

The Role of Big Data Analytic in Business Innovation

Khushboo Prashant Patil,
MCA (Computer Science)
JSPM University wagholi, Pune

Dr. Anita Pisote,
Assistant Professor
JSPM University wagholi, Pune

Abstract: *In the current business environment, the implementation of Big Data Analytics (BDA) becomes highly critical for innovating and improving performance. Thus, the aim of the paper is to examine the ways how organizations may process both structured and unstructured data to gain some useful insights and make more efficient business decisions. The growing amount of data created by social networking sites, Internet-based commercial transactions, and the Internet of Things makes the use of analytics highly crucial for businesses. Additionally, modern analytics allows firms to understand customer needs, optimize operations and launch new products or services thanks to sophisticated algorithms, which include machine learning and predictive analytics. Nevertheless, the paper will also address the need to use adequate methods for proper data processing and visualization. However, the usage of BDA faces several challenges related to data quality, data integration, expensive technology infrastructure, and scarcity of professionals.*

Keywords: big data analytics, business innovation, data-driven decision-making, machine learning, predictive analytics, digital transformation.

I. Introduction

In the constantly evolving world of information technology, data is now considered one of the most valuable commodities in the business world [1]. The tremendous increase in data production through social media, internet transactions, mobile applications, and the Internet of Things (IoT) has created what is termed "Big Data" [2]. big data is a term used to describe large amounts of data collected, structured, semi-structured, or unstructured, which cannot be processed by conventional data analysis methods [3]. To get valuable information from big data, organizations utilize big data analytics (BDA). BDA is a sophisticated analysis of complicated data using modern technology [4].

Big data analytics have transformed the way businesses are run. Through BDA, organizations can understand more about their customers' behavior, as well as market trends and other useful data to help them make effective decisions [5]. Competent firms benefit greatly in competitive markets [6] because of efficient use of data analytics.

Nevertheless, in spite of its significance, the application of big data analytics poses some problems for the organizations that are associated with data integration and data quality as well as with high cost of implementation and lack of experts in this field [7]. Moreover, organizations may experience difficulties in using

available data for innovations due to the absence of the necessary competencies.

In this research, the objective would be to discover the role of Big Data Analytics for business innovations. The application of big data analytics in various areas will be considered so that its influence on innovative business decision-making process can be explored.

II. Problem Statement

While companies have accepted (BDA) as an innovative motorist to the business for some time now, there continues to be numerous

organizations that have can't been suitable to use BDA to its maximum eventuality [5]. Despite the volume of data and the vacuity of significant advancements in logical technologies, companies constantly struggle with rephrasing their findings from analytics into creating innovative products [8]. These hindrances include lack of strategic alignment, lack of logical capability, lack of data integration, and lack of change readiness [6]. also, there's an adding gap between the theoretical eventuality of BDA (i.e., enhanced decision-making capability, process invention, competitive advantage) and the factual perpetration within a variety of business surrounds [9]. The absence of a complete frame or understanding about how BDA produces invention composites the difficulty of espousing the mechanisms, capabilities, and conditions in which BDA produces sustainable business inventions [10]. therefore, it's important for associations to understand the degree to which BDA can be responsibly employed in the current data-driven frugality [11].

III. Literature Survey

Expanding beyond earlier research, Wamba et al. [2017] and Mikalef et al. [2020] examine how BDA capability impacts business performance, focusing on a firm's dynamic capabilities to adapt and innovate in response to changes in the external environment [10]. They found companies that invested in robust analytic infrastructures were more likely to develop new products, services, or processes than firms that did not [6]. In addition, George et al. [2014] and Agarwal and Dhar [2014] included BDA in the larger framework of strategic management and organizational learning to create a context for thinking about BDA as an organizational rather than purely technological resource that is contingent on establishing the type of culture that encourages the use of data analytics and creating the skills to carry out analytic functions [12].

