Exploring Big Data Research: A Review of Published Articles from 2010 to 2018 Related to Logistics and Supply Chains

Wecka Imam Yudhistyra  
Transportation School, Faculty of Engineering,  
Suranaree University of Technology  
Email: D6040154@g.sut.ac.th (Corresponding Author)

Evri Marta Risal  
Project and Enterprise Management School,  
University College of London  
Email: evri.risal.18@ucl.ac.uk

I-soon Raungratanampaorn  
Transportation School, Faculty of Engineering,  
Suranaree University of Technology  
Email: isoonr@g.sut.ac.th

Vatanavongs Ratanavaraha  
Transportation School, Faculty of Engineering,  
Suranaree University of Technology  
Email: vatanavongs@g.sut.ac.th

ABSTRACT

The emergence of industrial revolution 4.0, digital era, social media networks, ride-sharing transportation networks, video on demand, 5G network technology, and other cutting-edge technologies has transformed the current research paradigm in science into the fourth stage where “big data” is analyzed using various statistical explorations. Conversely, there are still deficiencies in scientific articles exploring “big data” in the logistics and supply chain research communities. Thus, the objective of this study was to conduct a comprehensive review of the trends and the current status of big data analytics (BDA) in the logistics and supply chain research communities. An elaborate examination was performed by analyzing the big data papers published in logistics and supply chain journals. The results have uncovered three things: the changing roles of BDA; some daunting challenges of implementing big data analytics, along with its recommended solutions; and the promising future of logistics and supply chains. All the findings would be helpful for creating a good research design and for implementing big data analytics (BDA) in the related fields.

Keywords: big data, analytics, logistics, supply chain

1. INTRODUCTION

Today data is everywhere all the time. It is an era where data is influenced by the developments of information technology and systems that increase extensively. Some people call it big data or data deluge (Ghorpade-Aher et al., 2016). The availability of big data has affected many areas, ranging from sports, where the Oakland A’s team (United States) used an unconventional approach to build a winning baseball team using predictive analytics to win more games with less budget (Waller and Fawcett, 2013a); precaution case where Google sought to predict the timing of flu outbreaks geographically based on search term frequency (Mayer-Schönberger and Cukier, 2014); to the field of logistics and supply chain management (SCM) when big data has deep implications (Waller and Fawcett, 2013b) (Zhong et al., 2016). In case of the SCM field, big data is providing supplier networks with greater data accuracy, clarity, and insights, leading to more contextual intelligence shared across supply chains (Columbus, 2015).

Other scientific articles reviewing big data in logistics and supply chain exist, and some of them are the following: (1) Tiwari et al. (2018) tried to provide guidance for academicians and practitioners in utilizing BDA in different aspects of supply chains. However, their research solely contributed a general guideline that did not touch the deep aspects of big data utilization, and the result of this research was remarkably narrow since it was highly sensitive to the
keyword input to the search; (2) Wang et al. (2016) observed and emphasized that managers must apprehend big data business analytics and supply chain analytics as the strategic assets that must be integrated in all business activities to enable integrated corporate business analytics. However, this research did not insert data from the practitioners and professionals’ point of view; (3) Arunachalam et al. (2018) offered a systematic literature review about the capabilities of BDA in supply chain and developed the capabilities for an advance model. The result stimulated the academic researchers to initiate a new empirical research in BDA in the context of supply chain domain and provide a direction for future research. Regrettably, the conclusion was eminently general; (4) Brinch et al. (2018) conducted a Delphi study to gain an understanding about the terminology of big data and identify the application of big data in SCM using adjusted references or the framework of supply chain operations reference. The result of the study showed that the terminology of big data seemed to be more about collecting data rather than managing and utilizing data. Moreover, their research showed that supply chain executives seemed slow to adopt big data. The weakness of this study was on the method of the Delphi study, which was still on small-scale surveys; (5) Lai et al. (2018) discussed the factors that determined the motives of companies in adopting BDA in their daily operations. The empirical results of this study revealed that the benefits and support from the top management can significantly influence the intention to adopt BDA. Since the study of big data was conducted in the early phase, the interpretation of BDA may have multiple divergences from different perspectives, causing some ambiguities in apprehending the meaning of big data.

