A Guide for selecting big data analytics tools in an organization

Wandisa Nyikana
Department of Information Technology,
Cape Peninsula University of Technology
South Africa
NyikanaW@cput.ac.za

Tiko Iyamu
Department of Information Technology
Cape Peninsula University of Technology
South Africa
iyamut@cput.ac.za

Abstract

Selection of appropriate big data analytics (BDA) tools (software) for business purposes is increasingly challenging, which sometimes lead to incompatibility with existing technologies. This becomes prohibitive in attempts to execute some functions or activities in an environment. The objective of this study was to propose a model, which can be used to guide the selection of BDA in an organization. The interpretivist approach was employed. Qualitative data was collected and analyzed using the hermeneutics approach. The analysis focused on examining and gaining better understanding of the strengths and weaknesses of the most common BDA tools. The technical and non-technical factors that influence the selection of BDA were identified. Based on which a solution is proposed in the form of a model. The model is intended to guide selection of most appropriate BDA tools in an organization. The model is intended to increase BDA usefulness towards improving organization's competitiveness.

1. Introduction

Big data Analytics (BDA) tools are increasingly used by different organizations across sectors, ranging from e-commerce, government administration, education to science, technology, and healthcare [1,2]. BDA tools (software) are also used for marketing, insurance, telecommunications, retail and fraud detection [3]. E-commerce organizations use analytics tools for activities such as examining website traffic, purchases, and to determine customers' interests and types and frequencies of transactions [4]. Some institutions of higher learning employ BDA tools to assess and address institutional performance, students' performances, and overall progress, to predict the future in teaching and learning. In healthcare, BDA is used to analyze patients' profiles and patterns of health conditions [5,6]. From marketing perspective, many organizations employ BDA tools to gain deeper

understanding about their customers' behaviors, needs and preferences [7]. In a nutshell, BDA enables and supports organizations towards sustainability and competitiveness [8, 9]. Based on this premise, there is a need to investigate the impact which characteristic (volume, velocity, variety, and veracity) of BDA (descriptive insight, predictive insight, and prescriptive) have on innovation competency in an organization, and collaborative positive insights [10].

Despite the coverage and premise, there are challenges in selecting the most appropriate analytics tools, particularly by organizations in developing countries [11]. This can be attributed to primarily two factors: the practice of the concept is slow [8]; and availability of skilled personnel are scarce. For an organization to gain useful insights and value from big data, it requires the application of the most appropriate BDA tools for the analysis [12], which has been a challenge for many organizations. There is emphasis on the 'most appropriate' because no one tool fits all or can fulfils every requirement. Organizations apply BDA differently as their businesses and environments dictate. Some organizations integrate the tools with other software within various platforms such as the data warehouses [7]. Other organizations apply BDA tools in isolation, to test their potential value [13]. From a different angle, healthcare organizations primarily apply BDA tools to construct patients' unified datasets [14]. This is different from how other organizations employ the tools, where the focus is mainly on business strategy and organizational processes for effectiveness and competitiveness [15]. Irrespective of the focus and approach, BDA tools remain a challenge in many organizations, which perhaps triggered a multilevel analysis approach [11]. This seems to have contribute to solving only the analysis aspect of the challenges. The challenges of selecting most appropriate BDA tools remain.

The challenges persist even though the boundaries and uses of the descriptive, prescriptive, and predictive analytics tools are highlighted or clear to some degree [15]. One of the main challenges is how to filter down

HICSS

to the most important elements of big data for service delivery purposes [16]. Hence it is essential to examine the challenges from the predictive analytics approach perspective. However, the problem is the difficulty in combining datasets from different systems that are not connected, in creating relationships and gaining valuable insights [17]. These challenges inform views that there is no single approach for complementary formation in the application of BDA, which seem to influence selections [18]. Notwithstanding, the choice of BDA approach or tools, there are challenges in their use for service delivery, sustainability, effectiveness, and competitive advantage [19]. This challenge influences some organizations in selecting and using multiple BDA tools for their operations and strategic activities. Complementarily, the use of BDA tools seems to be a probability, or 'trial and error' approach towards fulfilling business needs.

For various reasons, some personnel struggle to complementarily apply analytics tools effectively [20]. This sometimes results in losses of datasets and inaccurate outputs [2]. Another implication is that not all BDA tools are supported by every database software, which influence the goals and objectives of an organization [1]. This therefore requires a comparative analysis, to experimentally assess and ascertain the similarities and differences between BDA tools. This evidently assist organizations and interested persons in selecting tools that are considered most appropriate. BDA tools need to be compared to get a better understanding of their capabilities including strengths and weaknesses [21]. Thus, the similarities and differences of BDA tools require investigation [22]. The comparison of tools provides specific direction about capability, relevance, and applicability within an environment [23].

The challenges identified above triggered the formulation of this research's objective, which is to propose a guide that can be used by organizations in selecting BDA tools. This paper is structured into seven main sections. It begins with introduction, followed by a review of literature, which is divided into three sections. The third and fourth sections present the methodology that is employed in the research and the analysis of the data, respectively. The findings from the analysis are discussed in the fifth section. Validation of the proposed model is discussed in the sixth section, and conclusion is finally drawn.

2. Literature review

Based on the objective, a review of literature related to this study was conducted in three main areas: big data analytics, big data analytics tools, and comparative analysis.

2.1. Big data analytics

There is a continuous increase in datasets of many organizations, which come in different forms, speed, types, and formats [24]. This surge in datasets has led to the advent of big data, defined by its distinctive characteristics: volume, velocity, and variety [25, 26]. Currently, the main categories of the BDA are descriptive analytics, predictive analytics, and prescriptive analytics [22, 27]. Also, value is one of the main characteristics of big data, in that it is the aspect that guide organizations toward improving customers' relationships [28]. Traditionally, the characteristics are often referred to as Vs [29]. Based on the Vs, big data can be considered as diverse and versatile, which means it can be applied to numerous areas, irrespective of simplicity or complexity [15]. Arguably, big data versatile nature propels organizations to make use of the concept for various reasons, such as efficiency, sustainability and competitiveness [11]. The diversity and versality of datasets sometimes make it difficult to employ a single analytics tool. This is like Nakashololo and Iyamu disposition that data heterogeneity poses a challenge for big data analytics [13].

