organisation of data, and the model should make better decisions. There are two essential aspects of these capabilities: a consistent plan to use data and analytics to compete and incorporate the necessary technological architecture and functions. Equally essential, an integrated approach to data sources, model building, and organisational change should be shaped by a clear understanding of the required business effect. This will assist in preventing the common trap of commencing by requesting what details the organisation can use. To assist the mission, leaders should spend adequate time and energy in aligning managers in the company.

**3.2.1. Select correct data**

The data field has undergone significant changes in the past few years. The amount of knowledge is increasing, and there is an acceleration in extending insights from emerging data. Better and large data allows businesses to have a comprehensive view. From the viewpoint of previously invisible skills, behaviours, and strategies that can be enhanced, the skill emerges. This implies that source data and required assistance are the two fields that companies can concentrate on them.

**3.2.1.1. Source data**

Companies have data and requirements to handle challenges, but managers do not have important decisions. By specifying business challenges and opportunities that require to be solved, businesses may promote a more detailed data view. The ability of external and new data sources requires managers to innovate. In the form of images and videos, output data can contain terabytes of unstructured data. In addition, data streams from sensors, processes monitor, and external sources are available.

**3.2.1.2. Necessary support**

Structures can hamper new types of data sources, storage, and analysis. Current architectures may prevent the incorporation of isolated information, and the management of unstructured data remains outside conventional capabilities. It typically takes years to solve these challenges fully. Organisation leaders can satisfy short-term requires for big data by priority criteria. This ensures that it is possible to recognize and connect the most relevant data for analysis easily. To synchronize and reconcile overlapping data and address missing information to clean-up operations can then be performed.

**3.2.2. Develop models to predict outcomes**

Data is important, but performance analysis and competitive advantage come from analytical models that allow management to forecast and maximize outcomes. More importantly, it is typically not from the data but from finding company opportunities and deciding how the model can boost efficiency that the most productive way to use a model. It can be noticed that in the actual data relationships that managers understand more generally, this quite substantial modeling can produce faster results and root models. Advanced statistical approaches would undoubtedly yield better models.

**3.2.3. Transform capabilities of the company**

Significant changes in organisational capabilities are essential for transforming companies to find new ways to compete. However, managers who do not understand or trust big data-based technologies may cause concern. Such challenges are usually caused by a mismatch between the organisation's current culture and skills and the evolving strategies for the effective use of analytics. The new method is inconsistent with decision-making methods, or it does not provide an exact blueprint for business objectives to be accomplished. Big data requires thoughtful systemic improvements and three fields of operation to be used. Business-relevant analytics that can be used as initial big data and analytics implementations fail because they are not in line with day-to-day processes and decision-making principles. Conversations with frontline managers can ensure that a set of trade-offs can be handled through analytics and technologies that support current decision-making processes.

To use emerging technology and algorithms every day, managers require apparent methods. Terabytes of data and complicated modeling are required to improve the ability to use big data. Most organisations would require developing their analytical skills and literacy, even with evident and functional models. Managers should consider the key to overcoming problems and finding possibilities for research to become daily operations. A multifaceted strategy, including preparation, role models for leaders, and rewards and indicators to encourage behaviour, is typically required to change culture and attitude.

This research indicates how to use big data and analytics; management should act immediately. Instead of making large-scale changes to collect data, use models, and change the organizational culture, managers can nevertheless conduct focused work. To retain versatility of initiative changes organisational culture. This is important because it will continue to evolve and change the data itself and the technologies used for management and analysis, generating new opportunities. More companies are mastering the critical skills of using big data, making it a decisive strategic advantage to support superior functions.

**3.3. Big data positive outcome**

The big data concept has been around for years; most companies recognize that analytics can be implemented to collect the data that streams into database platforms and gain significant insight. Nevertheless, important numbers in a spreadsheet were manually analysed to discover observations and patterns from streams into the database platform back in the 1950s, decades before anyone uttered big data, using these basic analytics. However, the benefits of big data analytics provide to the table are speed and efficiency. A few years ago, a business would have collected information that could be used for potential choices, run analytics, and discovered information. For urgent decisions, the firm may recognise insights. The ability to operate faster and remain agile provides a strategic edge to companies they did not have before. Big data analytics allows companies to exploit knowledge and use it to discover new possibilities. This leads to smarter business deals, more effective practices, higher income, and happy customers. Companies understand the way big data can be used. Studies can be used for the below ways:

**3.3.1. Reducing costs**

In storing vast volumes of data, big data applications such as predictive analytics have cost advantages. More efficient ways of doing business can be found.

**3.3.2. Quick decision making**

Companies can quickly interpret information and evaluate based on what has been observed with the speed of predictive and memory analysis and the ability to analyse new data sources.

Businesses has changed the way big data in every sector handle, interpret, and use data. One of the most promising fields where modifications can be implemented is the manufacturing department. Analysis of production will reduce the costs of forecasting future results, eliminate risks, and improve big data efficiency. Data can enable companies to base decisions on valid evidence and facts without excessive guesswork—the benefits of using data to drive decision-making become obvious. Using big data can assist managers in generating data and using it for decision-making more effectively.

Using data is important within the context of big data. A theoretical framework for using big data analysis is proposed in this chapter, which uses intelligent computing algorithms to predict future outcomes. The purpose of this method extracts the patterns and uncover valuable knowledge from the data collection. Data mining aims to discover hidden characteristics and patterns in large and complex data sets.

**Chapter 4 Research Methodology**

The research methods adopted to achieve the research objectives are outlined in this chapter. This refers to the literature review, theoretical framework, and research background. However, the literature review has become a suitable research method. Research methods are inconsistent with research questions and research questions. The researchers elaborated on using predictive analysis techniques (Lewis, 2015; Ribeiro et al., 2016) and the necessity of this analytics technique to solve research problems and objectives. Predictive analytics became an ideal method to provide answers to research. The researcher took the proper steps and actions to ensure the review's accuracy, accuracy, and reliability. Rajasekar et al. (2012) defined research as the systematic and logical growth of useful information to find scientific challenges. In this study, with the support of the research theme, various steps have been taken to solve the research problem.

The essential elements and specific methods were applied to this study. There were two methods dedicated to research, namely, research design and research method. In the research design process, the researcher strived to select the overall strategy for various research components, ensuring that the research problem is addressed in this study. The blueprint for data collection, measurement, and analysis has been developed. For the analysis, the collection of the data was considered important. This research began with analysing the data then finds patterns and trends. Hence, in this research methods paradigm, the driven factors were direct involvement and transits from experiences set of propositions to the company (Aiken et al., 2014).

**4.1. Research design**

Research designs refer to the framework's overall strategy of research methods and techniques chosen by the researcher. These methods defined the logical plan to tackle established research questions. The researcher was directly involved in participating in the company to understand big data analytics in production. This has constituted an outline for data measurement and analysis. This approach was based on the phenomenological method of gathering data using predictive analytic.

Company A is a private production company that employs about 5,000 employees. The firm has an area of two million square meters. It produces steel for various companies worldwide and has been selected for this research study because of the company reputation as innovative technology. The study was conducted on the company premises, with production management. The study's period was two months, from July to the end of August 2020. The research design was described as a master plan by Swaffield et al. (2011), demonstrating the techniques for conducting the research study. However, research design served as a method and procedure to collect and analyse the data required. Therefore, the method chosen was an intrinsic case study. The research designs illustrated in the research are shown in Figure 2.

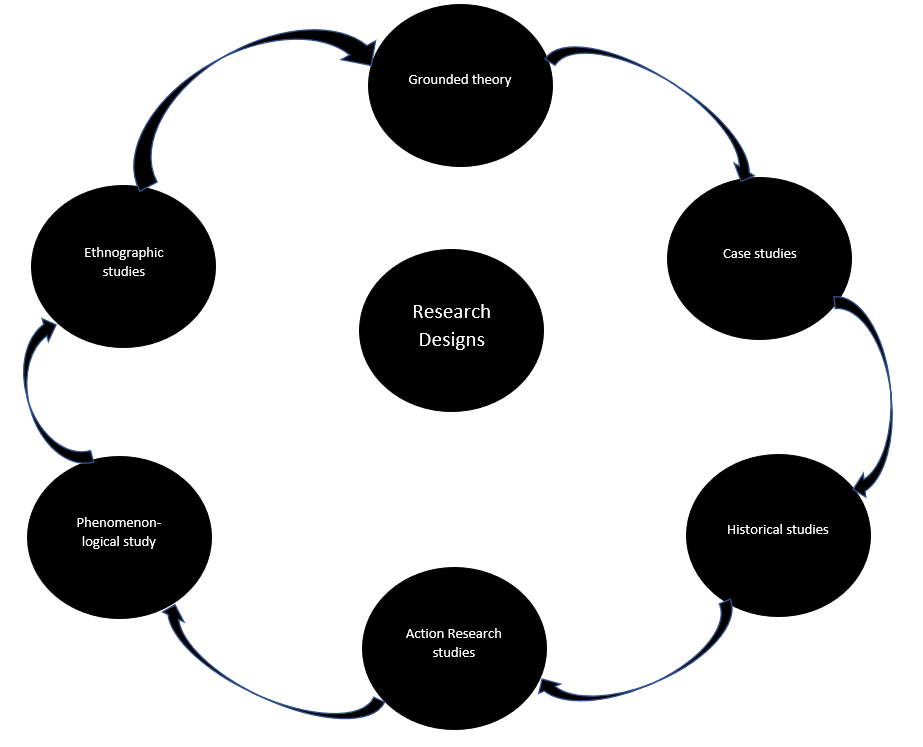


Figure 4: Types of research designs sources: (Ridder, 2017)

**4.2. Research methods**

For this method, a single case with an embedded approach was chosen for this study to be the most suitable method to gain insight into a single organisation's company. This approach illustrated a powerful collection of procedures. It has generated a large amount of typically qualitative data and provided considerable insight into the phenomenon (Easton, 2010). This constructive approach has placed the researcher being examined with the phenomenon approach's interest (Andrew et al., 2012). This phenomenon has described the situation that is observed to exist or happen. Yin (2013) accepted that a case study is a suitable approach to research explaining the aligned and single context and phenomena. The case study was conducted to clarify the usage of the case organisation of big data management. It was not done to represent any other manufacturing companies using these technologies (Stake, 2013).