Meanwhile, the purpose of this manuscript was to explore and analyze current conditions and trends regarding the implementation of BDA in logistics and supply chain research communities. This is important inasmuch as (a) the current research paradigm is at the fourth level where big data is analyzed using various kinds of statistical methods (Kitchin, 2014), this manuscript can provide insight into the current situations and trends of BDA in logistics and supply chain communities around the world; (b) researchers can be aware and realize their current position to minimize the gap between big data available now and the capability to analyze it (Marr, 2017), particularly in ASEAN region (Pujawan, 2016). By 2020, about 1.7 million megabytes of new information will be created per second for every human being on this planet (Taggart et al., 2016). In fact, there will be a definite shortage of skilled professionals in BDA available, while there will be a big demand for data analytics skill at this time. Meanwhile, the forecast calls for fantastic growth estimated to be 12% in 2024 for data analytics jobs as compared with growth projected to average 6.5% for other jobs (Marr, 2017).

2. RESEARCH METHODOLOGIES

2.1 Data Collection

Because of limited accessibility rights to collect scientific articles, this study focused on the journals indexed by Scopus on Elsevier’s database with “logistics” or “supply chain” on its “title.” To ensure the quality of the data collection, several journals that had been discontinued were removed, and they were filtered in terms of journal ranking, citations, and relevance to logistics and SCM for each subarea from January 1, 2011, to December 30, 2018. Although the definition of big data has been introduced since 2001, the focus of this study was the papers related to big data published over the past eight years in the journals, and this was in line with the results of a number of previous literature review studies that discovered that scientific articles related to big data were mostly published after 2010 (Tiwari et al., 2018) (Govindan et al., 2018) (Addo-Tenkorang and Helo, 2016) (Barbosa et al., 2018) (Wamba et al., 2018).

2.2 Methodology and Method

As explained earlier, this study explored, investigated, and conducted a comprehensive analysis of the current status and trend of the application of BDA techniques in logistics and SCM; therefore, the methodology for this research can be illustrated in Figure 1. The first part of the research methodology began with a literature review to reconceptualize the definition of big data, logistics, and supply chains. Rowley and Slack’s (2004) framework for literature review was adopted to be the guidance for conducting literature review processes. Likewise, these guidelines were used in recent scientific articles by Chen et al. (2014) and Wang et al. (2016). The second part of the research methodology was the journal searching containing “logistics” or “supply chain” as keyword titles. Next, the journals that had been discontinued and had no citations were removed from the journal searching result. Some journals have changed their names and are still publishing articles; however, there are also a number of journals that have stopped publishing scientific articles. This will help improve the accuracy of the article search process in journals.

The third part of the research methodology was the journal filtering in which this study used 26 BDA techniques listed by the McKinsey Global Institute (Manyika et al., 2011) as shown in Table 1 equipped with the explanation and description of each technique. The fourth part of the research methodology was the paper searching. The searching process in this step (for big data techniques used in each journal) was conducted manually by writing at least one of the 26 big data techniques and several combinations of keywords (including the writing in the title or the abstract) to distinguish the techniques included in the articles. Afterward, the journal was saved and manually reviewed to observe current conditions and trends in the future.

The last part of the research methodology was the result and discussion of this manuscript, which explained the reconceptualization of SCM in the industrial 4.0 era, the statistical data of manuscripts about the current situations and trends with BDA topics, the challenge of BDA implementation, and some knowledge discoveries or insights based on the literature review conducted in this research.
Table 1 Hypothesis testing