There are different formats, structured, and unstructured datasets, which are associated with the Vs of big data [30]. Structured datasets are organized and stored in ways that they can be retrieved and manipulated by machine [5]. While on the other hand, unstructured data is unorganized datasets which come from different sources such as media streams, social networks, e-commerce, and websites. This helps organizations to put the complexity of big data into perspective. In addition, big data helps organizations to enhance its decision-making which increases the organizational output but argued that this happens only when the correct analytics methods are used to extract usefulness from the datasets [31].

As big data increases so does its challenges, which manifest into complexities. Some of the problems that are associated with big data include its storage, visualizations, skills, searching, analyzing, security, and privacy violation [28], which sometimes make appropriateness of analytics challenging. One of the problems regarding big data is in its conflict with infrastructure deployment [31]. Due to the rapid growth of big data, some organizations find it difficult to trust their data sources, which are also argued to influence its analysis [9]. However, finding skilled personnel is a challenge for many organizations [25]. This could be attributed to the newness of the concept, which seems to be a common challenge in many organizations [32]. Integration of datasets is another challenge, finding a unified platform where both structured and unstructured data can coexist for ease of control and manageability

[27]. All these challenges make it difficult for many organizations in their quest to selecting and using big data analytics tools for their business purposes [1].

2.2. Big data analytics tools

Big data analytics can be described as the science that is used to examine raw datasets for specific purposes including drawing conclusions [4, 11, 17]. BDA tools focus on examining and extracting knowledge from big data [1]. In addition, the use of BDA tools includes decision-making [33], which requires formulation of requirements and objectives. Also, BDA tools are used for transforming processes, activities, and services in some organizations, purposely to increase business value [34]. From a different angle, the BDA are tools that focus on enabling organizations to anticipate on future opportunities, based on real-time insights, which are obtained from different data sources [9]. Also, "big data analytics is used in many industries to allow organizations to make better business decisions through descriptive, prescriptive or predictive approaches" [11: p.6].

The descriptive analytics are primarily for analysis of datasets covering current and previous events and processes [31]. The approach allows an organization to learn from past behaviors and practices [7]. The descriptive analytics uses standard and ad-hoc reporting, dashboards, querying and drilling, to draw interferences [35]. Some of its tools, such as Hadoop and Spark are sometimes mistaken for predictive, due to lack of technical know-how [19, 22]. Some of the commonly used descriptive analytics tools are in the areas of data visualization and business intelligence (BI). One of the challenges with the descriptive analytics is that they work backwards and are only capable of revealing things that have already happened, and do not have capability to find hidden relationships within datasets [7].

The predictive analytics focuses on predicting future behavior and expected trends by using past or historical datasets [32]. Mainly, the objective for adopting the predictive analytics is to allow the organizations to compete, grow, improve satisfaction, learn and act using the analytics [35]. These predictive analytics use a variety of statistical techniques, machine learning and data mining [1]. Examples of the tools include Google's Knowledge Graph, Apple's Siri [36]. Some of the challenges that are faced when using these analytics tools are in the areas of theory, models, and hypotheses [3].

The prescriptive tools combine both the descriptive and predictive models. This sometimes make it difficult for some employees to differential between the prescriptive and descriptive or predictive tools. The tools focus on the current state of an environment, and uses consistent predictions to make informed choices, in providing advice for future purposes [17]. This type of analytics is more concerned with future decisions and opportunities [7]. One of the challenges of the prescriptive analytics is that it uses several methods, which can be complex sometimes [37], giving some employees the tendencies to opt for alternatives that are of ease of use or familiar.

The market is saturated with numerous BDA tools from which organizations can choose. Some of the BDA tools include Tableau, Apache Hadoop, Apache Hive, Memcached, Cloudera, Hue and Splunk [19, 38]. On one hand, some organizations' considerations for selecting analytics tools include functionality, cost, reliability, and complexity [29]. On the other hand, compatibility with existing systems within an organization is another consideration. With such a wide range of requirements from both technical and non-technical (business) units, it is critical to have guidelines for comparing the analytics tools towards selecting the most appropriate for an organization's purposes. Results from comparative analysis of the tools can assist to recognize relevance and context exponential detailed level, for an organization.

2.3. Comparative analysis

A generic view about comparative analysis is first provided before it was put into the context of this study. Comparative analysis systematically creates awareness, capability, performance as well as similarities and differences about methods, which include concepts and frameworks [39]. Also, comparison includes assessment of the methods' main goals, objectives, and purposes [40], which covers both strategic and operational aspects of an entity or tool [41]. Comparative analysis focuses on suitability of techniques in handling different goals and objectives within context [83]. Thus, specific technical or non-technical requirements are required based on the complexity and focus of an environment [39].

The results from comparative analysis of tools help to gain deeper understanding within relevance, significance, and context, historically or empirical [42]. The comparative approach also promotes concept and theory building [42, 43]. Analysis from a comparative viewpoint provides insight into a general and specific causes underlying a phenomenon [44]. As a result, comparative analysis can be performed to examine current theories, improvement, or re-engineering of approaches [43].

Based on the above premise, comparative analysis is often used across different areas to identify and explain factors such as the economic, historical,

political, and sociocultural differences and similarities in a phenomenon [44]. For an example, it is used as a cross national comparison whereby cultural and social comparison is done to understand different societies [45]. In addition, the concept is used to compare the economic systems and processes [46]. Economic, historical, and sociocultural are some of the factors that potentially influence the selection and use of BDA tools [40, 47]. There is a need for alternative approach, which combines BDA tools [11]. But before that happens, the BDA tools should be compared to gain better understanding of their distinctive strengths and weakness of the various tools. From methodological viewpoint, such an approach increases the value of big data through an understanding of why and how datasets transform from one point to another [48].