Companies like Deloitte, Inc., Mckinsey & Co., and Ibm have stated that there are practical instances where practical instances take place when companies widen the meaning of Big Data Analytics to include big data analytics based innovations into their definition of big data analytics. These practical instances include: assistance in personalization; assistance in predictive maintenance; assistance in segmenting market by customer; and assistance in formulating new business models using BDA [8]. Instances which prove the capacity to make fast decision-making, innovative problem-solving, and responsiveness to customers' demands using BDA have been cited from actual cases like logistic optimization by ups and content

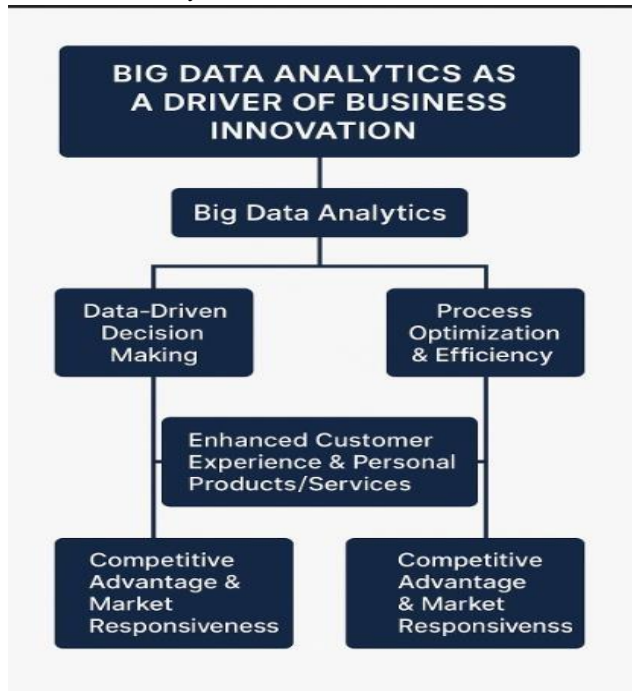


Fig.1 Big Data Analytics as a Driver of Business Innovation

IV. Limitations

A. The Problem of Data Quality and Integration

Most companies struggle when working with unstructured, incomplete, or unreliable data since they are wary of the outcomes that result from processing such information [7].

B. The Problem with Strategic Alignment

Many enterprises fail to see business intelligence (BI) projects as being aligned with their strategic goals, limiting the scope of how data analytics could help create innovative solutions [6].

C. The High Cost of Implementation

The high cost associated with purchasing hardware and software, as well as the cost of employing competent experts, often becomes an obstacle for firms seeking to capitalize on big data analytics (BDA) [13].

D. Scarcity of Skilled Workforce

There is a substantial mismatch between the supply and demand for talent in data science and analytics, making it challenging for companies to utilize and implement BDA solutions successfully [14].

E. Tech Issues

The integration of legacy IT infrastructure with new BDA solutions proves challenging, while the slow adoption of technology in most organizations complicates matters [15].

F. Privacy And Ethics Issues

Big data poses some significant issues relating to customer privacy and ethics in securing informed consent when collecting individuals' private information [16].

V. Research Methodology

The research paper highlights how the operate of big data impacts the management procedure using a case study and interviews with business leaders [10]. The emphasis is put on the importance of the role played by big data in innovation processes in organizations [9]. The researchers undertook an empirical analysis by using companies to understand how big data affects their performance [6]. The Structural Equation Model (SEM) was used to examine the influence of these variables [12]. Previous research work has been

done on this topic; however, it was mostly qualitative research. To fill this gap, the authors conducted a review of literature and formulated a theory. After that, they conducted an extensive literature review concerning the current scenario regarding innovations in the field of Big Data [4].