<table>
<thead>
<tr>
<th>Technique(s)</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>A/B testing</td>
<td>This is often referred as split testing, bucket testing, or A/B/N testing (if using multiple variants in testing). This is a technique for comparing two variants with various tests to be able to see which one has better performance to determine what treatment increases the performance of the variant.</td>
</tr>
<tr>
<td>Association rule learning</td>
<td>This is a common technique to find an interesting relationship among many variables in a large database using a variety of algorithms to generate and test the possible rules. It is also referred as market basket analysis used for data mining.</td>
</tr>
<tr>
<td>Classification</td>
<td>This is a set of techniques to identify the categories of data point in a collection for targeting categories in which data point belong to, based on data points that have already been categorized.</td>
</tr>
<tr>
<td>Cluster analysis</td>
<td>This a statistical method to classify objects into groups of similar objects where the previous characteristics of similarity are unknown.</td>
</tr>
<tr>
<td>Crowdsourcing</td>
<td>A combination of &quot;crowd&quot; and &quot;outsourcing,&quot; this technique collects data and ideas from a large group of people or community rather than an organization or own employees through an open call, usually through networked media such as the web.</td>
</tr>
<tr>
<td>Data fusion and data integration</td>
<td>This is a technique that integrates multiple data sources to obtain more consistent, accurate, and useful information rather than using data provided by individual data source.</td>
</tr>
<tr>
<td>Data mining</td>
<td>This is a technique of discovering hidden patterns from a large data set by combining methods from statistics, machine learning, and database management to analyze and extract the pattern.</td>
</tr>
<tr>
<td>Ensemble learning</td>
<td>This is a technique that uses multiple learning algorithms and predictive modeling to obtain better predictive performance than can be obtained from any of the constituent models. This is a type of supervised learning.</td>
</tr>
<tr>
<td>Genetic algorithms</td>
<td>This is a method for solving and improving both constrained and unconstrained optimization problems and is inspired by the process of natural evolution or &quot;survival of the fittest.&quot;</td>
</tr>
<tr>
<td>Machine learning</td>
<td>This is a subspeciality in the scientific study of algorithms, statistical models, and computer science (within a field historically called artificial intelligence [AI]) concerned with the design and development to improve performance from a specific task.</td>
</tr>
<tr>
<td>Natural language processing</td>
<td>This is a set of techniques from a subfield of information engineering, computer science (within a field historically called AI), and linguistics that use computer algorithms concerned to analyze human (natural) language, in particular, focus on how to use a computer to analyze and interpret a large amount of human (natural) language from a large data set.</td>
</tr>
<tr>
<td>Neural networks</td>
<td>These are powerful computational data models, inspired by the structure and workings of biological neural networks (i.e., the cells and connections within a brain), that find and represent patterns in data from complex input/output relationships.</td>
</tr>
<tr>
<td>Network analysis</td>
<td>This is a set of techniques that study the graph used to characterize relationships among discrete nodes that have attributes in a graph or a network. Connections among individuals in a community or organization are analyzed in social network analysis.</td>
</tr>
</tbody>
</table>
3. RESULT AND DISCUSSION

3.1 Logistics and Supply Chain

Primarily, it is important to know the definition and the scope of logistics and supply chain in this research since they are often used interchangeably, and many other terms such as material management, demand management, procurement, customer relationship management, manufacture inventory, packaging, and so on have similar definitions. Some claim that SCM is just another name for integration of business logistic management, and the broad scope of SCM has been promoted over the years (i.e., SCM may be concerned with product pricing and manufacturing quality). Although SCM promotes viewing the supply channel with the broadest scope, the reality is that enterprises do not practice this ideal. A single enterprise generally is unable to manage its entire product flow channel from raw materials to customers, and usually, it has a narrow scope. For example, the maximal managerial control that can be expected is over (a) the immediate physical supply channel, which is the time and space gap between the firm’s immediate material sources and its processing points; and (b) the distribution channel, which refers to the time and space gap between the firm’s processing points and its customers (Ballou, 2007). Because of the similarities in the activities between the two channels, physical supply (referred to material management) and physical distribution combined integrate those activities into business logistics. Afterward, business logistics developed into SCM as shown in Figure 2. Other terms such as value nets, value stream, and lean logistics describe similar scope and purpose with SCM.