The growth of a single phenomenon over a period is detailed in this case study. Using both qualitative and quantitative data on these case studies are commonly assumed to explore complex challenges. There are holistic, hierarchical, context-sensitive, and layered ones. It began with identifying the research problem, followed by selecting the case and data collection approach and analysis. Involved in the gathering and analysing of data in the field. Although this case study has focused on a particular phenomenon, it has challenged the use of data to form generalisations.

This study used qualitative research as a phenomenological approach to gathering appropriate data. This approach was concerned with understanding the challenges that arise (Cumming et al., 2015). The qualitative research study design primarily focused on the data derived from an active participant. The qualitative approach was appropriate for this study because it has used participant experiences in practice. This method is used to collect data from these diverse data sets possible to generate an ideal insight. This case study method was for the capacity it has in creating new knowledge. This case study has used both the current and historical data from the company data sets. The other methods could not capture this data (Easton, 2010; Yin, 2013).

**4.2.1. Research Approach**

Big data techniques have been developed and implemented to analyse a large volume of data generated by company A. The company requires quicker insight into the growing amount of data. Real-time data analysis allowed the organisation to view the past and predict the future. These were analytics streaming and knowledge of what happened and what could happen. While this was progressive, it was more challenging to implement and use these three analytics to benefit the company. In addition, the study sought to draw attention to the immense benefits brought to the company by big data.

**4.2.2. Research study approach**

The empirical method of this research was direct participation in nature, which was for studying the latest theories. The reviews included the literature in production departments related to big data and analytics and the proposed methods and techniques used to forecast future data. Increasing reliability and using the accumulated data were big data strategies. Increasing reliability and using the accumulated data were big data strategies. In particular, the analysis included papers published in renowned scientific journals, conference proceedings, books, theses, and any other related studies that could be deemed significant and follow the standard level of scientific research. The steps were to formulate research questions and then use techniques to address research questions and commerce collecting and analysing obtained data. Finally, the outcome of big data analytics is outlined in this research.

**4.2.2.1. Research strategy**

The research strategy was defined as how the research questions were answered based on procedures. The research study has used the mentioned methods such as surveys, experiments, case studies, and grounded theory (Walshe, 2011). This research has opted to incorporate a case study approach to determine the approaches used to address research questions, considering various strategies.

**4.2.3. Participant**

Company A invited the researcher to demonstrate how to use predictive analytic software to manage production data. Based on the researcher's potential skill gained from work experience as an active participant, the researcher demonstrated the company's procedure. However, the researcher used this predictive analytic to assist the company in adapting to the ever-changing world. With these methods' assistance, it was easier to use big data, although this procedure acquired professional personnel or data scientists with the qualification to handle the massive data. Therefore, the researcher has shown knowledge of how to use big data to benefit the company. Furthermore, the researcher has proved to the company that big data could be believed to predict the future occurrence rather than it being stored.

Company A selected data sets for the researcher to demonstrate, such as workers' absenteeism data, machine data, equipment data, product order data, and error log message data. In the first step, large data sets were imported from the storage platforms and databases into Microsoft Excel. The predictive analytic software was added to Microsoft Excel as an “Add-On”. The data were split along with delimiters such as semi-colon or a comma into a tabular shape. The researcher extracted items from the data entries to be searchable easily. The words and space were removed on the single column to allow the trimming of the sheet. The second step, on the sheet, was a formatting adjustment. It was crucial to harmonize the format data for consistent decimal separators, such as comma and dot. The correct format was checked to automatically ensure that it contained numerical data, which highlighted inconsistent data type and replaced unrecognised character. The sheet identifier has checked automatically: outliers, duplication, consistent and missing value rows were removed. The various data sets were combined for prediction analysis.

**4.2.4. Study strategy to the research**

This research's success required reliable data from broad historical data in the different storage platforms that existed. However, when preparing the case study, the time and budget variables were a concern. For this reason, the correct option to gain knowledge from the data was called direct participation. This was useful for using a predictive analytical approach for both the business and the participant.

**4.2.5. Sources of data collection**

The researcher selected the intrinsic case study method to understand more about data sets. This method has been involved in what has been observed happening in the participatory process. Although this case study is not a research method, the researcher selected the data collection method and analysis that generated the material suitable for qualitative techniques such as participant observation assessment. In assessing valuable information to understand the mechanism behind direct participation outcomes and analyse the company's improvements, the qualitative data approach played an important role. The researcher participated in the company being observed in the manner of demonstration. Company A selected five data sets on themselves from data archives and databases to be used by the participant. The criteria used by company A for these data sets were based on the spreadsheet's size and data integrity. Data integrity was referred to as the quality of the data about correctness, trustworthiness, and accuracy. According to Cooper and Schnindler (2014), there are two main sources of data collection. The primary and secondary data sources are those sources. In general, primary data is referred to as the data obtained from first-hand interactions. Secondary data referred to the data collected from the primary sources, such as data archives and databases. However, this research study has chosen to focus on secondary data to answer research questions. Therefore, this data source contributed to the research objectives and assisted in generating conclusions and recommendations.

**4.2.5.1. Secondary data**

According to Sekaran and Bougie (2016), secondary data are derived from the existing data. Data came in different types of data that were structured, semi-structured, and unstructured. This research used data sets that were designed to assess the aggregate of data. These structured data sets were easy to use because of the arrangement format.

**4.2.6. Data collection**

The collection of data examined the existing secondary source (Hodder et al., 2012). Data was collected from the Microsoft Access tool's data archives and databases. Microsoft Access was used to assist the stored data for reference, enabling the management to update, report, and query. According to Wilcox et al. (2012), data collection gathers and measures information that enables capturing quality evidence and converting it into data analysis using data collection instruments. The Microsoft Excel tool was used to import the data sets from Microsoft Access. The predictive analytic technique was then added to Microsoft Excel to assist with gaining deep insights and predictions. According to Baker (2010), a participant who acts as an active participant differs because it required special training to obtain this skill. The researcher has obtained this skill, participating in the company. The company is an exemplar of innovation that was recognized. This analysis's findings are not founded on a strong basis to claim the outcome for the scientific generalisations (Zainal, 2011). The company was recognised as an exemplar of innovation. Therefore, this study does not claim the outcomes from a strong basis for scientific generalisations (Zainal, 2011).

**4.2.7. Data analysis method**

The prediction results were analysed using the predictive analytic technique. Thomson (2014) describes this data analysis method to analyse data by organising it into categories based on themes. LeCompte and Schensul (2012) emphasised using data analysis to summarise the data into a story and interpret it to extract information. According to Rossman and Marshall (2016), data analysis is messy and time-consuming, they applaud that structure, and a creative and fascinating process can collect meaningful mass data. Maxwell and Chmiel (2014) described that this data analysis method is organised based on concepts. Therefore, the generated data analysis findings have transformed the raw data to create new knowledge.

Company A's goal was to discover knowledge and gain an ideal result from the selected data sets that emerged in databases. This has brought fast development of information to the new technology. Zhou et al. (2016) defined this process as an iterative sequence of five steps. This study adopted predictive analytics: importing data, preparing data, sampling the data, predictive modeling, predictive insight, and data visualization.

**4.2.7.1. Importing data**

This data was imported from the Microsoft Access database repository to determine the trends and patterns. Over the years, companies have used Microsoft Excel to analyse data. In the present time, Microsoft Excel was used for the powerful capability it has for data analysis.

**4.2.7.2. Preparation of data**

Over the years, there has been significant advancement of techniques. This advancement has not been matched in the preparation of data. Therefore, stronger new techniques were required. Davenport (2013) stipulated that preparation data consumes much time. However, data extracted from the database contains incomplete or missing values. The most significant step in big data analytics was data preparation, in which data processing was carried out to improve the quality of big data. Duplication records and outliers were included in the data extracted by the tabular format's predictive analytic technique. Duplications of records have been discarded, and data wrangling has manipulated and cleaned every volume and format into a functional format as a new process. Errors and missing values were omitted from the tabular format.

Data preparation operations included various methods used mainly for duplications and errors (Salmon et al., 2014). The presence of outliers in these selected big data sets has degraded the predictions' quality of knowledge patterns. Many methods, such as the anomaly detection and removal process, were used to identify and eliminate outliers from big data to generate high-quality data sets (Chen et al., 2017). These methods have improved the quality of the records (Moshtaghi et al., 2015).

Data preparation came from multiple sources with high volume and various attributes that were refined into information assets used for accurate analysis and valuable insights. It was applied to raw data that has missing values or outliers and attributes. This process ensured it had resolved the data issues to improve the quality of data. This data preparation has played a crucial role in producing accurate and reliable models with high accuracy and performance. It avoids misleading data preparation forecasts to provide accurate performance; it was necessary to spot data problems earlier. Missing values were filled out, and incomplete data was smoothed out as the key and essential phase of the data preparation task that handled the inconsistent data. Some rows did not value interest or conflicting data attributes, duplicate records, and some other random errors in the data set. As the first step in data planning, all these data quality concerns have been resolved.

For these reasons, data preparation was a crucial part of the predictive analytical technique, and most of the exercise time was required. Dozens of transformations repeated multiple times consisted of the planning process. Both of those transitions took a considerable amount of time and effort, despite advancements in data-working technologies. To the researcher, working with massive, complex data remained a challenge. Mosyakin et al. (2015) identified data preparation as messy and disorganized data that holds data scientists back, according to data science. It was recorded in the same study that 70 percent of the time was spent on data cleaning. Therefore, the preparation of data differs according to the analytical methods and applications used. As the technique to prepare the data, data profiling was used.

**4.2.7.2.1. Data profiling**

Data profiling examines the data from existing information sources. It has acted as the monitoring and cleansing data agent. The data that has been emerged after the cleansing process is significant patterns from large data sets (Jaseena and David, 2014). Data profiling collects metadata used to find data to be mined and imports data into different analytical tools, a significant preparatory activity (Munawar et al., 2020). Abedjan et al. (2015) conclude that data profiling can assist handle data sets using metadata generation for data sets.

To determine suitability for sourcing new data obtained from data sets using metadata. It was important to review the data that existed. It was necessary to determine whether sufficient knowledge existed to obtain the desired prediction results realistically. Data derived from data sets for manual data processing ensures the consistency of data quality in data entry. Automated systems for gathering data can be faulty, resulting in inaccurate or incorrect data. Data quality measurement and evaluation criteria were general, categorized into main elements; accuracy and uniqueness (Huang and Lilienfeld, 2016) as defined in Figure 5.