3. Research methodology

Based on the objective of the study, which is to compare and differentiate among BDA tools for organizations' purposes, the qualitative method was selected from the interpretivist viewpoint, because it allows subjective understanding of humans' interactions and actions. The interpretivist approach focuses on subjective meaning of reality through human and social interactions [49]. Two factors influence the decision to select the method: (1) it helps to gain an understanding about IS phenomena, which are often complex [50], and naturally, it focuses on interpretivism and constructivism approaches to study cases, social situations to reveal the meanings that people associate with their experiences [51]; and (2) it primarily focuses on exploring and obtaining depth of an understanding about phenomenon being studied [52]. These rationales helped in this study, to explore and gain broader understanding of the BDA tools, to carry out comprehensive comparison in accordance with strengths and weaknesses.

From the qualitative perspective, the document analysis technique was employed in collecting data for this study. This entails collection and study of related, existing materials or data. The sets of data included empirical studies and published peer-reviewed articles. The approach allows gathering of research articles from databases. In this study, data was gathered from academic databases, which include Google Scholar, Ebscohost and Scopus. These databases host most IS research articles. The focus was on articles published within a historical period of ten years, between 2009 and 2019. Articles published before 2009 were considered. This study was conducted in 2019. The "spread of historical perspectives, is in terms of the consistency of the meaning that has been associated to the concepts, as well as the challenges and confusions that are caused overtime" [53: p.53]. This was to ensure a holistic coverage of the key areas of the study, which include similarities and differences between the BDA tools. The following keywords were used in the search for the articles: big data analytics, descriptive analytics, predictive analytics, and prescriptive analytics. From the 177 articles that were collected, only 57 were directly related to this study, from the perspectives of capability, challenges, and management of the tools, as shown in Table 1. The perspectives were based on the objective of the study.

Table 1. Selection of article

Area	Description	No.
Capability	This group of articles focus on BDA features and the capabilities to enhance organizational functions, goals, and objectives.	29
Challenges	Emphases were more on the challenges in applying the BDA tools, to find solutions in organizations. These include understanding of the obstacles and co-existence of the tools.	11
Management	The articles primarily concentrated on how BDA tools are managed, to validate processes, predict outcomes, and interpret circumstances, in ensuring accessibility, reliability, and timeliness.	17
Total		57

The data was analyzed from the interpretivist perspective, by following the hermeneutics approach, which is concerned with reading, understanding, and interpreting text or linguistic materials [54]. The hermeneutics approach is more interested at developing and understanding of data, which is refers to as the hermeneutic circle. The approach is a circular structure of interpreting and understanding the parts or details that make up the whole phenomenon being studied [55, 56]. The hermeneutic circle as the pillar of interpretivist research as it allows a researcher to use the text to build context [57]. The hermeneutics approach was therefore used to develop an understanding of how BDA tools are similar or different to each other. This was done in circle; repetitive manner as prescribed by the hermeneutic approach: the data was read, puts in perspective, pattern within context. This helps to gain better understanding of the strengths and weaknesses of BDA, by analyzing the technical and non-technical factors that are associated with the tools.

4. A model for selecting big data analytics tools

As explained above, the hermeneutics approach was employed in the analysis of the technical and non-technical aspects of the analytics tools, descriptive, predictive, and prescriptive [58]. Technical artefacts are objects that have been created by humans to fulfil practical functions. Non-technical artefacts include people and processes. Thereafter, the findings and discussion are summarized in Table 2. The analysis focuses on the three most dominant types of analytics tools, descriptive, predictive, and prescriptive, to gain better understanding about their specific strengths and weaknesses for organizations' benefits. From the analysis of the data, 14 factors were identified to influence the usefulness of BDA tools in organizations. Based on the objective, the factors were subjectively identified from the 3 categories shown in Table 1, as follows: Capability: Correlation, Forecasting, Innovation, Visualization, and Reporting; (2) Challenges: Root cause, Simulation, Real-time, and Filtering; and (3) Management: Historical data, Optimization, Transparency, Decision support, and Planning. The factors are presented in Table 2 and discussed below. The discussion focuses on weaknesses and strengths.

The strengths and weaknesses of the BDA tools are indicated with 'X' (for strength) and 'O' (for weakness) in Table 2. The factors of strengths and weaknesses were derived from comparing the BDA tools. The functionalities of the BDA tools reflect through the highlighted factors. This was done to show the capabilities of each BDA tool, towards illustrating expectation from both technical and non-technical viewpoints. The BDA tools have strengths to integrate information from historical and real-time data to forecast trends that can assist organizations to identify opportunities and threats. The tools have strengths to provide visuals and reports that are used to share the organizational information in a transparent manner and assist with planning and decision making. Optimization and innovations are the other strengths of analytics tools towards bringing value to the organization.

Table 2. Strengths and weaknesses

#	Tools / Functions	Descriptive	Predictive	Prescriptive
1	Correlation	O	X	O
2	Forecasting	O	X	O
3	Historical data	X	O	O
4	Optimization	O	O	X
5	Real-time	X	O	O
6	Root cause	X	O	O
7	Simulation	O	O	X
8	Visualization	X	O	O
9	Innovation	O	O	X
10	Reporting	X	X	O
11	Transparency	X	O	O
12	Decision support	X	X	O
13	Planning	O	X	O
14	Filtering	X	O	O

Of the 14 identified factors, descriptive tools are stronger in most of the functions followed by predictive tools. Prescriptive tools are weaker in most of the functions listed below with the strengths is in only innovation, simulation, and optimization. Another example which demonstrates that the BDA tools are distinctive in their functions, strengths, and weaknesses, is shown in [10], which revealed that data velocity, variety, and veracity enhance data-driven insight generation, while data volume have no impact. This table shows that the BDA tools have different capabilities hence it is crucial for an organization to be aware of the strengths and weaknesses, in their decisions to select and adopt any of each tool.