Among the previous terms, there were no absolute definitions because the evolution and the background of logistics and SCM come from a set of functional activities (transportation, storage management, information services, etc.). Many manuscripts explained the definition of SCM based on logistics terminology and a unified definition. In the writing of Jules Dupuit, a French engineer, the idea of trading one cost for another (transportation costs for inventory costs) was evident in the selection between road and water transport (Ballou, 2007).

The fact is that carriage by road being quicker, more reliable and less subject loss or damage, it possesses advantage since businessmen often attach a considerable value. However, it may well be that the saving 0 fr.87 induces the merchant to use the canal; he can buy warehouses and increase his floating capital in order to have a sufficient supply of goods on hand to protect himself against slowness and irregularity of the canal, and if all were told the saving of 0 fr.87 in transport gives him/her advantage of a few centimes, he/she would favor of the new route (Dupuit, 1952).
Around 1961, the first textbook to suggest the benefits of coordinated logistics management was introduced. Logistics is a branch of military science that deals with procuring, maintaining, and transporting materials, personnel, and facilities (Dupuit, 1952) (Pienaar, 2009). This definition put logistics into a military description. A better representation of its definition was described by the Council of Logistics Management, a professional organization of logistics managers, educators, and practitioners formed in 1962 for the purpose of continuing education and fostering the interchange of ideas:

Logistics is that part of the supply chain process that plans, implements, and control the efficient, effective flow and storage of goods, services, and related information from the point of origin to the point of consumption in order to meet customers’ requirements (Council of logistics management, 2000).

This definition conveys the idea that product flows should be managed from the point where they exist as raw materials to the point where they are finally discarded (Chen et al., 2014). It is also worthwhile to explore several definitions for the scope and content of logistics and supply chain. Chow and Heaver (1999) described it as groups consisting of producers, suppliers, distributors, retailers and transportation, and other logistics management service providers engaged in supplying goods to consumers (Pienaar, 2009). Ayers (2001) defined it as life cycle processes involving physical goods, information, and financial flows to meet the customer needs with goods and services from suppliers (Ayers, 2001). Grant et al. (2006) characterized it as corporate business processes with integrated systems from end users through suppliers that provide information, goods, and services, which add value for customers (Grant et al., 2006). Mentzer et al. (2001) explained it as a group of entities (i.e., organizations or individuals) directly involved in the supply and distribution flow of goods, services, finance, and information from source to destination (customers) (Mentzer et al., 2001). Hanfield (2017) specified it as the activities to maximize customer value and get a sustainable competitive advantage (Handfield, 2017). LeMay et al. (2017) mentioned it as the coordination of a network through organizations and individuals to get, use, and deliver material goods; to acquire and distribute services; and to make their offerings available to markets, customers, and clients (LeMay et al., 2017). Pienaar W. (2009) specified it as a common description of the integration process that involves organizations to transform raw materials into finished goods and to transport them to the end user (Pienaar, 2009). Some researchers argue that there are many readily available definitions of SCM going beyond the fundamental concept of definition. Meanwhile, the definition of logistics and SCM in this research does not focus on the term in the linguistic definition and refers directly to the definition, which is agreed upon by experts as the following (Rushton et al., 2017):

Logistics = Material Management + Distribution

An extension to the logistics idea includes the supply of raw materials and components as well as the delivery of products to the final customer. Thus, supply chain can be defined as follows:

Supply Chain = Suppliers + Logistics + Customers

In general, it can be said that logistics is the part of SCM that manages the flow of materials and information from the supplier and customers to the end consumers. It includes inventory management, warehousing,
marketing, sales, customers, management information systems, technology, transportation, and compliance as its key logistics activities. The activities that constitute the supply chain process vary from enterprise to enterprise, depending on an enterprise’s particular organizational structure, management’s honest differences of opinion about what constitutes the supply chain for its business, and the importance of individual activities to its operation (Ballou, 2007).