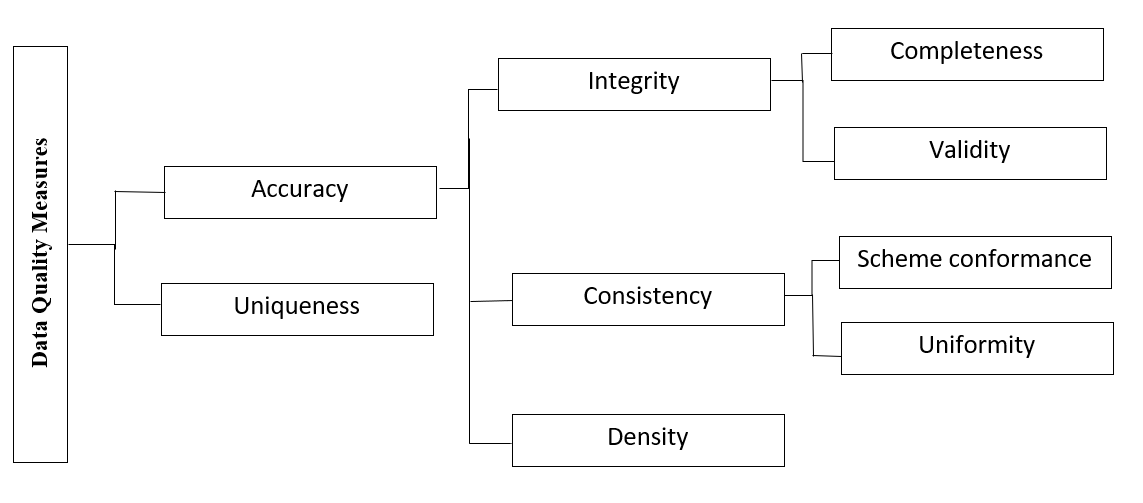


Figure 5: Data quality measures, adapted from Otto et al. (2012)

Accuracy over quality parameters is defined as an aggregate value: integrity, consistency, and density. Intuitively, this explains how the data is a reliable, coherent, and complete representation of the elements. The study defined each precision criterion:

**Integrity**: This refers to the overall accuracy and reliability of information throughout the entire life cycle. Integrity within the databases demands both completeness and validity.

* **Completeness**: A comprehensive representation provides complete data and contains no missing values. Through correcting anomalies and not deleting values, this research achieved completeness in data cleansing. There was typically standard data for this reason, which could be a reference to measure completeness.
* **Validity**: There were no constraints were violated; information was validated. The multiple validity enhancement mechanisms include mandatory fields, unique values enforcement, data schema, or structure.

**Consistency**: Syntactic anomalies and contradictions are concerned with that quality. Regarding data consistency, the main challenge was choosing which data source for reliable agreement between different sources.

* **Schema conformance**: This is particularly true for relational database systems, which rely on the data set to adhere to the domain format.
* **Uniformity**: This was associated directly with irregularities.

**Density**: The quotient of missing values in the data was of concern to this criterion. No non-existent values are represented by null values that have the precise meaning of not being identified.

The precision measure was collectively defined by the above three parameters of honesty, accuracy, and density. Uniqueness is the other big quality measure important for assessing the quality of data. Uniqueness was fulfilled when no duplicates were present in the results. Another criterion that was considered for data quality was timeliness. This criterion refers to the data generated by the spreadsheet from the database.

**4.2.7.2. Sampling big data**

Data sampling is a statistical analysis used to define patterns and trends in the broad data collection being analysed to select, manipulate, and analyse a representative subset of data points. This approach was intended to help this study minimize the amount of data and speed up data processing. The big data sampling technique reduced unstructured and semi-structured data to a manageable processing size from large data sets (Lomotey and Deters, 2014). Few data were used through sampling to obtain the overall characteristics of the entire data sets. The role of sampling in big data analysis was studied by Albattah (2016). Albattah (2016) thought that the full volume of data could be done by sampling.

**4.2.7.3. Predictive Insights**

Predictive insight is a system used to identify and detect faults. This system has used mathematical models on Microsoft Excel to detect future behaviour. Predictive analytic uses the past and present data combined with data algorithms. The analytical technique ran more algorithms on the data set for prediction, although a repetitive process because it involves training. Predictive analytics models worked throughout an iterative process. It commerced with pre-processing, then mined the data to understand company objectives, followed by data preparation. Once the preparation was completed, it was iterated once more.

Studies used predictive analysis as the method to predict future unknown events to create informed insight. In this study, the predictive analytic was used on historical data and had fed it into Microsoft Excel to find trends and critical patterns. Predictive analytic was assisted in creating a practical model, but it also played a role in testing the theory. This predictive analytic has allowed company A to become proactive based on the data, not on assumptions.

**4.2.7.4. Data visualization**

Visualization of data is used to visualize events, such as patterns and situations, that cannot be observed directly. Researchers have agreed that the primary meaning is a vision in the dominant sense. Advanced analysis and easy-to-understand visualizations, it was an efficient way to distinguish the value of relationships. Visualization of data is practically applied to the knowledge field. It has developed a significant collection of methods to achieve a qualitative understanding. The following plots were the basic technique:

**4.2.7.4.1. Visualizing big data**

The company can generate and collect data themselves. This data visualization became a valuable addition presented to company A and an easy path to understanding data. Moreover, it can be enjoyable and demanding to visualize details. With this technique, it was simple to use in presenting the data. As a common approach to information volume and complexity, visualization is an interest in company A.

According to Friedman (2014), data visualization's main goal was to communicate information effectively, displayed through graphical means. In this study, the big data was presented in the tabular view—the default type of visualization used for the selected data. However, the data table was an efficient format for data analysis on items placed in a column, while categorical in the rows. Recently, data visualization is being used in almost any industry. This was translated and presented data to company A in a simple way. This table method was useful for presenting data of precise value to the findings.

**4.3. Research credibility**

The predictive analytic method has been exploded over the last ten years. This technique is still valid in this present time and continues to provide credible predictions. In this study, it has predicted more data sets without any errors results. A measure of the predictive value in each application attached to a set of data was used statistically. This technique is estimated based on parameter estimates' statistical significance and the performance to hold out data set consistency to these measures.

**4.4. Research validity**

The validity of the predictive analytic technique corresponds accurately to the real world. Validity was important because it helped determine the prediction and ensure a researcher uses ethics and methods that have genuinely measured the idea. This technique was appropriate for answering the questions (Richardson et al., 2013). The method used statistical to validity the conclusions of the research study found on Microsoft Excel.

**4.4.1. Statistical validity**

This approach used mathematical measurement to consider the importance of this method to determine the cause and effect between variables. This has been referred to as the correct implementation of big data techniques. The spreadsheet's obtained probability was examined to the degree that the inference reached using a statistical approach was correct.

**4.5. Justification**

These methods were suitable for this research to provide the ideal future occurrence prediction outcomes. Although it was important to understand this dissertation study's methods, it was equally important to understand why research methods are important. These methods were important in developing the study's research questions to gather data and select an analytic approach. This has assisted to assess and gaining knowledge on using the data sets. These methods of analysis have been useful for the management of production. The company's big data effect assisted boost production data, increase profit, and improve big data goals.

**4.6. Ethical considerations**

The ethical principles set by the generic ethics of science were honoured by this research. In this research, the company informed the participant of all measures to perform the demonstration. The organisation was more important than the research, and it was usually valued. The company was informed that the demonstration was voluntary and that it would not impact the brand. Johnson and Christensen (2014) described ethics as principles and guidelines that assist in upholding procedure. These guidelines are the core principles of ethics in the research study: meaning voluntary participation, ability to anonymity, and data storage as stipulated by Boylan (2012). As part of the study conditions, it was decided that the name of the case organization, shown in Appendix A, would not be disclosed, only the findings. Burnes and By (2012) draw attention to the emerging tendency to observe research almost entirely in ethical terms. The research was conducted in an incredibly positive manner according to ethical issues. Thus, the organisation is only known as company A. Any public information that could damage the company's authenticity and could result in public unrest was removed. The demonstration process was used only for the research purpose. The information was kept confidential, and findings were not manipulated against the company favouring the competitors, and no misinterpretation nor view has been done. Company A has ethical consideration regarding the company's information, meaning the outcomes of this research have followed the company policies. The study retains the company name; some data had to be withheld or changed slightly (Adelman-McCarthy, 2011). Therefore, the researcher had access to the historical data, known as raw data, and the company supervisor. The future dissemination of the dissertation’s research findings and historical data post-study has followed ethical guidelines.

**4.7. Summary**

There were methods used in this research study for the company using secondary data source analysis. These methods came with advantages to strengthen the study. This research has described the methodology to gain insight on how to manage data using big data analytics.

**Chapter 5: Findings and Results**

The research methodology and strategies used to collect and analyse information are discussed in this chapter. The findings of the study are based on researcher participation in company A. This study's predictive analytic technique was carefully believed to ensure that the findings presented where it is possible. These findings are related to the research questions that guide the research. The results section states the findings after predictive analytics were used without bias or interpretation and arranged in the logical sequence.

**5.1. Findings of the study**

The findings are based on the researcher's participatory phase with company A. This analysis contains six records of observations to ensure rigour and objectivity in the data. The study used data sets of machine downtime, equipment downtime, machine error log message, product order data, worker absenteeism data, and alarm management data to predict the accuracy probability for each data sets used in this study.

**5.1.1. Machine downtime**

The data sets contain the machine downtime records with the thousand range of dates. These records show the machine time breakups during the working hours. This was for machine maintenance not done; as a result, it caused unplanned downtime. The other cause of the machine breakdown was excessive tool changeover. The company may stop production from carrying out maintenance for machinery to perform more efficiently. These downtimes on the machines have caused significant concern for the company. The company tried to prevent machine downtime by reading the machine performance patterns but could not find the solution. The company deployed the researcher to come up with the solution for the challenges.

**5.1.2. Equipment downtime**

The equipment is a tool that was used in operation. These tools posed a significant risk to operations, despite the company’s advancements in control systems. The company has pieces of equipment that were older, which no longer suitable for modern time. The system's machine has no longer supported this equipment unsuitable for the systems. The company required several spare parts on hand, with this old equipment to be used for production.

**5.1.3. Machine error log messages**

The machine error message is messages that appear on the current machine status showing the different logs. These error messages made it difficult for operators to understand the information. The machine was believed to alert the operator on the machine screen status, but it missed some alert logs. The data set was in a time series, consisting of log messages and failure records for 984 days. The challenge was to predict which day was a failure day in advance. The machine log data failures on the features constructed from log files displayed the error messages. Therefore, the task was to conduct a prediction process before these error messages appear on which day.