5. Interrelationship between the factors

Based on Table 2, Figure 1 was developed, to illustrate the relationship and dependency between technical and non-technical factors, which together influence BDA tools' weaknesses and strengths. As shown in Figure 1, there are 9 and 5 technical and non-technical factors, respectively. The technical strengths of the functions are correlation, forecasting, real-time, root cause, visualization, simulation, optimization, transparency, and filtering. The non-technical strengths are planning, decision making, reporting, historical data, and innovation. From Table 2, the strength of BDA tools relies more on its technical than non-technical factors. The figure demonstrates the inevitable relationship between the technical and non-technical factors, which consequently influence the selection, deployment, use, and management of BDA tools within an environment. Together, both technical and non-technical factors are determinant, inseparable in influencing the weaknesses or strengths of the BDA tools.
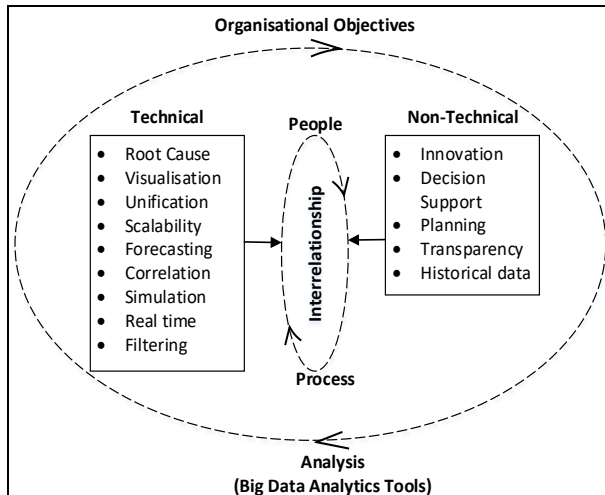
Figure 1.Interrelationship between Technical and non-technical factors

Figure 1 shows the relationship between technical and non-technical factors. As shown in the figure, the relationship is based on process, and driven by people. The diagram clearly indicates that neither the technical nor non-technical factors work in isolation in the application of BDA for organization's purposes. Employees apply the analytics tools to analyze the organizational data, for reasons such as sustainability, competitiveness, process re-engineering, value. Purposely, people drive the use of analytics tools in an organization to achieve business goals [7]. Another significant factor is that Figure 1 reflects the importance of aligning BDA tools with the business processes to achieve the business objectives. Essentially, for the big data analytics tools to add new insights for the organization, they need to be linked to organizational strategy and be embedded with business processes and events [20].

6. Discussion of the findings

The findings from the analysis as presented above is categorized into approach (descriptive, predictive, and prescriptive), and discussed:

6.1. Descriptive analytics

Descriptive analytics are tools that help organizations to understand what happened (past incidents or events) and what is happening (current state) in their environments [31]. The descriptive analytics enact the ability to discover market opportunities and understand the financial strengths of a business [2, 59, 60]. Some of the strengths of the analytics tools are root cause; visualization, current state; and unification:

Root cause: In descriptive analytics, dashboards are created and used to monitor the performance of an organization over time [31]. This enables an organization to trace root cause of incidents and episodes, toward enhancing competitiveness. An understanding of root causes to failures and challenges enables an organization to make better decisions and avoid repetition of mistakes [1]. In addition, identification of root causes is used to improve the quality of products and services, which help to reduce costs and risks that often leads to failures and challenges [61]. This can be associated with its ability to allow datasets to be viewed in a consistent manner, which provides a complete visibility and transparency of business processes [9, 1].

Visualizations: Visualizations are employed in descriptive analytics, to demonstrate the criticality, significance, and essentiality of datasets from its complexity to simplicity. visualization provide useful insights and knowledge that are sometimes hidden in the large and complex datasets, from different types and sources [62]. Another important strength of visualizations come from demonstrating complex datasets in a simplest form by using pictures, graphs, and diagrams [63, 35]. Thus, visualizations improve stakeholders' understanding of datasets as they relate to their business situation. Visualizations provide real-time datasets that can be used to promote quick decision making [64].

Filtering: The use of descriptive analytics enables organizations to discover their current business situation through reports and alerts that are produced by the tools [65, 31]. The tools are strong at filtering through unstructured datasets and organize them into a form that improve accessibility [64]. Also, the descriptive tools have the capability to filter through and find relationships between datasets, which is often impossible to other analytics approaches [66]. In addition to the strengths of the tools, they are potent in summarizing datasets in a more meaningful and easily understandable manner [48].

Despite the benefit from the strengths as mentioned and discussed above, the descriptive analytics tools pose challenges through its' weaknesses. Two of the main weaknesses are: lack of responsiveness; and lack of scalability:

Real-time: The descriptive analytics tools are not usually responsive at real-time. This means that the tools can only identify and highlight business failures and losses that have already happened. The tools focus on describing and tracing the events and activities backwards hence they are considered reactive rather proactive [7, 31].

Descriptive tools can be challenging in that they are associated with poor performance related lack of real-

time response. As the accumulation of the big data increases, the systems response time become slow. Also, the tools struggle with scalability based on the increasing size, variety, and veracity of the big data. This affects dimensions of the datasets that is analyzed for organizational purposes.

6.2. Predictive analytics

Predictive analytics are used to predict possible future occurrences, through data that is currently available [59]. The primary strengths of the predictive analytics tools are its futuristic, forecasting and correlation:

Planning: Planning is a strength of the predictive analytics, used to guide prediction of activities that are likely to happen in the future, as well as the problematic conclusion of the unknown events. In the context of planning, predictive analytics tools can be used to make business decisions independently without humans' involvement in the interpretation of activities and events. The evolution of the predictive analytics tools provides organizations with high level of information transparency for different kinds of insights and transformative purposes, which are associated with machine learning [67]. This type of transformation makes case for the development of machine learning algorithms that can handle huge datasets, which the existing algorithms were not designed to handle [22]. The cognitive technologies such as machine learning can soon shift business operations towards embedded business processes, automation, and robotics [1].