3.2 Big Data

Big data is a term that describes the large volume amount of data in business (Jeble et al., 2018). Gartner (at the time, it was named META Group), a consultant firm, defined big data as “high-volume, high-velocity and high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making” (Beyer and Laney, 2012). Later, the IBM scientists added two more dimensions of big data—“high-veracity” (IBM, 2013) and “high-variability” (Andreu-Perez et al., 2015). Currently, the definition of big data has five dimensions. The description and explanation of big data with four dimensions on this research referred to the study of Katal (2013), which has a lot of citations. The description of big data with five dimensions is explained in detail below (Katal et al., 2013) (Andreu-Perez et al., 2015):

1. “Volume” – Currently, data that have reached petabyte volume is already problematic. With the increasing utilization of smartphones, internet and web technology, and wearable devices (such as smartwatches, smart glasses, smart home, etc.), data is predicted to increase to zettabyte volume in the next few years.
2. “Velocity” – The flow of data movement is rapid recently. It also causes challenge for data (previously collected and captured) that change faster than before and continuously in motion.
3. “Variety” – When data are collected not only from one specific source but also from various data formats such as web pages, images, sensors, e-mails, social media, and so on, which are structured and unstructured data, this complicates the conventional analysis technique to analyze big data.
4. “Veracity” – The focus in this dimension is the ambiguity within data (typically from noise and abnormality).
5. “Variability” – This refers to the establishment of a regular contextualization structure of data flows that can be relied upon even in conditions that cannot be predicted to be extreme. This is defined as the need to obtain data that are meaningful considering all possible circumstances.

The five “V” pillars mentioned above are the core components of big data that can produce a new “V” pillar called value if organizations have the ability to obtain greater valuable information from big data through deeper insight from superior data analytics (see Part 3.4).

3.3 Journal Filtering and Big Data Papers Published in Logistics and Supply Chain Journals

The searching verification was performed to figure out the list of scientific journals indexed by Scopus on Elsevier’s database with “logistics” and “supply chain” keywords. The journals that had been discontinued and did not have a citation score were excluded as shown in Table 2.

As a result, 23 journals had been chosen, where 5 of 19 search results journals with “logistics” and 2 of 11 journals with “supply chain” keywords had been stopped from Elsevier’s database. Among the 23 selected journals indexed by Scopus on Elsevier’s database, there were merely 43 published papers related to big data or approximately 0.63% of the total papers (an extensive gap from the total 6,929 papers published in the journals). The journal that published the least scientific articles was the International Journal of Construction Supply Chain Management with 16 papers, while the journal that published the most scientific articles was the Transportation Research, Part E: Logistics and Transportation Review with 1,063 papers. The Journal of Business Logistics and the International of Logistics Management were the journals that published the big data papers in the logistics and supply chain area the most with 11 journals using the term “big data” in their title, abstracts, or keywords. Interestingly, no more than 5% of the total papers were published by the two journals, even though they were the most widely published journals related to big data.

<table>
<thead>
<tr>
<th>Table 2 Discontinued journal list</th>
</tr>
</thead>
<tbody>
<tr>
<td>JOURNAL</td>
</tr>
<tr>
<td>Year</td>
</tr>
<tr>
<td><strong>Cellular Logistics</strong></td>
</tr>
<tr>
<td><strong>EURO Journal on Transportation and Logistics</strong></td>
</tr>
<tr>
<td><strong>Logistics and Transportation Review</strong></td>
</tr>
<tr>
<td><strong>Naval Research Logistics Quarterly</strong></td>
</tr>
<tr>
<td><strong>EURO Journal on Transportation and Logistics</strong></td>
</tr>
<tr>
<td><strong>Supply Chain Manufacturing and Logistics</strong></td>
</tr>
<tr>
<td><strong>Supply Chain Systems Magazine</strong></td>
</tr>
</tbody>
</table>

*Note that the Logistics and Transportation Review journal is continued as Transportation Research Part E: Logistics and Transportation Review, while the Naval Research Logistics Quarterly is continued as Naval Research Logistics.