**5.1.4. Product Order data**

The competition amongst companies increased daily, meeting customer demand orders is a high priority to the company, making it difficult to predict when the demand usually occurs. The company ends up with inventory stock products that do not match what was in demand or have sufficient products because it requires meeting the customer. These excessive products have caused company A capital to be bound due to producer surplus. The designated storage has been increased with excess inventory. The out of stock products resulted in reduced customer satisfaction and store loyalty. For these reasons, customers have shifted to other competitors. The data sets contained the date range of order data. The researcher was deployed to predict which days do the customers usually demand orders.

**5.1.5. Worker absenteeism data**

Worker absenteeism is an employee's absence from work. The company faces a big challenge leading to backlogs, work pilling, and work delays. Company A found it challenging to manage the number of absenteeism workers who do not show up for work. This has mostly impacted both workers and the company’s financial bottom line. Managers of the company spent much money considering hiring temporary workers replacements for those days. This has resulted in workers working overtime to cover other duties to retain production running. The company tried to determine legitimate reasons the number of workers does not show up to work. Although these workers were given certain days off, this had gone beyond the amount of excessive absenteeism. However, some legitimate reasons were intentional and habitual. This has led company A to collect absentee data based on historical data to predict frequent long-term production absence. The company tried to identify which days usually have higher absenteeism and deployed the researcher to predict the absenteeism days from the data set records.

**5.1.6. Alarm management data**

An alarm is a device that provides an audible, visual, or another signal approximately awareness of danger or condition. This device is on the company premise and is believed to scare or alert the operators and workers or threaten situations such as detecting a fire. Alarm signals were sent only when the alarm threshold of both detectors is reached. The company used the alarm for theft and sensors to alert when there is danger. 90% of false alarms were caused by heat, rain, human-made disturbances, transmission errors, and unknown. However, false alarms pose several challenges for the company. This has resulted in disruptions to company A's outsourced security company and wasted time, and anxiety had led to a stressful work environment for the operator. False alarms eroded the confidence workers of the company have in the workplace safety and culture.

The real alarm data set consists of all false alarms. The aim was to reduce all 3000 alerts of false alarm data by predicting accuracy. The false alarm alert began from July 2020 to the end of August 2020. The challenge of these false alarms’ prediction was to conceptualise the same as the anomaly detection Chandola et al. (2010) to predict the alarm failures probability as Salfner et al. (2010) stipulated. Hence, the company took advantage of predictive analytics methods (LeCun et al., 2015) to remove these false alerts.

**5.2. Results**

The results found during the organization's participatory evaluation and the outcome of the six data sets are discussed in this section. The purpose of the results section is to clarify what the data sets have revealed. This summarises all data sets by predicting data in this study.

**5.2.1. Machine downtime results**

The machine downtime revealed 80% of break-ups based on the data set records. Downtime costs the company thousands an hour. It was revealed that machine maintenance was the main cause of breakdowns. With this probability, the company could have spent more money on the machine. The company took the following paths approximate collective decision-making:

* Provide a machine monitoring system as an effective way to understand what happened underneath and minimise unplanned machinery.
* Incorporating preventative maintenance tasks into everyday schedules is an excellent way to avoid unplanned downtime.
* Train and empower workers who are likely to prolong downtime events through maintenance and accurate documentation potentially.

**5.2.2. Equipment downtime results**

The equipment's probability was based on the data sets used; 92% of the equipment breaks down has contributed to the unexpected failure of equipment to work. Some equipment was 15-20 years old, and the machine was no longer supported by manufacturing. Parts became unavailable or were made of the country and took weeks to be delivered. The company made the following decisions to minimise risk:

* To track equipment's lifetime. If the piece of equipment has been running for a long time beyond the planned useful life, it is far more likely to malfunction. The benefit of retaining track of how old every piece of equipment is and what a lifetime of very substantial maintenance it has had.
* To improve worker communication to ensure that everyone knows which tasks are the responsibility.
* To simplify maintenance to ensure that everyone knows one key to reducing downtime, ensuring equipment obtains preventative maintenance.
* To document equipment performance. A piece of equipment provides sign equipment that would go before it fails.

**5.2.3. Error machine log message results**

The results obtained from the predictive analytic technique based on the data set used for prediction revealed that out of 984 days, the error messages pop-up on the machine display monitor for 872 days repetitively. This has shown that 88% of errors appeared on the log file, resulting in the machine not operating. The decision was taken to erase these error messages on the display of the machine. The following changes were proposed:

* Hire professional technicians to remove all errors that occur on all applications of the computer.

**5.2.4. Product order data results**

The data set has 60 rows and 13 columns. The data set was collected for 60 days. The data set had twelve predictive attributes and a target that was an order for daily demand. The company had to discover which days were the highest peak order demand from customers. The predictive analytic technique estimated the peaks were in the first week on monthly. The decision was made urgent for product order demand. The company produced a product every first week of the month to meet customer requirements.

This approach predicted product demand and the correct quantities of inventory storage to anticipate the customer's demand. This company has learned that in the future, they may produce the product when there is customer demand, which is usually in the first week on a monthly or operates only performed when there is a data signal.

**5.2.5. Worker absenteeism data**

The data set consisted of 500 workers with presence and absence numbers. The results found out that most of the workers have not shown up to work. The predicted days were on Saturdays and Sundays; those were the days that workers had gone absent. The company proposed that some workers should work on weekends to avoid or minimise the number of absenteeism. The company also created an absence policy to manage absenteeism. These rules were made for absentees.

**5.2.6. Alarm management data results**

The alarm data set comprised of false and true alarm records that were conducted to be predicted using the predictive analytic technique. The values of each alarm alert were range between 1 to 10 minutes. In these records, it was predicted that 92% false alarm alerts. An outsourced security company has defined a false alarm as an alarm signal transmitted when no alarm condition exists. False alarms were classified according to the conditions that caused them: environment, animals, operators’ errors, human-made disturbances, and equipment malfunction, and unknown causes. The following decisions were made to ensure that there was no occurrence of false alarms:

* All alarms should be processed as a legitimate alarm condition.
* All alarms caused by events should be accounted for, controlled, or removed. When a device malfunctions, an operator makes an error or an unaccounted event such as an environmental occurrence that causes an alarm condition. It indicates that some part of the system is not functioning as planned.
* All facilities should implement 100 percent reliable systems to detect real alarm conditions and not transmit alarm signals when no alarm conditions exist.
* Training, management, and procedures require to be reviewed.
* The sensors should be self-checking and resistant to tampering and sabotage.

These results demonstrated were a directly typical alarm for prioritisation and had a considerable potential to reduce costs significantly.

This chapter focused on determining the big data analytic strategy's effectiveness to use data. Methods of data analytics, research conclusions, and a discussion of the results have been presented. The findings of this research are consistent with the results of other related Big Data Analytics studies. The results were identified and presented as a methodological phenomenon in the study. The results of the findings for outcomes are found using the predictive analytical methodology in the coming chapter. The contribution of this study and limitations will be discussed.

**Chapter 6 Recommendations and Conclusions**

This final chapter draws together the main conclusions from an intrinsic case study of the study. It discusses a follow-up overview of the study and finally makes general recommendations on responding to big data analytic. The preceding intrinsic case study illustrates the impact of analytics on big data. Based on the introduction of this study, the following conclusions are made. This study aimed to examine the range and essence of decision-making in managing the production environment. The conceptions, procedures, and results of the previous chapters' findings are summarised in this chapter.

**6.1. Research process summary**

This study conceives to address the fundamental questions: *Does the company understand how to manage big data analytics?*

The involvement of researchers in the sector has changed the mindset around data storage. This skill set has altered how management handles multiple generated data. It appears that the organization has adopted a new climate. The participation of the researcher in the company has changed the mentality regarding storing data. However, this skill set has changed how the management manages various data generated. The company appears that has brought a new environment. The knowledge of how to use this technique has changed the management perception towards managing big data. This was important to the company.

*What is the analytical technique that would be suitable for managing big data?*

This technique has leveraged the production data to the extent that the management can predict company data. Therefore, this has created many opportunities in the company for job opportunities that are related to data.

*How would the company govern the data and change the future data outcome?*

The company has deployed data scientists to govern the data predicted for patterns and future occurrences.

The theoretical framework, informed by the literature, situates industry-based learning on big data and theory. The supporting literature underscores the gains made in using tested theory to enrich big data. The research design and methods of obtaining data from the company that addressed the research question are the contemporary embedded design of the qualitative approach in qualitative data collected and analysed simultaneously. The case study method was applied because the central issues raised were concerned with company A from which most of the data was collected to gain insights.

**6.2. Summary findings**

This section of the research study addresses the summary of data sets that were used for prediction. The finding reflects the significant results of the study. A component of the findings' summary is to discuss each of the findings, using a predictive analytic technique that justifies rather than distorts the findings' intent. The following data sets were summarised as critical findings.

**6.2.1. Data set theme 1**

This predictive analytic has empowered production managers to detect the potential defects before materializing and avoid unnecessary maintenance for parts that do not require repair. It has prevented unnecessary downtime by allowing potential challenges fixed without stopping the machinery. These analytics have helped the production managers spot potential problems by comparing historical machine data with the current performance. However, there is much more than identifying future downtime failure; it also includes predictive analytics uses by monitoring how machinery performs. Using these new strategies have emerged, changing maintenance forever.

**6.2.2. Data set theme 2**

This analytic technique has analysed the historical data trends by minimizing the risk of the equipment and maximizing opportunity, which has allowed managers to make proactive rather than reactive decisions—being proactive focuses on eliminating challenges before it appears. The reactive approach is based on responding to events after they have happened. These approaches assisted in determining the condition of equipment and predicted when equipment should be maintained. This technique has eliminated the number of failures that occurred during production as any breakdown of equipment. Unexpected equipment failures have triggered process flow interruptions, damaged equipment, product losses, and diverted staff from preventive maintenance (PM) duties during downtime. These implications have resulted in costly overtime hours and decreased completion rates. Nonetheless, as separate records, for instance, the work of failure were kept. More predictive analytics contributes to more downtime analysis.

It can be verifiable and condensed to explain the equipment for replacement. Equipment breakdown patterns were exposed by statistical analysis of downtime history. Tracking downtime has a significant influence on how, when, and where resources for maintenance are used. Time spent in logging equipment can also easily be justified by the time and saved for maintenance.