Methodologies in the Role of Big Data Analytics in Business Innovation

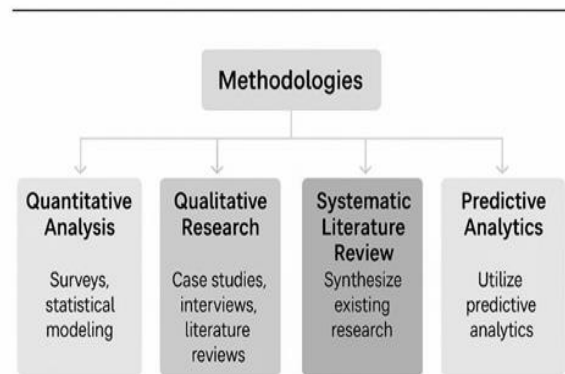


Fig.2. Methodologies

A. Three Types of Analytics:

I: Descriptive Analytic : Descriptive analytics is one of the most basic forms of data analysis that includes summarizing and describing data from previous periods in order to know what happened [17]. In other words, it is like looking into the past through a rear-view mirror. Some examples of descriptive analytics may include sales report from the previous quarter, web traffic from the previous month, or the average age of a customer, and more. Descriptive analytics involves the make use of data aggregation, data mining, statistical analysis, or data visualization including the use of charts, graphs, and dashboards [4].

II: Diagnostic Analytics (Why did it happen?): As the name implies, diagnostic analytics is the level of data analysis that goes beyond just understanding what happened in the past but tries to explain why it happened [17]. This involves using sophisticated methods and techniques to analyze trends from the past and any anomalies. Some questions that could be answered using diagnostic analytics are as follows: 'Why did sales decline last month?', 'Why do we see an increased rate of web traffic bouncing?', and 'Why have there been an increased number of complains about a certain product?', for example.

III: Predictive Analytics:

Predictive analytics uses the past information provided by data to predict the future [15].

a. Predictive analytics is about answering the following question: "What will happen?"

Example applications include forecasting future sales, customer retention, equipment maintenance, future market trends, and credit risk assessment.

Methods used: machine learning algorithms, statistical approaches like linear regression, time-series forecasting, data mining techniques [18].

IV: Prescriptive Analytics (What should be done?)

Prescriptive analytics is the highest level of analytics that exists [17].

a. It not only predicts but also prescribes activity that should be taken in order to obtain an optimal outcome.

b. Prescriptive analytics includes knowledge gained from descriptive, diagnostic, and predictive analytics.

Examples include optimizing prices for products, making recommendations based on consumer preferences, optimizing delivery routes within a company's supply chain, and scheduling preventive maintenance of machinery.

Approaches used: Optimizing, simulations, decision trees, artificial intelligence, and machine learning algorithms [6].

B. System Architecture:

The system architecture of the proposed study reflects the path taken by data through various source systems (data sources) to further different sources where transactional data/output has been analyzed [17]. The following layers have been developed within the context of conversion of raw data into output [19].

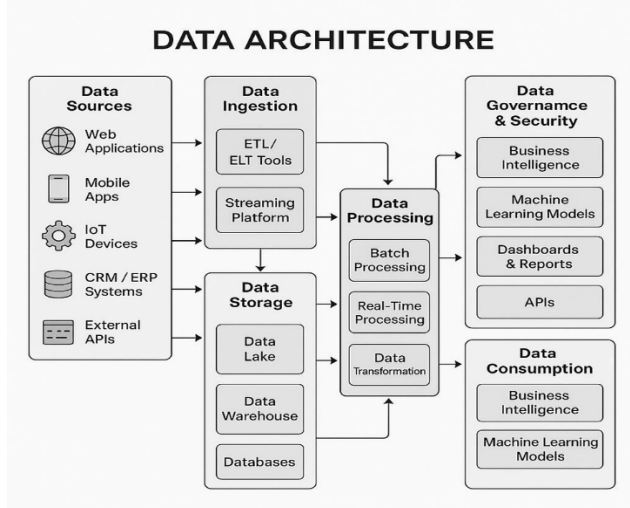


Fig.3. Data Architecture

Layer I: Data Source Layer

This layer is the first part of the system where raw data is created [19]. Data comes from many heterogeneous sources such as:

- Social Media Platforms (Behavioural Data and Feedback from Users)
- Transactional Systems (Sales, Billing, ERP Systems) (making up a real-time data stream)
- Web Logs and Mobile Applications
- The data can be structured (Databases), semi-structured (JSON, XML), or unstructured (text, images, video). The various types of data has become part of the defining features of big_data [3].