Increasingly, many organizations rely on the predictive tools to enhance their products and services as well as to gain competitive advantage [59]. In the context of health care, it depends on the predictive tools to map their interventions in improving patients' outcomes. The challenge is that it is highly likely that there would be shortage of skilled personnel in area of predictive analytics. Shortage of skilled personnel result in significant pay rise for data analysts and scientist, as well as increase in training of persons, if an organization is to stay competitive using predictive approach [67].

Forecasting: The predictive tools focus on forecasting of trends, to identify threats and opportunities [68]. The predictive tools are often used by e-commerce businesses to forecast future buying behaviors of customers [48]. From another viewpoint, many e-commerce organizations employ the predictive analytics tools to forecast new products for opportunities purposes, and to manage products fluctuating prices [32].

Through forecasting, the predictive analytics tools are used to defect failure deficiency, which help to improve the quality of the products [62]. The analytics

tools are used in various industries, such as education, health, and engineering. Some institutions of higher learning employ predictive analytics tools to predict students' performance based on current academic results, as well as course enrolment [4, 69]. The predictive analytics tools are used in the health sector to predict patients' health conditions based on their history as well as current health status, lifestyle and behaviors [1]. The tools are used to forecast the threats such as those related to jet engine failures based on the alerts from sensors [48]. In manufacturing industry, these tools can forecast failures of equipment and provide transparency of manufacturing process and quality of product.

Correlation: The predictive analytics have an ability to integrate information from different data sources [14] and create relationships within the data [59]. The tools are used in the various industries such as healthcare, retail, and financial institutions for correlation of activities. In healthcare, the predictive tools are used to integrate clinical information such as medical records, physician notes and laboratory results from different healthcare units, to improve quality of care. In retail industry, decisions related to product orders, product prices and design of promotions can be established by correlating customers' demographics, the time they spend in the store, and the areas they visit [48]. Financial institutions correlate information such as income flow of borrowers and past financial behaviors of the borrower, to detect the credit risks and possibilities of borrower defaulting payments [70].

The predictive analytics tools are widely used by different industries and organizations to forecast future events and happenings. However, the tools do have weaknesses such as high maintenance, error handling, handwriting recognition, lack of skills, and slow innovation:

Decision support: The predictive analytics tools can be used as an approach for decision support, to continuously maintained and monitored business processes change in a complex environment. The cost of maintenance is expensive, which must be considered when selecting analytics tool. When these tools are not continuously maintained they can make project irrelevance and predictions become unreliable [71].

Without decision support mechanism, the predictive tools often struggle with error handling when large number of parameters are applied [48]. Some errors are encountered due to variables being correlated [14]. This happens when large number of variables are evaluated against processes, events, and activities, and in the process some of them become unrelated to the environmental context.

Reporting: In reporting, the predictive tools do not have the capability to recognize and transform human

handwriting into text [72]. This challenge exists because there were no specific algorithms that can address it, as at the time of this study. This is a technical deficiency which many employers are not aware of.

Thus, the predictive tools require soft skills such as analytical and creativity know-how to interpret the insights that are projected by these tools [71]. This is to close the gaps during reporting. However, this type of skills is scarce across industries. Technical skills for activities such as data mining are also a challenge since they are needed to apply the predictive tools [1]. BDA can be used to create new knowledge and new potentialities, which is not simply a matter of capabilities, but crucially need skills [8].

Historical data: The predictive tools lack theoretical basis as they cannot translate historical data sets using subjective viewpoint, which have negative effect on innovation [48, 3]. Also, the analytics tools don't allow the decision-makers to discover and learn things themselves, which leads to poor creativity and slow innovation within an environment where historical data is relied upon [60].

6.3. Prescriptive analytics

The prescriptive analytics tools are crucial for many organizations because the tools are not only used to suggest recommendations, but also to create rules [62]. The prescriptive analytics tools' cruciality primarily come from simulation, innovation, and optimization.

Simulation: The prescriptive tools are used to develop simulation models for the purposes of lowering prediction of business risks and introduction of new solutions [71]. The simulation models imitate business processes' behaviors and predict outcomes [62]. The process of introducing a new solution involves identification of other possible alternatives and using different scenarios to achieve the desired business goals and objectives [73]. The knowledge and opinions of experts are crucial as decisions they make could lead to introduction of new solutions [59].

Innovation: The prescriptive tools examine the business processes and operations to discover insights. These insights are used to innovate business process models to create value and sustainability for the business and maximize profits [64]. One of the strengths of the BDA is in its innovative capability to enable transformation within an organization [19].

Optimization: The prescriptive tools incorporate results from predictive tools and thoroughly investigates the cause-and-effect relationship between the results and business processes [31]. The findings are used to regenerate the prescriptions and optimization of business models [64]. Some organizations enhance their performances by optimizing their business processes

[59]. Prescriptive tools provide innovative and optimization opportunities for organizations, towards sustainability and competitiveness. Despite these benefits, challenges are associated with the prescriptive analytics tools. The most common of the challenges are complexity and implementation limitation:

Transparency: Prescriptive tools allow the use of several methods towards finding solutions, which makes its application and process complex [37]. Many organizations have complex processes because of continuous changes to the algorithms of the predictive analytics tools [74]. The tools can only provide transparency to complex business problems due to the database constraints caused by the number of dimensions the database can record accurately [60].