**6.2.3. Data set theme 3**

The logging errors encourage debugging, making it easier to correct bugs with smart error logging. The errors were solved more quickly and easily with the information available from error logs. This technique l has correctly predicted the alert of notification. It had high accuracy, was associated with a unique log message with a numeral value, and memorized the combination of certain messages, locations, and log types that result in alerts. This has handled each message entirely rather than learning patterns with the text. The machine changed settings for each message related to the challenge, repairing corrupted files, and updating drivers to restore the machine to a normal routine.

Terms and data formats were synchronized to assist in easy analysis and reduce error. The normalization of statistics and data from multiple sources is meaningful and precise. In addition, log data can be centralized and structured in practical ways, understood by humans, and interpreted by machine learning and this technology and processes. Aggregating log data from different sources correlate logs to distinguish trends and patterns more quickly.

**6.2.4. Data set theme 4**

Customers were critical to the company meeting the customer's demand can be the difference between a company's success and one that fails. The company considered several changes when devising a product or service. These include cost, location, and promotion. The company used work order management software, which can be a useful tool for organising workload. It was important not to overproduce a product, as this can lead to losing money. An informed decision on the product requires checking the past data. Beyond these basic expectations, the company can offer customer questionnaires to fill out to understand individual requires further. The make-to-order strategy for production is driven by demand, and it only produces items when orders.

**6.2.5. Data set theme 5**

Workers' absenteeism is a serious challenge to any company; it increases the risk of low production. The changing of shifts for workers has worked well for the company. This has resulted in the company observing workers' full attendance, which has increased the production output. It has motivated the company to consider a bonus scheme for workers who have a lower number of absences.

**6.2.6. Data set theme 6**

These false alarms have threatened the outsource security company and company A. The data set was used to reduce the false alarms from the company. Alarm sensors a condition such as heat and rain triggered the false alarm. Predictive analytic classifies alarms with accuracy. The alarm is back to the normal routine; the outsourcing company has no challenges regarding the alarm from company A.

**6.3. Implications for research**

Big data has dramatically changed the world and is poised to become a crucial component of the production industry. Over the years, big data has changed the outlook of companies. This assists organisations in improving company bottom lines, gaining ability, spot meaning innovations, and transforming. Big data is reshaping and changing how organisations function and operation. Studies have shown the advantages of using it in different fields. This study explored how the data sets used big data analytics techniques (Oussous et al., 2018) as detailed yet rapidly evolving data. Big data can alter research and enhance the manufacturing industry's outcomes and recognise company habits using big data analytics during data prediction. Big data analytics provide production management's ability to match the company's interest in the historical data available; this analytic is chosen as the ideal prediction technique.

Big data provides production management an understanding of each data set and the predictive analytics technique with the most beneficial effect on an individual data set. Using big data for guiding instructions in big data has a significant role in enhancing production data. Big data impact many aspects of society, resulting in societal benefits. It provides benefits to use data, as using big data enables the management to decide predictively.

This process has embedded predictive analytics in applications, production managers to monitor equipment's condition and performance, and predict failures before it occurs. This research study implicates the management of production for using big data. This research study has presented two implications for production managers: accuracy and usability of big data. There is much hype surrounding big data, emphasizing this to commerce a big data initiative. Poor data management can doom big data initiatives regularly. Companies have no hope of fully mastering big data and achieving success at various data use levels without fundamental management practices.

**6.3.1. Data quality**

One common challenge to the organisation is to gather data from various sources, sometimes being of a lower quality than others. The company should handle data; thus, poor data is thrown out, yet poor management practices have not happened. Poor data quality may come from information that is old, lost, and misused. It is no longer sufficiently reliable to delete the old data; substituting it with output information would lead to better quality data.

**6.3.2. Relevance of data**

The accuracy of data quality between data content and the production data of the respective data will be obtained. The organisation requires to handle the relevant information, gather as much data as possible, and work from there. However, some useful lessons could be missed by companies’ deficiency structured data. It comes down to understanding what knowledge is important and maximizing use. This attempts to incorporate an unstructured data set that would only confuse the predictive analytics model the company tried to create.

**6.3.3. Data cohesion**

Organisation data management practices lead to excellent data collection. Spending a great deal on the new technology, not on the data, is a sign of poor data management being practiced. Suddenly, this engagement in predictive analytics allows obtaining more out of the data. These elements cooperated to present an interface to a hidden structure.

Managing poor data can have a significant effect on the morale of workers. In manual data clean-up, workers who have been recruited for high-skill jobs are unlikely to find satisfaction. A poor and incomplete collection of data will result in a revenue loss. Effective management is required for the big data initiative stage and missing any move will not produce the desired results. These changes and challenges provide the organisation and the way it functions closer to the governor of data. This implies using big data, evidence, and facts and has a final goal in pursuing truth.

**6.4. Implication for practice**

The research results were useful to company A. The case organisation pioneers in solutions; it has provided the company with up-to-date information approximately using the technology. The tremendous opportunity that it offers, important not to be an illusion for them. Big data can improve the efficiency of comprehensive data analytics; this includes predictive analytics. These analyses can provide predictive statistics on the entire data set on a large and complete scale. The data opportunities are continually changing, and with that comes the require for an iterative approach toward data mining and analysis. A better big data analytics system can agile and iterative as new technology and data opportunities emerge. Big data itself can assist drive this evolution, though the researcher did not find studies implicating big data analytics for company intelligence purposes.

Drawing big data analytics interference from large sets of data assists in identifying hidden patterns. To uncover these insights, big data analysts use data mining, predictive analytics, and optimisation. Observing past results, the advancement of predictive analytic technology opens new possibilities for predicting future events. Big data recently allows data scientists to analyse vast volumes of data and hope that accuracy can only improve future predictions. However, predictive analytics accuracy affects the practice of big data.

Big data is becoming increasingly prevalent and affects management, how research is done, and policy implementation. To advance the vision of promoting the management discipline requires to optimize the advantages of big data. Current practice requires managing to be driven by a data policy. To step into technology, the practice of big data reached production. In the traditional data analytics landscape, it is easy. Big data analytics is a rapidly developing phenomenon with a limited understanding of the implications. Therefore, recognizing and exploring ethical implications is critical for data analytics users. This predictive analytical technique has posed many ethical problems, particularly when companies externally monetize data for different purposes from which analytics can be carried out. The ethical framework of big data analytics has been updated entirely. Therefore, big data analytics has significantly stepped up in research for cognitive computing.

**6.5. Contribution to the study**

This study aimed to analyse the effect of big data on organisational efficiency with the assist of analytics. Strong evidence was found in this study that big data has a positive effect on organisational efficiency. The results of the massive, big data sets selected in this study converge to the conclusion that organisations benefit in different forms from a big data initiative. However, to adapt to this revolution, resources must be put in place, and managers require to develop new skills and behaviours. These findings assist in strengthening and support the idea that big data analytics transformation is beneficial for an organisation.

**6.6. Limitations and delimitations of the study**

The study focuses on delimitations on the production data sets generated by the organisation. The study only focuses on how to predict future outcomes using the data sets for a single organisation. The delimitations work addresses the following points:

* Being selected as an active participant in this study required to conduct this demonstration to the company was mainly chosen from the production company with high data generated.
* Companies of service providers were excluded from this work.
* Improvement procedures for the use of big data have been related to predicting practice, irrespective of other approaches that may be useful and relevant.

The limitations of the current study set the course for future research opportunities. The study excluded unstructured data and incomplete blank data from the data sets, which may have affected the study's outcomes. The conditions were beyond the control of this study that placed restrictions on being supervised. On the other hand, when carrying out this analysis, it defined the limitations as uncontrolled variables. The time constraint was the first factor; this work was based on a case study and took more time to assimilate much analysis. The second factor was the scarce resources and literature in the production environment for extracting data from databases, works correlated with expected, structured observation.

**6.7. Recommendation for future research**

The study's recommendations are directed to the company studied and future reference to the companies that require to use this study using the predictive analytic technique to predict future outcomes. Upon building the rich underpinning of the research findings gained in the dissertation study, this has shown the potential to use big data analytics to predict future occurrences. Big data is a vital area that presents potential advantages and innovation, a fantastic sector with a bright future for production data when the methodology is correct. The big data challenge is mainly due to the size, which requires proper storage, management, integration, cleaning, processing, and analysis. The complexity of dealing with conventional data management is intensified by the sheer volume, velocity, and variety of data. It generates a need to research and explore new analytics methods that could help solve these challenges and encourage the positive role of big data analytics in the production industry sectors.

This research reveals opportunity strength to research based on the qualitative case study approach, as the most analytical approach for analysing. The company is using and appreciating the value of predictive analytics. In the history of data analysis, the availability of big data has created a remarkable moment. Big data analytics’ strategic value makes sense governed by what is stored, analysed, and accessed managing data. This study recommends future work to look at the rise of unstructured data and data cleansing in production. It suggests that dealing with these would benefit the companies to predict without errors. This would open or create opportunities for the company to hire data scientists to clean up the big messy data. Future research studies have focused on using storage such as Hadoop and cloud platforms to store various forms of data. However, new analytical methods require to be built to manage big data challenges, such as unstructured data and data cleaning for processing when the volume of data is high (Oussous et al., 2018).

**6.8. Conclusion**

The purpose of this study was to provide a literature review on the use of analytics for big data. This began with a general background on the topic, including definitions and characteristics of big data, followed by a review of big data analytics methods and tools. Using predictive analytics has assisted management in managing big data for using this technique to gain insights. This dissertation aimed to generate new knowledge and better understand how big data analytics initiatives in the production industry can be applied. The primary objective in these two months of the case organisation, and thus the most important impact of the big data initiative, was to achieve process and efficiency improvements in the impact of development. Using predictive analytics to predict large, diverse data sets for future outcomes enables identifying opportunities that drive a positive impact in the area.

Company A has taken advantage of big data analytics to gain insight and made the production a conducive environment. Findings have contributed to the company and have turned these technologies into methods to manage data effectively. Using the predictive analytics model to company A, current data to predict the ever-increasing data based on historical data has presented a new dimension. It has solved the challenges of managing the data and has presented new opportunities and adopted these techniques for future use. This has led the company to create a job opportunity for data scientists, data analysts, and data engineers to make informed management decisions. This has increased knowledge of the companies and understanding of managing the data. Predictive analytics tools have a significant impact on companies and managing large volumes of data.