Meanwhile, the authors with the most published papers in the journals were Stanley E. Fawcett and Matthew A. Waller with 4 manuscripts (see Appendix A), and most of the authors came from the United States with 38 authors, followed by the United Kingdom with 15 authors, and Australia with 13 authors. Furthermore, scientific articles were published each year: 2 from 2014 to 2015, 5 in 2016, 3 in 2017, and 25 in 2018. Complete statistical data on manuscripts using big data analysis techniques in selected journals and a comparison between the number of big data articles published and the total number of articles published in selected journals are shown in Figures 3, 4, and 5.
A comparison between the number of articles related to big data and the total number of articles published in logistics and supply chain journals per year showed the massive gap one more time (2 big data papers: 752 papers or 0.26% of total paper published in 2013 and 26 big data papers: 1,158 papers or 2.24% of total paper published in 2018). To analyze the trend toward the big data papers published in logistics and supply chain journals, five time series forecasting techniques were performed (i.e., exponential smoothing, linear trend model, quadratic trend model, exponential trend model, and autoregressive model). Four of the five models (except exponential smoothing) used the consecutive coded value 0 (2013) through 5 (2018) as the X (coded year) variable. The result of these models produced the following forecasting equations as shown in Figure 6. The R2 and the adjusted R2 for the quadratic trend model were the highest when they were compared with the other models. To further verify and to show the actual value of the big data paper published in logistics and supply chain journals, along with the predicted value of the papers, the residual, the error sum of squares (SSE), the standard error of estimates (SYX), and the mean of absolute deviation (MAD) for each of the four models, comparing the magnitude of the residuals in four models was performed. For this time series, the comparison of the SYX and MAD indicates that the first-order autoregressive model provides the poorest fit, and the quadratic trend model is the best model. Using the best model, the result of the analysis for 2019 was 35 papers, which were rather higher than the number of papers from 2011 to 2018 related to big data in logistics and supply chain journals.

Meanwhile, simple exponential smoothing (SES) merely requires deficient data storing, quickness and simplicity in computing, and emphasis on the most up-to-date information that was also exerted to predict the amount of research papers in 2019 related to big data within the scope of logistics and supply chain journals. It is likely that at least 8 papers related to big data will be published in 2019 within the scope of logistics and supply chain journals. Alpha (α) = 0.1 provided better accuracy with its lowest mean squared error (MSE) and MAD score on SES forecasting technique. Every year (from 2013 until the end of 2018), movement on the number of published scientific articles increased (except in 2017). In 2018, there was a drastic rise in the pattern movement with 24 journal papers. The result of the analysis for 2019 was 8 papers, which was higher than the number of papers from 2011 to 2017 related to big data in logistics and supply chain journals. The trend increased from the previous year as shown in Figure 7.

Thus, it can be concluded that the research topic related to big data had an extensive gap compared with the other topics within logistics and supply chain journals based on real data and information that had been collected and analyzed.

Furthermore, the result of this study explored the application of each of the 26 big data techniques. It shows that in total, these techniques appeared 1,980 times within the 23 selected journals. It is important to notice that many of the publications applied more than one of these techniques. Table 3 presents the selected research areas, the associated journals, and the results of the database search. From the information shown, there were 1,560 scientific articles related to the logistics area and 420 scientific articles related to the supply chain area with a total of about 1,980 scientific articles related to big data techniques used in logistics and supply chain journals. The five most widely used techniques in logistics journals were optimization, simulation, regression, genetic algorithms, and classification (with 1,560 journals). Meanwhile, the five most widely used techniques in supply chain journals were optimization, regression, simulation, classification, and genetic algorithms (with 420 journals). Moreover, Table 3 shows the usage of big data techniques per year. Optimization was the most frequently used technique every year in both areas (with 680 scientific articles). From 2011 to the end of 2018, the number of scientific articles that used this technique had increased dramatically. Nevertheless, there were still no scientific articles that used association rule learning, natural language processing, and unsupervised learning.