Layer II: Data Ingestion Layer

This layer involves ingesting and sourcing data from diverse sources [4]. Data Ingestion Layer provides two main ways for ingesting the data:

- Batch Processing – after some specified interval of time, bulk ingestion of the data is done
- Real-time Streaming – continuous processing of data

Many tools can be used for ingesting the data such as Apache Kafka, Apache Flume, APIs etc.

Essentially, the responsibility of the Data Ingestion Layer lies in capturing the data efficiently and without any loss [15].

Layer III: Data Storage Layer

Once the data ingestion phase is done, then comes the part of storing the data using the efficient storage techniques meant for handling voluminous amount of data [19]. The storage layer will have the following element:

- NoSQL databases such as MongoDB, Cassandra
- The storage layer takes care of scalability, high availability, and flexibility of storing enormous amount of two structured including unstructured data [7].

Layer IV: The Data Processing Layer

Data processing layer objective is for process data that received and transform that be capable of use the data effectively [17]. Functions of the data processing layer include:

- Data cleaning – Elimination of inconsistency and errors in data
- Data transformation – Transformation of raw data into a structured format
- Data integration – Data consolidation from multiple disparate sources into one unified data source

Distributed computing tools like Apache Spark and Hadoop MapReduce framework are used for handling large-scale data

processing [15]. Data processing layer processes the data in preparation for analytics.

Layer V: The Analytics Layer

Data Analytics layer is the key component of any data analytics system. This is where all the important insights are generated [17]. The following types of analytics are included in the Analytics layer:

- Descriptive analytics – Historical data analysis with the aim of explain behavior or function of the business
 - Predictive analytics – using machine learning methods and algorithms for the predicting potential events or trends in the business's future
 - Prescriptive analytics – Prescribing certain actions using historical data and predictive models [4].
- for this layer, firms employ a range of algorithms and machine learning methods like Decision Trees, Random Forest, and Regression analysis [15]. The Analytics layer serves to help firms recognize patterns in customer behavior as well as in the market, or even within customer behavior itself [6].

Layer VI: Visualization & Presentation Layer

Visualization and presentation layer would deliver analyses to the recipients in the format which can be digested, such as dashboards, reports, graphs and charts [17]. Tools like tableau, Power BI and matplotlib are some examples of software used at the visualization layer. These tools help in visualizing data in order to support decision makers in their process [4].

Application / Business Layer

This layer helps in applying the insights generated through data analysis into the business context [5]. This layer provides support to:

- decision making through strategic insights,
- development of products/services with the help of insights generated through data,
- enhancing customer experience by developing insights from the data, and
- increasing operational efficiencies by means of data applications [20].

Insights from above mentioned layers help in gaining a competitive edge for an organization.

Layer VII: Security & Governance Layer (Cross Layer Component)

The security & governance layers of an organization are used to provide protection to all layers in the entire process of data management by providing the following services [16]:

- Data privacy, integrity, and confidentiality
- Access control and authentication services
- Regulatory compliance

The inclusion of cross layer composite components is essential for the ethical and secure use of data; it is important that such components be used in data-driven organizations today [7].

C. Mathematical Modeling for Big Data Analytics in Business Innovation

Mathematical Modelling for Big Data Analytics and Business Innovation

Mathematical modelling gives an approach as to how data turns into information (or rather knowledge) through big data analysis [15]. This allows you to analyze how a change in one variable will affect another and also the relationships between customer behavior, company performance, and innovations generated from each. The mathematical representation of the system is as given below:

A. Model Representation

$$I = f(D, P, A)$$

where,

I = Innovation output

D = Data gathered from several sources

P = Transformation of data

A = Application of analytics tools

The above shows that innovation by businesses depends on the quality of the data that has been collected, its transformation process, and application of analytics tools [6].