The results demonstrated were about using big data analytics on a storage platform to gain insight and make decisions that would benefit the company. This analytics has proven that these approaches used on company A’s historical data effectively manage the production data.

**7. References**

Abbott, D., 2014. *Applied predictive analytics: Principles and techniques for the professional data analyst*. John Wiley & Sons.

Abedjan, Z., Golab, L. and Naumann, F., 2015. Profiling relational data: a survey. *The VLDB Journal*, *24*(4), pp.557-581.

Abouzied, A., Bajda-Pawlikowski, K., Huang, J., Abadi, D.J. and Silberschatz, A., 2010, June. HadoopDB in action: building real world applications. In *Proceedings of the 2010 ACM SIGMOD International Conference on Management of data* (pp. 1111-1114).

Agarwal, R. and Dhar, V., 2014. Big data, data science, and analytics: The opportunity and challenge for IS research. IS research Journal.

AGGARVAL, G.B., JACKSON, E. and BONEH, C., 2012. An Analysis of Private Browsing Modes in Modern Browsers. USENIX 2010.

Aiken, L.H., Sloane, D.M., Bruyneel, L., Van den Heede, K., Griffiths, P., Busse, R., Diomidous, M., Kinnunen, J., Kózka, M., Lesaffre, E. and McHugh, M.D., 2014. Nurse staffing and education and hospital mortality in nine European countries: a retrospective observational study. *The lancet*, *383*(9931), pp.1824-1830.

Akter, S., Wamba, S.F., Gunasekaran, A., Dubey, R. and Childe, S.J., 2016. How to improve firm performance using big data analytics capability and business strategy alignment?. International Journal of Production Economics, pp. 113-131.

Albattah, W., 2016, November. The role of sampling in big data analysis. In *Proceedings of the International Conference on Big Data and Advanced Wireless Technologies* (pp. 1-5).

Andrews, J.O., Tingen, M.S., Jarriel, S.C., Caleb, M., Simmons, A., Brunson, J., Mueller, M., Ahluwalia, J.S., Newman, S.D., Cox, M.J. and Magwood, G., 2012. Application of a CBPR framework to inform a multi-level tobacco cessation intervention in public housing neighborhoods. *American Journal of Community Psychology*, *50*(1-2), pp.129-140.

Anuradha, J., 2015. A brief introduction on Big Data 5Vs characteristics and Hadoop technology. Procedia computer science, 48, pp.319-324.

Baker, R.S.J.D., 2010. Data mining for education. *International encyclopedia of education*, *7*(3), pp.112-118.

Bancroft, A., Karels, M., Murray, Ó.M. and Zimpfer, J., 2014. Not Being There: Research at a Distance with Video, Text and Speech', Big Data? Qualitative Approaches to Digital Research (Studies in Qualitative Methodology, Volume 13).

Bathla, G., Aggarwal, H. and Rani, R., 2018. A Novel Approach for Clustering Big Data based on MapReduce. International Journal of Electrical & Computer Engineering (2088-8708), 8(3).

Bestavros, A., Lin, K.J. and Son, S.H. eds., 2012. *Real-time database systems: Issues and applications* (Vol. 396). Springer Science & Business Media.

Bhogal, J. and Choksi, I., 2015, March. Handling big data using NoSQL. In 2015 IEEE 29th International Conference on Advanced Information Networking and Applications Workshops (pp. 393-398). IEEE.

Bibri, S.E. and Krogstie, J., 2020. The emerging data–driven smart city and its innovative applied solutions for sustainability: the cases of london and Barcelona. Energy Informatics, 3(1), pp.1-42.

Borgman, C.L., 2015. *Big data, little data, no data: Scholarship in the networked world*. MIT press.

Boyd, D. and Crawford, K., 2011, September. Six provocations for big data. In *A decade in internet time: Symposium on the dynamics of the internet and society*.

Boyd, D. and Crawford, K., 2012. Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, communication & society*, *15*(5), pp.662-679.

Boyd-Graber, J., Mimno, D. and Newman, D., 2014. Care and feeding of topic models: Problems, diagnostics, and improvements.. Handbook of mixed membership models and their applications Journal , Volume 225255.

Bragazzi, N.L., Dini, G., Toletone, A., Brigo, F. and Durando, P., 2016. Leveraging big data for exploring occupational diseases-related interest at the level of scientific community, media coverage and novel data streams: the example of silicosis as a pilot study. PLoS One, 11(11), p.e0166051.

Breed, D.G. and Verster, T., 2019. An empirical investigation of alternative semi-supervised segmentation methodologies. South African Journal of Science, Volume 115, pp. pp.92-98.

Brinkmann, S., 2016. Methodological breaching experiments: Steps toward theorizing the qualitative interview. *Culture & Psychology*, *22*(4), pp.520-533.

Bughin, J., 2016. Big data: Getting a better read on performance. *The McKinsey Quarterly*.

Burnes, B. and By, R.T., 2012. Leadership and change: The case for greater ethical clarity. *Journal of business ethics*, *108*(2), pp.239-252.

Chandler, D., 2015. A world without causation: Big data and the coming of age of posthumanism. *Millennium*, *43*(3), pp.833-851.

Chandola, V., Banerjee, A. and Kumar, V., 2010. Anomaly detection for discrete sequences: A survey. *IEEE transactions on knowledge and data engineering*, *24*(5), pp.823-839.

Chaudhuri, K., Monteleoni, C. and Sarwate, A.D., 2011. Differentially private empirical risk minimization. *Journal of Machine Learning Research*, *12*(3).

Chen, C.P. and Zhang, C.Y., 2014. Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. *Information sciences*, *275*, pp.314-347.

Chen, J., Sathe, S., Aggarwal, C. and Turaga, D., 2017, June. Outlier detection with autoencoder ensembles. In *Proceedings of the 2017 SIAM international conference on data mining* (pp. 90-98). Society for Industrial and Applied Mathematics.

Chen, X., Lu, X. and Yang, G., 2012. Sources identification of heavy metals in urban topsoil from inside the Xi'an Second Ringroad, NW China using multivariate statistical methods. *Catena*, *98*, pp.73-78.

Constantiou, I.D. and Kallinikos, J., 2015. New games, new rules: big data and the changing context of strategy. Journal of Information Technology, Volume 30, pp. 44-57.

Cooper, D.R. and Schindler, P.S., 2014. *Business research methods*. McGraw-Hil.

Cui, Q., Gong, Z., Ni, W., Hou, Y., Chen, X., Tao, X. and Zhang, P., 2019. Stochastic Online Learning for Mobile Edge Computing: Learning from Changes.. IEEE Communications Magazine journal, Volume 57, pp. 63-69.

Cumbley, R. and Church, P., 2013. Is “big data” creepy?. Computer Law & Security Review, 29(5), pp.601-609.

Cumming, G.S., Allen, C.R., Ban, N.C., Biggs, D., Biggs, H.C., Cumming, D.H., De Vos, A., Epstein, G., Etienne, M., Maciejewski, K. and Mathevet, R., 2015. Understanding protected area resilience: a multi‐scale, social‐ecological approach. *Ecological Applications*, *25*(2), pp.299-319.

Das, T.K. and Kumar, P.M., 2013. Big data analytics: A framework for unstructured data analysis. *International Journal of Engineering Science & Technology*, *5*(1), p.153.

D'Assunção, F.L.C., Sousa, J.C.N., Felinto, K.C.A., de Medeiros, T.C., Leite, D.T., de Lucena, R.B. and de Oliveira Lima, J., 2014. Accuracy and repeatability of 3 apex locators in locating root canal perforations: an ex vivo study. *Journal of endodontics*, *40*(8), pp.1241-1244.

Dave, M. and Gianey, H.K., 2016, December. Analysis of big data for data-intensive applications. In *2016 International Conference on Recent Advances and Innovations in Engineering (ICRAIE)* (pp. 1-6). IEEE.

Davenport, T.H. and Dyché, J., 2013. Big data in big companies. *International Institute for Analytics*, *3*, pp.1-31.

Davenport, T.H., 2013. Analytics 3.0. *Harvard business review*, *91*(12), pp.64-72.

De Mauro, A., Greco, M., Grimaldi, M. and Ritala, P., 2018. Human resources for Big Data professions: A systematic classification of job roles and required skill sets. Information Processing & Management, 54(5), pp.807-817.

Dean, J., 2014. *Big data, data mining, and machine learning: value creation for business leaders and practitioners*. John Wiley & Sons.

Deb, S., 2014. Information technology, its impact on society and its future. *Advances in Computing*, *4*(1), pp.25-29.

Dekhtiar, J., Durupt, A., Bricogne, M., Eynard, B., Rowson, H. and Kiritsis, D., 2018. Deep learning for big data applications in CAD and PLM–Research review, opportunities and case study. *Computers in Industry*, 100, pp.227-243.

Duan, Y., Edwards, J.S. and Dwivedi, Y.K., 2019. Artificial intelligence for decision making in the era of Big Data–evolution, challenges and research agenda. *International Journal of Information Management*, 48, pp.63-71.

Easton, G., 2010. Critical realism in case study research. *Industrial marketing management*, *39*(1), pp.118-128.

Eberendu, A.C., 2016. Unstructured Data: an overview of the data of Big Data. *International Journal of Computer Trends and Technology*, *3*(1), pp.46-50.

Elgendy, N. and Elragal, A., 2014. Big data analytics: a literature review paper. s.l., Springer, cham, pp. 214-227.

Elgendy, N. and Elragal, A., 2016. Big Data Analytics in Support of the Decision Making Process. s.l., Elsevier.

Elgendy, N., 2013. Big Data Analytics in Support of the Decision Making Process. MSc Thesis. p. 164.

Elragal, A. and El-Gendy, N., 2013. Trajectory data mining: integrating semantics. *Journal of Enterprise Information Management*, *26*(5), pp.516-535.

Fan, J., Han, F. and Liu, H., 2014. Challenges of big data analysis. *National science review*, *1*(2), pp.293-314.

Feldman, D., Schmidt, M. and Sohler, C., 2020. Turning big data into tiny data: Constant-size coresets for k-means, PCA, and projective clustering. *SIAM Journal on Computing*, *49*(3), pp.601-657.

Finkelstein, A.S., 2015. *Carry on: American women and the veteranist-commemoration of the First World War, 1917-1945* (Doctoral dissertation).