![Figure 3](image-url) Complete statistical data on manuscripts using big data analysis techniques in selected journal
Figure 4 Complete statistical data on the distribution of manuscripts using big data analysis techniques in selected journals

Figure 5 Comparison between the number of big data articles published and the total number of articles published in selected journals

Figure 6 Forecasting equations

<table>
<thead>
<tr>
<th>Year</th>
<th>Big Data Papers Published</th>
<th>Predicted</th>
<th>Residual</th>
<th>Predicted</th>
<th>Residual</th>
<th>Predicted</th>
<th>Residual</th>
<th>Predicted</th>
<th>Residual</th>
</tr>
</thead>
<tbody>
<tr>
<td>2013</td>
<td>2</td>
<td>-1.00</td>
<td>-3.10</td>
<td>4.68</td>
<td>2.68</td>
<td>0.30</td>
<td>-1.70</td>
<td>#N/A</td>
<td>#N/A</td>
</tr>
<tr>
<td>2014</td>
<td>4</td>
<td>2.28</td>
<td>1.72</td>
<td>1.12</td>
<td>-2.88</td>
<td>0.46</td>
<td>-3.54</td>
<td>13.50</td>
<td>11.50</td>
</tr>
<tr>
<td>2015</td>
<td>4</td>
<td>5.65</td>
<td>1.65</td>
<td>1.03</td>
<td>-2.97</td>
<td>0.61</td>
<td>-3.39</td>
<td>12.77</td>
<td>8.77</td>
</tr>
<tr>
<td>2016</td>
<td>5</td>
<td>9.02</td>
<td>4.02</td>
<td>4.40</td>
<td>-0.60</td>
<td>0.76</td>
<td>-4.24</td>
<td>10.04</td>
<td>5.04</td>
</tr>
<tr>
<td>2017</td>
<td>3</td>
<td>12.39</td>
<td>9.39</td>
<td>11.24</td>
<td>8.24</td>
<td>0.91</td>
<td>-2.09</td>
<td>7.31</td>
<td>4.31</td>
</tr>
<tr>
<td>2018</td>
<td>25</td>
<td>15.76</td>
<td>-10.24</td>
<td>21.54</td>
<td>-4.46</td>
<td>1.06</td>
<td>-24.94</td>
<td>4.58</td>
<td>-21.42</td>
</tr>
</tbody>
</table>

SSE       | 224.42       | SSE       | 112.41    | SSE       | 671.28    | SSE       | 712.04    |
Sxx       | 7.490        | Sxy       | 6.121     | Sxx       | 0.290     | Sxy       | 10.808    |
MAD       | 5.02         | MAD       | 3.64      | MAD       | 6.65      | MAD       | 10.21     |

Figure 7 Prediction of the big data papers published in 2019 within the scope of logistics and supply chain journals
Table 3 Amount of papers that used big data techniques from McKinsey per year published in logistics and supply chain journals

<table>
<thead>
<tr>
<th>Big Data Techniques</th>
<th>Logistics Journals</th>
<th>Supply Chain Journals</th>
</tr>
</thead>
<tbody>
<tr>
<td>A/B testing</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Association rule learning</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Classification</td>
<td>70</td>
<td>36</td>
</tr>
<tr>
<td>Cluster analysis</td>
<td>41</td>
<td>13</td>
</tr>
<tr>
<td>Crowd sourcing</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Data fusion and data integration</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Data mining</td>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td>Ensemble learning</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Genetic algorithms</td>
<td>103</td>
<td>29</td>
</tr>
<tr>
<td>Machine learning</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>Natural language processing</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Network analysis</td>
<td>58</td>
<td>9</td>
</tr>
<tr>
<td>Neural networks</td>
<td>18</td>
<td>4</td>
</tr>
<tr>
<td>Optimization</td>
<td>570</td>
<td>110</td>
</tr>
<tr>
<td>Pattern recognition</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Predictive modelling</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Regression</td>
<td>224</td>
<td>109</td>
</tr>
<tr>
<td>Sentiment analysis</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>Signal processing</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Simulation</td>
<td>353</td>
<td>78</td>
</tr>
<tr>
<td>Spatial analysis</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>Statistics</td>
<td>57</td>
<td>19</td>
</tr>
<tr>
<td>Supervised learning</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Time series analysis</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>Unsupervised learning</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Visualization</td>
<td>18</td>
<td>6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>2011</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2012</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>2013</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>2014</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2015</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2016</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2017</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2018</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