The prescriptive tools are unable to automatically implement solutions or recommendations without the intervention of human effort [74]. Humans are expected to employ BDA tools in implementing solutions, which they can skillfully interpret the prescribed solutions, otherwise, the recommendations become void or useless [61]. As discussed above, the BDA tools have different focuses, impacts, capacities, and value creations. The descriptive and predictive approaches can be used to improve innovation competency, unlike the prescriptive approach [10]. The BDA tools are the BDA tools pose issues of capability and require continuous planning for decision-making through transparency[8].

Wherever Times is specified, Times Roman or Times New Roman may be used. If neither is available on your word processor, please use the font closest in appearance to Times. Avoid using bit-mapped fonts if possible. True-Type 1 fonts are preferred.

7. Validation of the framework

In the areas of big data analytics, frameworks have been developed and implemented in recent years. This includes [75] for smart cities, [76] for peer-to-peer botnet detection, and [77] for data management. None of these frameworks focus on selection of BDA. This makes it difficult or impossible to validate the model proposed in this study.

What helps is that IS frameworks or models can be validated through different ways, such as, case descriptions, or alignment of components of a framework [78]. Another significant aspect is the flexibility associated with validation of the framework or model, from research viewpoint. It is from this angle that validation of framework or model can be objective or subjective. The subjective approach was on that basis employed in this validation. The primary factors of the proposed model (Figure 1), which are technical, people, and non-technical were therefore used as the

components for validation, against similar components of other frameworks or models.

Technical factors consist of technology solutions. The significance of technical factors on integration can be based the 'backbone' of availability and use of the BDA in an organization [79]. Availability of big data is critical for the purpose of achieving the aim and objectives of analytics. This is influenced by the interaction that happens between technical and non-technical factors in the use of a BDA tools for an organization's purposes. It is therefore a primary criterion to understand how the BDA tools facilitate, enable, and support such activity (interaction). The peer-to-peer approach was ironically used to demonstrate the role of people in BDA tools [76]. The non-technical factors drive requirements in the selection of BDA tools in an organization. Framework explains the essentialities of non-technical factors in the use of BDA tools for decision-making and achieving smart solutions. Based on this validation, the proposed model can be considered suitable for guiding selection of BDA tools in an organization.

8. Conclusion

Through Table 2 and Figure 1, the study proposes a solution, which can be used to guide organizations in selecting BDA tools for business purposes. The study can be used by both small and large organizations by customizing the outcome. For this reason, the study can be of interest and benefit to both business and academic domains. To the business, IT specialists, data architects, individuals, and organizations who have particular interest in the concept of BDA. The interest can be viewed from the angle of roles and responsibilities. Primarily, the elements that distinguish data analysts or architects is the ability to identify and select appropriate analytics tools, which can be used to leverage business activities for organizational benefits. It can also be used to guide IT managers in defining requirements for support and enabling infrastructures. From academic perspective, the study can be used as lecture and case studies' material. In addition, it adds to existing literature, which is a contribution to the academic's domain.

Even though this study is comprehensive because it completely achieved its objectives, there are room for further studies. The study lays foundation for comparative analysis of tools (software) in each areas of big data analytics. Such studies can focus on exerting the depth of functionalities for the benefits of organizations that deploy the tools. Also, a study can be carried out to address complementary use of big data analytics tools in an organization.

9. References

[1]. Vassakis, K., Petrakis, E. & Kopanakis, I. (2018). Big Data Analytics: Applications, prospects and challenges. In *Mobile Big Data*. Springer, Cham: 2-20.

[2]. Wang, Y., Kung, L., Wang, W. Y. C., & Cegielski, C. G. (2018). An integrated big data analytics-enabled transformation model: Application to health care. *Information & Management*, *55*(1), 64-79.

[3]. Grable, J.E.& Lyons, A.C. (2018). An introduction to big data. *Economics & Investment Management,* 72(5), 17-20.

[4]. Picciano, A.G. (2012). The Evolution of big data and learning analytics in *American Higher Education. Journal of Asynchronous Learning Networks*, 16(3), 9-20.

[5]. Raghupathi, W. & Raghupathi, V. (2014). Big data analytics in healthcare: Promise and potential. *Health Information Science and Systems*, 2(3), 1-10.

[6]. Sagiroglu, S. & Sinanc, D. (2013). Big data: A review. In *International Conference on Collaboration Technologies and Systems.* San Diego, CA: 42-47. IEEE.

[7]. Watson, H.J. (2014). Tutorial: Big data analytics: Concepts, technologies, and applications. *Communications of the Association for Information Systems*, 34(65), 1247-1268.

[8]. Kuoppakangas, P., Kinder, T., Stenvall, J., Laitinen, I., Ruuskanen, O. P., & Rannisto, P. H. (2019). Examining the Core Dilemmas Hindering Big Data-related Transformations in Public-Sector Organizations. *NISPAcee Journal of Public Administration and Policy*, *12*(2), 131-156.

[9]. Grover, V., Chiang, R.H.L., Ting-Peng, L. & Dongsong, Z. (2018). Creating strategic business value from big data analytics: A research framework. *Journal of Management Information Systems*, 35(2), 388-423.

[10]. Ghasemaghaei, M., & Calic, G. (2019). Does big data enhance firm innovation competency? The mediating role of data-driven insights. *Journal of Business Research*, *104*, 69-84.

[11]. Iyamu, T. (2018). A multilevel approach to big data analysis using analytic tools and actor network theory. *South African Journal of Information Management*, 20(1), a914.

[12]. Al-htaybat, K. & Alberti-alhtaybat, L. V. (2017). Big data and corporate reporting: Impacts and paradoxes. *Accounting, Auditing & Accountability Journal*, 30(4), 850-873.

[13]. Nakashololo, T. & Iyamu, T. (2019). Understanding big data analytics and the Interpretive approach for analysis purposes. *Journal of Contemporary Management,* 16(1), 272-289.

[14]. Hernandez, I. & Zhang, Y. (2017). Using predictive analytics and big data to optimise pharmaceutical outcomes. *American Journal of Health-System Pharmacy,*74(18), 1494-1500.