Fournier-Viger, P., Gomariz, A., Campos, M. and Thomas, R., 2014, May. Fast vertical mining of sequential patterns using co-occurrence information. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining* (pp. 40-52). Springer, Cham.

Friedman, V., 2014. Data visualization and infographics (2008).

Gandomi, A. and Haider, M., 2015. Beyond the hype: Big data concepts, methods, and analytics. *International journal of information management*, *35*(2), pp.137-144.

Gärtner, B. and Hiebl, M.R., 2018. Issues with Big Data. *The Routledge Companion to Accounting Information Systems (S. 161-172). New York: Routledge*.

Gartner, E. and Hirao, H., 2015. A review of alternative approaches to the reduction of CO2 emissions associated with the manufacture of the binder phase in concrete. *Cement and Concrete research*, *78*, pp.126-142.

George, G., Osinga, E.C., Lavie, D. and Scott, B.A., 2016. Big data and data science methods for management research.

Ghadimi, A., Saidur, R. and Metselaar, H.S.C., 2011. A review of nanofluid stability properties and characterization in stationary conditions. *International journal of heat and mass transfer*, *54*(17-18), pp.4051-4068.

Ghani, N.A., Hamid, S., Hashem, I.A.T. and Ahmed, E., 2019. Social media big data analytics: A survey. *Computers in Human Behavior*, 101, pp.417-428.

Ghasemaghaei, M. and Hassanein, K., 2015. Online information quality and consumer satisfaction: The moderating roles of contextual factors–A meta-analysis. Information & Management, 52(8), pp.965-981.

Gnanadesikan, R., 2011. *Methods for statistical data analysis of multivariate observations* (Vol. 321). John Wiley & Sons.

Grover, V., Chiang, R.H., Liang, T.P. and Zhang, D., 2018. Creating strategic business value from big data analytics: A research framework. *Journal of Management Information Systems*, 35(2), pp.388-423.

Guenter, B., Joshi, N., Stoakley, R., Keefe, A., Geary, K., Freeman, R., Hundley, J., Patterson, P., Hammon, D., Herrera, G. and Sherman, E., 2017. Highly curved image sensors: a practical approach for improved optical performance. *Optics Express*, *25*(12), pp.13010-13023.

Günther, W.A., Mehrizi, M.H.R., Huysman, M. and Feldberg, F., 2017. Debating big data: A literature review on realizing value from big data. *The Journal of Strategic Information Systems*, *26*(3), pp.191-209.

Guo, H., Liu, Z., Jiang, H., Wang, C., Liu, J. and Liang, D., 2017. Big Earth Data: a new challenge and opportunity for Digital Earth’s development. International Journal of Digital Earth, 10(1), pp.1-12.

Gupta, M. and George, J.F., 2016. Toward the development of a big data analytics capability. *Information & Management*, *53*(8), pp.1049-1064.

Hashem, I.A.T., Yaqoob, I., Anuar, N.B., Mokhtar, S., Gani, A. and Khan, S.U., 2015. The rise of “big data” on cloud computing: Review and open research issues. Information systems, 47, pp.98-115.

Henke, N., Bughin, J., Chui, M., Manyika, J., Saleh, T., Wiseman, B. and Sethupathy, G., 2016. The age of analytics: Competing in a data-driven world. *McKinsey Global Institute*, *4*.

Hodder, R.K., Freund, M., Bowman, J., Wolfenden, L., Campbell, E., Wye, P., Hazell, T., Gillham, K. and Wiggers, J., 2012. A cluster randomised trial of a school-based resilience intervention to decrease tobacco, alcohol and illicit drug use in secondary school students: study protocol. *BMC Public Health*, *12*(1), pp.1-8.

Huang, B. and Von Lilienfeld, O.A., 2016. Communication: Understanding molecular representations in machine learning: The role of uniqueness and target similarity.

Jacobson, E., Thomas, L. and Zhang, G., Apple Inc, 2014. Integrated repository of structured and unstructured data. U.S. Patent 8,914,414.

Jagadish, H.V., Gehrke, J., Labrinidis, A., Papakonstantinou, Y., Patel, J.M., Ramakrishnan, R. and Shahabi, C., 2014. Big data and its technical challenges. *Communications of the ACM*, *57*(7), pp.86-94.

Jaseena, K.U. and David, J.M., 2014. Issues, challenges, and solutions: big data mining. *CS & IT-CSCP*, *4*(13), pp.131-140.

Jeble, S., Kumari, S. and Patil, Y., 2016. Role of big data and predictive analytics. *International Journal of Automation and Logistics*, *2*(4), pp.307-331.

Johnson RB, Christensen LB., 2014. *Educational Research: Quantitative, Qualitative, And Mixed Approaches*. Fifth edition. Sage, Thousand Oaks CA.

Kaisler, S., Armour, F., Espinosa, J.A. and Money, W., 2013, January. Big data: Issues and challenges moving forward. In *2013 46th Hawaii International Conference on System Sciences* (pp. 995-1004). IEEE.

Kandel, S., Paepcke, A., Hellerstein, J. and Heer, J., 2011, May. Wrangler: Interactive visual specification of data transformation scripts. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 3363-3372).

Kibria, M.G., Nguyen, K., Villardi, G.P., Zhao, O., Ishizu, K. and Kojima, F., 2018. Big data analytics, machine learning, and artificial intelligence in next-generation wireless networks. IEEE access, 6, pp.32328-32338.

Kim, K., Lee, D. and Essa, I., 2011, November. Gaussian process regression flow for analysis of motion trajectories. In *2011 International Conference on Computer Vision* (pp. 1164-1171). IEEE.

Kiron, D., Prentice, P.K. and Ferguson, R.B., 2014. The analytics mandate. MIT Sloan management review, p. 1.

Kitchin, R. and McArdle, G., 2016. What makes Big Data, Big Data? Exploring the ontological characteristics of 26 datasets. *Big Data & Society*, *3*(1), p.2053951716631130.

Kitchin, R., 2017. Big data—Hype or revolution. *The SAGE handbook of social media research methods*, pp.27-39.

Kouladouros, K., Gärtner, D., Münch, S., Paul, M. and Schön, M.R., 2015. Recurrent intussusception as initial manifestation of primary intestinal melanoma: Case report and literature review. *World Journal of Gastroenterology: WJG*, *21*(10), p.3114.

Kwon, E.D., Drake, C.G., Scher, H.I., Fizazi, K., Bossi, A., Van den Eertwegh, A.J., Krainer, M., Houede, N., Santos, R., Mahammedi, H. and Ng, S., 2014. Ipilimumab versus placebo after radiotherapy in patients with metastatic castration-resistant prostate cancer that had progressed after docetaxel chemotherapy (CA184-043): a multicentre, randomised, double-blind, phase 3 trial. The lancet oncology, 15(7), pp.700-712.

Lamba, H.S. and Dubey, S.K., 2015, September. Analysis of requirements for big data adoption to maximize IT business value. In 2015 4th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO)(Trends and Future Directions) (pp. 1-6). IEEE.

Larson, E.B., 2013. Building trust in the power of “big data” research to serve the public good. *Jama*, *309*(23), pp.2443-2444.

LaValle, S., Hopkins, M.S., Lesser, E., Shockley, R. and Kruschwitz, N., 2010. Analytics: The new path to value. *MIT sloan management review*, *52*(1), pp.1-25.

LaValle, S., Lesser, E., Shockley, R., Hopkins, M.S. and Kruschwitz, N., 2011. Big data, analytics and the path from insights to value. MIT sloan management review, p. 21.

Lebied, M., 2016. Top 11 Business Intelligence and Analytics Trends for 2017. Business Intelligence. Lechevalier, D., Narayanan, A. and Rachuri, S., 2014, October. Towards a domain-specific framework for predictive analytics in manufacturing. In *2014 IEEE International Conference on Big Data (Big Data)* (pp. 987-995). IEEE.

LeCun, Y., Bengio, Y. and Hinton, G., 2015. Deep learning. *nature*, *521*(7553), pp.436-444.

Lewis, S., 2015. Qualitative inquiry and research design: Choosing among five approaches. *Health promotion practice*, *16*(4), pp.473-475.

Loebbecke, C. and Picot, A., 2015. Reflections on societal and business model transformation arising from digitization and big data analytics: A research agenda. The Journal of Strategic Information Systems, 24(3), pp.149-157.

Lomotey, R.K. and Deters, R., 2014, April. Towards knowledge discovery in big data. In *2014 IEEE 8th International Symposium on Service Oriented System Engineering* (pp. 181-191). IEEE.

Ludlow, A.D., Benítez-Llambay, A., Schaller, M., Theuns, T., Frenk, C.S., Bower, R., Schaye, J., Crain, R.A., Navarro, J.F., Fattahi, A. and Oman, K.A., 2017. Mass-discrepancy acceleration relation: a natural outcome of galaxy formation in cold dark matter halos. Physical Review Letters, 118(16), p.161103.

Ludlow, D., Khan, Z., Soomro, K., Marconcini, M., José, R.S., Malcorps, P., Lemper, M., Perez, J.L. and Metz, A., 2017. From top-down land use planning intelligence to bottom-up stakeholder engagement for smart cities-a case study: DECUMANUS service products. *International Journal of Services Technology and Management*, *23*(5-6), pp.465-493.

Ma, C., Zhang, H.H. and Wang, 2014. Machine learning for big data analytics in plants. XTrends in plant science Journal, pp. 798-808.

Marz, N. and Warren, J., 2014. A new paradigm for Big Data. *Big Data: Principles and best practices of scalable real-time data systems. Shelter Island, NY: Manning Publications*.

Maxwell, J.A. and Chmiel, M., 2014. Notes toward a theory of qualitative data analysis. *The SAGE handbook of qualitative data analysis*, pp.21-34.

Mayer-Schönberger, V. and Cukier, K., 2013. Big data: A revolution that will transform how we live, work, and think. Houghton Mifflin Harcourt.

Mayer-Schönberger, V. and Cukier, K., 2013. *Big data: A revolution that will transform how we live, work, and think*. Houghton Mifflin Harcourt.

Mazzei, M.J. and Noble, D., 2017. Big data dreams: A framework for corporate strategy. Business Horizons, 60(3), pp.405-414.

Mergel, I., Rethemeyer, R.K. and Isett, K., 2016. Big data in public affairs. Public Administration Review, 76(6), pp.928-937.

Mikalef, P., Boura, M., Lekakos, G. and Krogstie, J., 2019. Big data analytics capabilities and innovation: the mediating role of dynamic capabilities and moderating effect of the environment. *British Journal of Management*, *30*(2), pp.272-298.