B. Data Gathering

$$D = D1 + D2 + D3 + Dn$$

where,

D1 = Social Media data

D2 = Transactional data

D3 = Internet of things data

Dn = other external sources

The equation shows that big data is made up of several sources of data which are heterogenous and independent [3].

C. Data Processing

After gathering the data, it is necessary to clean it up before analyzing it. Hence, $P = D \times \eta$

Where η = Data processing efficiency ($0 < \eta \leq 1$)

The higher the data processing efficiency, the better the value of data for analysis, therefore, the higher the chance of innovative developments through big data analysis systems [4].

D. Predictive Analytics Model

There are some models used in predicting the results in the future through machine learning algorithms that have a common format for representation. In this case, the general representation is:

$$Y = f(X) + \epsilon$$

Where:

Y = Value of the outcome being predicted (sales; demand for products); X = Characteristics of the event (client details; previous events); ϵ = Error rate of the prediction done by the algorithm.

E. Innovation Impact Model

The influence of analytics on innovation can also be represented mathematically through an equation:

$$I = \alpha A + \beta P + \gamma D$$

Where

α, β, γ = The weights allocated by the analyst to show the significance of each parameter

A = Analytics ability

P = Processing efficiency

D = Data quality

In the model above, the innovation impact depends on the collective influence of various parameters compared to only using the data alone [6].

D. Data Collection and Processing:

A. Data Collection:

Data collection refers to the important first step in the whole process of analysis and interpretation of the gathered data for assist in decision making and use for analytics purposes [4]. In regard to innovation within the business environment, the organization must collect its data from the internal and external environment of company in order for get the necessary information regarding customer behavior, market trend identification and the total performance of the company [8]. Organizations can collect their data from many sources including but not limited to Internet applications, mobile applications, Internet of things devices, Customer Relationship Management Systems, Enterprise Resource Planning Systems, and Third Party APIs [3]. The data collected from these sources can take any form ranging from structured to semi-structured data. The main forms of data that can be collected include images, video and textual data. Due to the numerous variations in the nature and amount of data collected from many sources, organizations need efficient data collection technologies [15]. There are two major types of data collection techniques used by organizations, namely batch or real-time data collection.

Batch data collection refers for the procedure of collecting large amounts of data on a regular basis. Real-time data collection, on the other hand, involves collecting data continuously as it becomes available. By leveraging Application Programming Interfaces

(APIs), web scraping software, and real-time data streams, firms can improve their efficiency and effectiveness in collecting data from several different sources [20].

Valid, complete, and accurate data is compulsory to ensure credible analysis. The quality of data influences not only the findings but also any business decisions that are made in the light of such findings [18]. Consequently, data validation is necessary at the point of data collection by firms.

B. Data Processing

Once the data has been collected, it moves on to the data processing stage, where the raw data will be processed or transformed into a meaningful format [17]. The reason why this stage is important is that raw data usually lacks many components and has inconsistencies and/or does not have any structure at all.

The data processing stage entails four critical activities:

- Data Cleaning – detecting and eliminating duplicate or erroneous data entries and ensuring that there are no gaps within the entire set of records in the dataset.
- Data Transformation – transforming the data inside of structured form that are appropriate of analysis.
- Data Integration – consolidating the data that comes from different sources into one set of records.
- Data Reduction – discarding excess records while retaining useful data within the entire record set [4].

The data processing stage can be performed using data analyzing tools like Apache Hadoop and Apache Spark, which are capable of handling vast volumes of data [15]. Previously the data has appeared processed, it will be ready for analysis using either statistical analysis or machine learning approaches [7]. Once the data has been adequately processed, that will provide valuable insights for making strategic decisions [17].

C. Importance of Innovations in the Business Field Organizations that are good at collecting and analyzing data are able remove including the data collected, distinguish patterns and trends, and base their decision-making processes on the data [8]. Innovation is thus facilitated by helping organizations create new products, improve their customers' experience, and optimize their operations [6]. To conclude, data collection and analysis is important for big data and is directly related to how well an organization performs in implementing its business innovation strategy [9].