[15]. Sharma, S. (2015). *An extended classification and comparison of NoSQL big data models*. arXiv preprint arXiv:1509.08035

[16]. Katal, A., Wazid, M. & Goudar, R. H. (2013). August.

Big data: Issues, challenges, tools and good practices. In *Sixth international conference on contemporary computing (IC3)*: 404-409. IEEE.

[17]. Daniel, B. (2015). Big data and analytics in higher education: Opportunities and challenges. *British Journal of Educational Technology: Journal of the Council for Educational Technology,* 46(5), 904 -920.

[18]. Najafabadi, M.M., Villanustre, F., Khoshgoftaar, T.M., Seliya, N., Wald, R. & Muharemagic, E. (2015). Deep learning applications and challenges in big data analytics. *Journal of Big Data,* 2(1), 2-21.

[19]. Wang, Y., Kung, L. & Byrd, T. A. (2018). Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations. *Technological Forecasting and Social Change*, 126, 3-13.

[20]. Lavalle, S., Lesser, E., Shockley, R., Hopkins, M.S. & Kruschwitz, N. (2011). Big data, analytics and the path from insights to value. *MIT Sloan management review*, 52(2), 21-32

[21]. Cheng, Y., Qin, C. & Rusu, F. (2012). Glade: big data analytics made easy. In *Proceedings of the ACM SIGMOD International Conference on Management of Data.* 697-700.

[22]. Yaqoob, I., Hashem, I. A. T., Gani, A., Mokhtar, S., Ahmed, E., Anuar, N. B. & Vasilakos, A. V. (2016). Big data: From beginning to future. *International Journal of Information Management*, 36(6), 1231-1247.

[23]. Tsai, C.W., Lai, C.F., Chao, H.C. & Vasilakos, A.V. (2015). Big data analytics: a survey. *Journal of Bid Data,* 2(1), 21.

[24]. Daniel, B. (2018). Reimaging research methodology as data science. *Big Data and Cognitive and Cognitive Computing,* 2(1), 4.

[25]. Fredriksson, C., Mubarak, F., Tuohimaa, M. & Zhan, M. (2017). Big data in the public sector: A systematic literature review. *Scandinavian Journal of Public Administration*, 21(3), 39-61.

[26]. Hofacker, C.F., Malthouse, E.C. & Sultan, F. (2016). Big data and consumer behaviour: imminent opportunities. *Journal of Consumer Marketing*, 33(2), 89-97.

[27]. Abbasi, A., Suprateek, S. & Chiang, R.H.L. (2016). Big data research in information systems: toward an inclusive research agenda. *Journal of the Association for Information Systems,* 17(2), i-xxxii.

[28]. Bariki, L., Arvind, T. & Hari, S.R. (2017). Big data analytics: Challenges, research issues, tools and application survey. *International Journal of Engineering Sciences & Research Technology*, 6(11), 548-554.

[29]. Acharjya, D.P. & Kauser, A.P. (2016). A Survey on big data analytics: Challenges, open research issues and tools. *International Journal of Advanced Computer Science and Applications*, 7(2), 511-518.

[30]. Tonidandel, S., King, E.B. & Cortina, J.M. (2018). Big data methods: Leveraging Modern Data Analytic Techniques to Build Organizational Science. *Organizational Research Methods*, 21(3), 525-547.

[31]. Sivarajah, U., Kamal, M.M., Irani, Z. & Weerakkody, V. (2017). Critical analysis of big data challenges and analytical methods. *Journal of Business Research*, 70, 263-286.

[32]. Zakir, J., Seymour, T. & Berg, K. (2015). Big data analytics. *Issues in Information Systems*, 16(2), 81-90.

[33]. Sahu, S.K, Jacintha, M.M. & Singh, P. (2017). Comparative study of tools for big data analytics: An analytical study. In *International Conference on Computing, Communication and Automation (ICCCCA)*. Greater Nodia, India: 37-41. IEEE.

[34]. Bonthu, S. & Bindu, K.H. (2018). Review of leading data analytics tools. *International Journal of Engineering & Technology*, 7(3.31), 10-15.

[35]. Saggi, K.M. & Jain, S. (2018). A survey towards an integration of big data analytics to big insights for value-creation. *Information Processing and Management*, 54(5), 758-790.

[36]. Kumar, A., Niu, F. & RE, C. (2013). Hazy: Making it easier to build and maintain big-data analytics. *Communications of the Association for Computing Machinery,* 56(3): 40-49.

[37]. Basu, A.T.A.N.U. (2013). Five pillars of prescriptive analytics success. *Analytics magazine*, 812.

[38]. Chang, B., Tsai, H., Tsai, C., Kuo, C. & Chen, C. (2016). Integration and optimization of multiple big data processing platforms. *Engineering Computations*, 33(6), 1680-1704.

[39]. Zdravkovic, J., Stirna, J., & Grabis, J. (2017). A comparative analysis of using the capability notion for congruent business and information systems engineering. *Complex Systems Informatics and Modeling Quarterly*, (10), 1-20.

[40]. Kambatla, K., Kollias, G., Kumar, V. & Grama, A. (2014). Trends in big data analytics. *Journal of Parallel and Distributed Computing*, 74(7), 2561-2573.

[41]. Bai, C., Dhavale, D., & Sarkis, J. (2014). Integrating fuzzy C-Means and TOPSIS for performance evaluation: An application and comparative analysis. *Expert Systems with Applications*, 41(9), 4186-4196.

[42]. Freiberger, O. (2018). Elements of comparative methodology in the study of religion. *Religions,* 9(2), 38.

[43]. Bloemraad, I. (2013). The promise and pitfalls of comparative research design in the study of migration. *Migration Studies,* 1(1), 27-46.

[44]. Von Schnurbein, G., Perez, M. & Gehringer, T. (2018). Nonprofit comparative research: Recent agendas and future trends. *International Journal of Voluntary and Nonprofit Organizations*, 29(3), 437-453.