Mikalef, P., Pappas, I.O., Krogstie, J. and Giannakos, M., 2018. Big data analytics capabilities: a systematic literature review and research agenda.. Journal of Information Systems and e-Business Management, Volume 3, pp. 547-578.

Mills, K.A., 2019. Big Data for Qualitative Research. Routledge.

Minelli, M., Chambers, M. and Dhiraj, A., 2013. *Big data, big analytics: emerging business intelligence and analytic trends for today's businesses* (Vol. 578). John Wiley & Sons.

Moshtaghi, O., Kelley, K.S., Armstrong, W.B., Ghavami, Y., Gu, J. and Djalilian, H.R., 2015. Using Google Glass to solve communication and surgical education challenges in the operating room. *The Laryngoscope*, *125*(10), pp.2295-2297.

Mosyakin, S.L. and Tsymbalyuk, Z.M., 2015. Pollen morphology of the southern African tribe Teedieae, an early-branching lineage of crown Scrophulariaceae. *Willdenowia*, *45*(1), pp.65-75.

Muller, A.E., Skurtveit, S. and Clausen, T., 2016. Many correlates of poor quality of life among substance users entering treatment are not addiction-specific. Health and quality of life outcomes, 14(1), p.39.

Müller, O., Junglas, I., Brocke, J.V. and Debortoli, S., 2016. Utilizing big data analytics for information systems research: challenges, promises and guidelines. European Journal of Information Systems, pp. 289-302.

Munawar, H.S., Qayyum, S., Ullah, F. and Sepasgozar, S., 2020. Big Data and Its Applications in Smart Real Estate and the Disaster Management Life Cycle: A Systematic Analysis. *Big Data and Cognitive Computing*, *4*(2), p.4.

Neal‐Boylan, L., 2012. An exploration and comparison of the worklife experiences of registered nurses and physicians with permanent physical and/or sensory disabilities. *Rehabilitation Nursing*, *37*(1), pp.3-10.

Ohlhorst, F.J., 2012. *Big data analytics: turning big data into big money* (Vol. 65). John Wiley & Sons.

Osanloo, A. and Grant, C., 2016. Understanding, selecting, and integrating a theoretical framework in dissertation research: Creating the blueprint for your “house”. Administrative issues journal: connecting education, practice, and research, 4(2), p.7.

Otto, B., Hüner, K.M. and Österle, H., 2012. Toward a functional reference model for master data quality management. *Information Systems and e-Business Management*, *10*(3), pp.395-425. Pandya, D., Srivastava, A., Doherty, A., Sundareshwar, S., Needham, C., Chaudry, A. and KrishnaIyer, S., 2018, April. Increasing Production Efficiency via Compressor Failure Predictive Analytics Using Machine Learning. In Offshore Technology Conference. Offshore Technology Conference.

Parks, M.R., 2014. Big data in communication research: Its contents and discontents. *Journal of Communication*, *64*(2), pp.355-360.

Rajasekar, S., 2012. The systems approach to curriculum development.

Rennie, R. and Law, J. eds., 2016. A Dictionary of Chemistry. Oxford University Press.

Ribeiro, M.T., Singh, S. and Guestrin, C., 2016, August. " Why should I trust you?" Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 1135-1144).

Richardson, M., Burges, C.J. and Renshaw, E., 2013, October. Mctest: A challenge dataset for the open-domain machine comprehension of text. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing* (pp. 193-203).

Ridder, H.G., 2017. The theory contribution of case study research designs. Business Research, 10(2), pp.281-305.

Ross, J.W., 2017. How to Develop a Great Digital Strategy: To Succeed Today, Companies Need a Unique Value Proposition that Incorporates Digital Technologies in a Way that is Difficult for Competitors to Replicate. MIT Sloan Management Review.

Rossman, G.B. and Rallis, S.F., 2016. *An introduction to qualitative research: Learning in the field*. Sage Publications.

Russom, P., 2011. Big data analytics. *TDWI best practices report, fourth quarter*, *19*(4), pp.1-34.

Salfner, F., Lenk, M. and Malek, M., 2010. A survey of online failure prediction methods. *ACM Computing Surveys (CSUR)*, *42*(3), pp.1-42.

Salmon, J., Harmany, Z., Deledalle, C.A. and Willett, R., 2014. Poisson noise reduction with non-local PCA. *Journal of mathematical imaging and vision*, *48*(2), pp.279-294.

Sekaran, U. and Bougie, R., 2016. *Research methods for business: A skill building approach*. John Wiley & Sons.

Sharma, R., Mithas, S. and Kankanhalli, A., 2014. Transforming decision-making processes: a research agenda for understanding the impact of business analytics on organisations.

Sharma, R., Mithas, S. and Kankanhalli, A., 2014. Transforming decision-making processes: a research agenda for understanding the impact of business analytics on organisations.. European Journal of Information Systems, Volume 23.

Sharpe, C., Wiest, T., Wang, P. and Seepersad, C.C., 2019. A comparative evaluation of supervised machine learning classification techniques for engineering design applications. Journal of Mechanical Design, 141(12).

Shin, S.C., Yang, W.J. and Lee, S.M., Samsung Electronics Co Ltd, 2015. *Domain name system, medium, and method updating server address information*. U.S. Patent 8,964,761.

Sicular, S., 2013. Gartner's big data definition consists of three parts, not to be confused with three" v" s. *Gartner, Inc*, *27*.

Siegel, E., 2013. *Predictive analytics: The power to predict who will click, buy, lie, or die*. John Wiley & Sons.

Slack, K., 2012. Mission impossible?: Adopting a CSR-based business model for extractive industries in developing countries. *Resources Policy*, *37*(2), pp.179-184.

Stake, R.E., 2013. *Multiple case study analysis*. Guilford press.

Sukanya, M. and Biruntha, S., 2012, August. Techniques on text mining. In *2012 IEEE International Conference on Advanced Communication Control and Computing Technologies (ICACCCT)* (pp. 269-271). IEEE.

Sun, Z., 2020. Business Analytics Intelligence: An Emerging Frontier for Innovation and Productivity.

Sun, Z., Sun, L. and Strang, K., 2018. Big data analytics services for enhancing business intelligence. Journal of Computer Information Systems, 58(2), pp.162-169.

Swaffield, S. and Deming, M.E., 2011. Research strategies in landscape architecture: mapping the terrain. *Journal of Landscape Architecture*, *6*(1), pp.34-45.

Thomson, R.E. and Emery, W.J., 2014. Data analysis methods in physical oceanography. Newnes.

Tiwari, S., Wee, H.M. and Daryanto, Y., 2018. Big data analytics in supply chain management between 2010 and 2016: Insights to industries. *Computers & Industrial Engineering*, *115*, pp.319-330.

Tofail, S.A., Koumoulos, E.P., Bandyopadhyay, A., Bose, S., O’Donoghue, L. and Charitidis, C., 2018. Additive manufacturing: scientific and technological challenges, market uptake and opportunities. *Materials today*, *21*(1), pp.22-37.

Tole, A.A., 2013. Big data challenges. *Database systems journal*, *4*(3), pp.31-40.

Turner, S., 2016. What is a refugee camp? Explorations of the limits and effects of the camp. *Journal of Refugee Studies*, *29*(2), pp.139-148.

Vidgen, R., Shaw, S. and Grant, D.B., 2017. Management challenges in creating value from business analytics. *European Journal of Operational Research*, *261*(2), pp.626-639.

Villars, R.L., Olofson, C.W. and Eastwood, M., 2011. Big data: What it is and why you should care. *White paper, IDC*, *14*, pp.1-14.

Wagner-Muns, I.M., Guardiola, I.G., Samaranayke, V.A. and Kayani, W.I., 2017. A functional data analysis approach to traffic volume forecasting. *IEEE Transactions on Intelligent Transportation Systems*, *19*(3), pp.878-888.

Wamba, S.F., Gunasekaran, A., Akter, S., Ren, S.J.F., Dubey, R. and Childe, S.J., 2017. Big data analytics and firm performance: Effects of dynamic capabilities. *Journal of Business Research*, *70*, pp.356-365.

Wang, B., Rhauderwiek, T., Inge, A.K., Xu, H., Yang, T., Huang, Z., Stock, N. and Zou, X., 2018. A Porous Cobalt Tetraphosphonate Metal–Organic Framework: Accurate Structure and Guest Molecule Location Determined by Continuous‐Rotation Electron Diffraction. Chemistry–A European Journal, 24(66), pp.17429-17433.

Wang, H., Xu, Z., Fujita, H. and Liu, S., 2016. Towards felicitous decision making: An overview on challenges and trends of Big Data.. Information Sciences journal, Volume 367, pp. 747-765.

Watson, H.J., 2014. Tutorial: Big data analytics: Concepts, technologies, and applications. Communications of the Association for Information Systems, 34(1), p.65.

Wilcox, A.B., Gallagher, K.D., Boden-Albala, B. and Bakken, S.R., 2012. Research data collection methods: from paper to tablet computers. Medical care, pp.S68-S73.

Wixom, B.H., Yen, B. and Relich, M., 2013. Maximizing Value from Business Analytics. MIS Quarterly Executive.

Yang, C., 2012. Restructuring the export-oriented industrialization in the Pearl River Delta, China: Institutional evolution and emerging tension. Applied Geography, 32(1), pp.143-157.

Yin, R.K., 2013. Validity and generalization in future case study evaluations. *Evaluation*, *19*(3), pp.321-332.

Zainal Shah, N., 2011. *Critical thinking and employability of computer-related graduates: The Malaysian context* (Doctoral dissertation, Dublin City University).

Zhang, J. and Huang, M.L., 2013, December. 5Ws model for big data analysis and visualization. In 2013 IEEE 16th International Conference on Computational Science and Engineering (pp. 1021-1028). IEEE.

Zhong, C., Deng, Y., Hu, W., Qiao, J., Zhang, L. and Zhang, J., 2015. A review of electrolyte materials and compositions for electrochemical supercapacitors. *Chemical Society Reviews*, *44*(21), pp.7484-7539.

Zhou, Y., Wang, J. and Zhang, L., 2016. *Basic theory of fractional differential equations*. World scientific.

**8. Appendices**

Appendix A: Authorization letter

This letter is the authorising document that serves as proof to the recipient to perform a certain action on the company. Company A issued this letter to the researcher to conduct the study.