VI. Result

Big data analytics holds immense promise for delivering tremendous value by improving decision-making, optimizing processes, and strengthening innovation [5].

- Successful firms take advantage of big data for enhance their supply chain management and client experiences and gain competitive advantages [20].
- Big data analytics can help out firms improve responsiveness, agility, and commerce performance leading to excellent market positioning [6].
- Research indicates that data-based firms have been able to boost operational efficiency, client satisfaction, and top line growth. Innovations driven by huge data lead to a shift in the business landscape, enabling new product developments, services, and commerce operations [10].
- The utilize of big data to improve processes will help firms beat their competition regarding profitability and market share [4].
- Organizations that harness innovation using big data will perform much better compared to others based on profit margins and market offerings [9].
- The utilization of big data in decision making processes increases the chances of making data-driven decisions [17].
- Deciphering and responding to decisions made based on large amounts of information is an issue [18].
- In spite of these difficulties, big data analytics will change decision-making within organizations, allowing for efficient

resource management, product development, and consumer interactions [15].

analytics by various industries to achieve maximum benefits from big data [16].

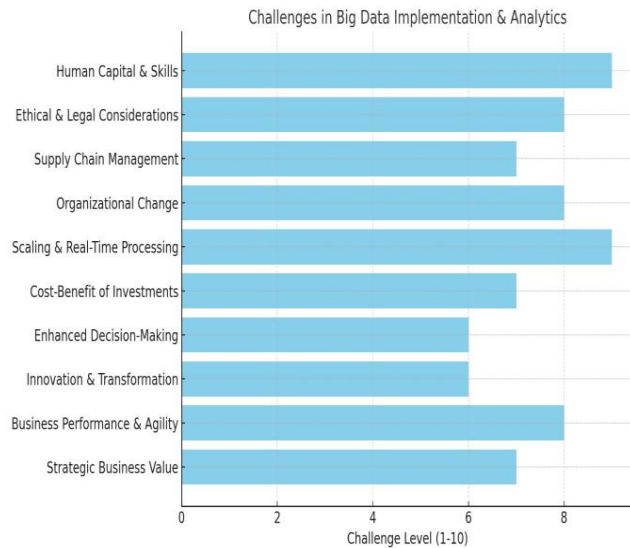


Fig.4. Challenges in Big Data Implementation And Analytics

VII. Conclusion

The results obtained from the ongoing study prove the significance of Big Data Analytics (BDA) in encouraging innovation and achieving better performance among modern organizations operating under the data-rich environment [5]. In other words, updated organizations generate large amounts of data through different sources due to the continuous development of digital technology [8]. Nevertheless, the main problem that can be observed at present within modern organizations with respect to obtaining value from data collection is that value is created not only due to data -collection as well as the analysis of data and turning it into valuable information [4].

BDA allows organizations to analyze customers' behavior, identify emerging market trends, and optimize internal business processes [17]. Furthermore, the application of various advanced analytical approaches, like machine learning and predictive analytics, helps to transition from reactive decision-making [6] to proactive decision-making. Overall, BDA improves organizations' efficiency and supports innovation through facilitating the creation of innovative products, services, and business models.

However, apart from the benefits, there were several problems related with the execution of big data technology highlighted from the research. First, there were challenges with the quality of data being analyzed; second, there were problems with data integration, which is a challenge most organizations encounter while linking up their different systems; third, the amount of executing big data analytics is vast, and finally, a lack of qualified personnel hinders the process [18]. From the results obtained from the study, one can see that an organization must have a good infrastructure and framework to successfully implement BDA [12].

From the information presented in the article, it is evident that companies implementing big data analytics have a competitive advantage compared to other firms [9]. Such companies have an ability to recognize the changes occurring within the business market environment and are able to act on it to provide their clients with additional services, resulting in the overall growth of the business. Big data analytics creates an opportunity for small-medium enterprises (SMEs) operating in Emerging Economies (EEs) such as India [8].