[45]. Gomez, M.V. & Kuronen, M. (2011). Comparing local strategies and practices: Recollections from two qualitative cross-national research projects. *Qualitative Research,* 11(6), 683-697.

[46]. Jezek, F. (2014). Comparative research in the area of corporate income tax. *Procedia-Social and Behavioural Sciences,* 110, 997-1005.

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

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

[49]. Chen, W., & Hirschheim, R. (2004). A paradigmatic and methodological examination of information systems research from 1991 to 2001. *Information systems journal*, *14*(3), 197-235.

[50]. Tsang, E. W. (2014). Case studies and generalization in information systems research: A critical realist perspective. *The Journal of Strategic Information Systems*, 23(2), 174–186.

[51]. Yilmaz, K. (2013). Comparison of quantitative and qualitative research traditions: Epistemological, theoretical, and methodological differences. *European journal of education*, *48*(2), 311-325.

[52]. Palinkas, L. A., Horwitz, S. M., Green, C. A., Wisdom, J. P., Duan, N., & Hoagwood, K. (2015). Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. *Administration and Policy in Mental Health and Mental Health Services Research*, 42(5), 533–544.

[53]. Iyamu, T., Nehemia-Maletzky, M., & Shaanika, I. (2016). The overlapping nature of business analysis and business architecture: What we need to know, 19(3), 169-179.

[54]. Suominen, V., & Tuomi, P. (2015). Literacies, hermeneutics, and literature. *Library trends*, *63*(3), 615-628.

[55]. Boell, S. K., & Cecez-Kecmanovic, D. (2014). A hermeneutic approach for conducting literature reviews and literature searches. *Communications of the association for information systems*, *34*(12), 257-286.

[56]. Boland, R.J., Newman, M. & Pentland, B.T. (2010). Hermeneutical Exegesis in information systems design and use. *Information and Organization*, 20(1), 1-20.

[57]. Darby, J.L., Fugate, B.S. & Murray, J.B. (2019). Interpretive research: A Complementary approach to seeking knowledge in supply chain management. *International Journal of Logistics Management, 1-37.*

[58]. Neale, J. (2016). Iterative categorization (IC): a systematic technique for analysing qualitative data. *Addiction*, 111(6), 1096-1106.

[59]. Delen, D., & Demirkan, H. (2013). Data, information and analytics as services. Decision support systems, 55(1), 359-363.

[60]. Banerjee, A., Bandyopadhyay, T. & Prachi, A. (2013). Data analytics: Hyped up aspirations or true potential?. *Vikalpa,* 38(4),1-12.

[61]. Lade, P., Ghosh, R. & Srinivasan, S. (2017). Manufacturing analytics and industrial internet of things. *IEEE Intelligent Systems*, 32(3), 74-79.

[62]. Venkatram, K. & Geetha, M.A. (2017). Review on big data & analytics-concepts, philosophy, process and applications. *Cybernetics and Information Technologies*, 17(2), 3-27.

[63]. Nielsen, S. (2015). The impact of business analytics on management accounting, *Available at SSRN 2616363*.

[64]. Rehman, M.H., Chang, V., Batool, A. & creation in sustainable enterprises (2016). *International Journal of Information Management,* 36(6), 917-928.

[65]. Joseph, R.C. & Johnson N.A. (2013). Big data and transformational government. *IT Proffesional*, 15(6), 43-48.

[66]. Boulila, W., Farah, I.R. & Hussain, A. (2018). A novel decision support system for the interpretation of remote sensing big data. *Earth Science Informatics*, 11(1), 31-45.

[67]. Lee, I. (2017). Big data: Dimensions, evolution, impacts, and challenges. *Business Horizons,* 60(3), 293-303.

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

[69]. Rajni, J. & Malaya, B. (2015). Predictive analytics in a higher education context. *IT Proffesional*, 17(4), 24-33.

[70]. Calof, J., Richards, G. & Smith, J. (2015). Foresight, competitive intelligence and business analytics- tools for making industrial programmes more efficient. *Foresight-Russia,* 9(1), 68-81.

[71]. Appelbaum, D., Kogan, A., Vasarhelyi, M. & Yan, Z. (2017). Impact of business analytics and enterprise systems on managerial accounting. *International Journal of Accounting Information Systems,* 25, 29-44.

[72]. Wu, C., Buyya, R. & Ramamohanarao, K. (2016). Big data analytics=machine learning+cloud computing. *axXiv preprint arXiv: 1601.03115*.

[73]. Evans, J.R. & Lindner, C.H. (2012). Business analytics: the next frontier for decision sciences. Decision Line, 43(2), 4-6.

[74]. Lepenioti, K., Bousdekis, A., Apostolou, D. & Mantzas, G. (2018). Prescriptive analytic: A survey of approaches and methods. In *International Conference on Business Information Systems.* Greece :449-460. Springer. Cham

[75]. Osman, A. M. S. (2019). A novel big data analytics framework for smart cities. *Future Generation Computer Systems*, *91*, 620-633.

[76]. Singh, K., Guntuku, S. C., Thakur, A., & Hota, C. (2014). Big data analytics framework for peer-to-peer botnet detection using random forests. *Information Sciences*, *278*, 488-497.

[77]. Fiore, S., Palazzo, C., D'Anca, A., Foster, I., Williams, D. N., & Aloisio, G. (2013). A big data analytics framework for scientific data management. In *International Conference on Big Data* (pp. 1-8). 6-9 October. Silicon Valley, CA, USA. IEEE.

[78]. Mueller, B., Viering, G., Legner, C., & Riempp, G. (2010). Understanding the economic potential of service-oriented architecture. Journal of Management Information Systems, 26(4), 145–180.

[79]. Elgendy, N. & Elragal, A. (2014). Big data analytics: a literature review paper. In *Industrial Conference on Data Mining* (pp. 214-227). Springer International Publishing. 16-20 July, St Petersburg, Russia.