From the above discussion, it can be said that this study shows how Big Data Analytics is not only an emerging technology but a strategic asset as well, and that it will give new ways to innovate in the future. The research should focus on creating frameworks, improving data governance, and the acceptance of big data

VIII. References

- [1] V. Mayer-Schönberger and K. Cukier, "Big Data: A Revolution That Will Transform How We Live, Work, and Think," Houghton Mifflin Harcourt, 2013.
- [2] J. Manyika et al., "Big Data: The Next Frontier for Innovation, Competition, and Productivity," McKinsey Global Institute, 2011.
- [3] S. Akter and S. F. Wamba, "Big Data Analytics in E-Commerce: A Systematic Review and Agenda for Future Research," *Electronic Markets*, vol. 26, no. 2, pp. 173–194, 2016.
- [4] M. Chen, S. Mao, and Y. Liu, "Big Data: A Survey," *Mobile Networks and Applications*, vol. 19, no. 2, pp. 171–209, 2014.
- [5] T. H. Davenport and D. J. Patil, "Data Scientist: The Sexiest Job of the 21st Century," *Harvard Business Review*, vol. 90, no. 10, pp. 70–76, 2012.
- [6] R. Kitchin, "The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences," Sage Publications, 2014.
- [7] S. Sagioglu and D. Sinanc, "Big Data: A Review," *Proc. IEEE Int. Conf. Collaboration Technologies and Systems*, pp. 42–47, 2013.
- [8] H. Chen, R. Chiang, and V. Storey, "Business Intelligence and Analytics: From Big Data to Big Impact," *MIS Quarterly*, vol. 36, no. 4, pp. 1165–1188, 2012.
- [9] S. Gandomi and M. Haider, "Beyond the Hype: Big Data Concepts, Methods, and Analytics," *Int. Journal of Information Management*, vol. 35, no. 2, pp. 137–144, 2015.
- [10] A. McAfee and E. Brynjolfsson, "Big Data: The Management Revolution," *Harvard Business Review*, vol. 90, no. 10, pp. 60–68, 2012.
- [11] J. Bughin, M. Chui, and J. Manyika, "Clouds, Big Data, and Smart Assets: Ten Tech-Enabled Business Trends to Watch," *McKinsey Quarterly*, 2013.
- [12] S. F. Wamba et al., "How 'Big Data' Can Make Big Impact: Findings from a Systematic Review and a Longitudinal Case Study," *Int. Journal of Production Economics*, vol. 165, pp. 234–246, 2015.
- [13] M. George, D. Haas, and A. Pentland, "Big Data and Management," *Academy of Management Journal*, vol. 57, no. 2, pp. 321–326, 2014.
- [14] P. Russom, "Big Data Analytics: TDWI Best Practices Report," *TDWI Research*, 2011.
- [15] S. Yin and O. Kaynak, "Big Data for Modern Industry: Challenges and Trends," *Proceedings of the IEEE*, vol. 103, no. 2, pp. 143–146, 2015.
- [16] J. Fan, F. Han, and H. Liu, "Challenges of Big Data Analysis," *National Science Review*, vol. 1, no. 2, pp. 293–314, 2014.
- [17] D. Laney, "3D Data Management: Controlling Data Volume, Velocity, and Variety," *META Group Research Note*, 2001.
- [18] S. Akter, S. F. Wamba, A. Gunasekaran, R. Dubey, and S. J. Childe, "How to Improve Firm Performance Using Big Data Analytics Capability and Business Strategy Alignment," *Int. Journal of Production Economics*, vol. 182, pp. 113–131, 2016.
- [19] R. Mikalef, I. O. Pappas, J. Krogstie, and P. Pavlou, "Big Data Analytics Capabilities: A Systematic Literature Review and Research Agenda," *Information Systems and e-Business Management*, vol. 16, no. 3, pp. 547–578, 2018.
- [20] M. Gupta and J. F. George, "Toward the Development of a Big Data Analytics Capability," *Information & Management*, vol. 53, no. 8, pp. 1049–1064, 2016.