3.4 BDA in SCM

As previously mentioned, “big data” is a term that describes high volume, high variability, high variety, high veracity, and high velocity of data and inundates organizations on a day-to-day basis. The important issue in big data is that it must be combined with a variety of analytical approach (i.e., statistical analysis, optimization, visualization, clustering, etc.) to extract valuable information from the large data set. This combination is referred as BDA as shown in Figure 8. The role of BDA in the immense impact of e-commerce, social media, self-driving cars, wearable device, barcode scan, drones, internet, and cloud storage era can transform the definition and the management of supply chain. Since the current supply chain cannot be equated with its traditional function (where a single purely operational function was reported to sales or manufacturing and focused on ensuring the supply of production and delivery lines to customers, and it was also combined with other independent SCM functions in several companies), the focus of the SCM function has shifted to advanced processes. First, it enlarges the data set for analysis beyond the traditional enterprise resource planning and supply chain management system internal data. Second, it applies powerful statistical techniques by combining internal and external data sources to analyze. Third, it creates a new approach that improves supply chain decision analysis, strategic choice, and operation optimization.

In relation to the papers publishing the implementation of BDA in logistics and supply chain journals, there were several scientific articles employing big data analytical techniques that were published in the said journals from 2011 to 2018. Based on the data on social media (in these cases, it was Twitter), sentiment analysis techniques, natural language processing, clustering analysis, data visualization, and supervised learning were combined to obtain a rich picture of interaction between the company and customers, to predict market demand, and to find out key issues that affect customer satisfaction (Bhattacharjya et al., 2016) (Bhattacharjya et al., 2018) (Singh et al., 2018). These studies showed the use of a set of tools to gather insights from a large amount of unstructured data from conversations on social media platforms. Bohács and Rinkács (2016)
purely performed data visualization to analyze and model the complex systems. To make a good decision in a distributor company, Hamister et al. (2018) integrated data visualization, optimizing modeling, and forecasting safety stock using real data. To improve the value density of raw data is a crucial step that allows analysis to acquire meaningful data, including optimization using data in the real world. Miller et al. (2018) implemented data visualization and statistical approach of big data techniques to devise a framework that examined the services of the supply chain setting of health care with public data regarding hospital-level patient satisfaction.

The deficient number of papers related to the implementation of BDA published in logistics and supply chain journals means there are many opportunities to explore several captivating issues using BDA in the logistics and supply chain areas by the research communities. Some of the main areas of logistics and supply chains that can be explored using BDA are as follows:

1. Production, inventory, and operation planning can be improved using BDA in planning processes and demand capability. For example, the visibility of point of sales data, inventory data, and production planning data can be analyzed to identify mismatch between supply and demand with real-time analysis. By incorporating temperature and data about sunny days, ice cream companies are able to predict the demand more accurately. In addition, other companies are able to perform the similar thing and drive actions such as price changes, the timing of promotions, shaping demand, or the addition of new lines based on the customer behavior in the right place for the right time to realign things with BDA.

2. In workload optimizations and picking zone allocation in warehouses based on efficiency—for example, to improve storage efficiency and picking productivity—a big data 3-D model visualization analysis can be used to model and simulate new configuration. Sensor data in warehouses can be exerted to anticipate real-time transportation requirement.

3. Connecting real-time routing allocation between transportation and warehouses is utilized to improve the operational optimization such as trucks that use consumption fuel analytics to enhance driving efficiency and use GPS to reduce waiting time in warehouses in real time.

4. Routing optimization for transportation vehicles based on traffic congestion, weather, and driver characteristics is easily analyzed with BDA before the era of autonomous truck comes.

5. Real-time out-of-stock detection and prevention for point of sales are aligned with the production planning analysis. The combination between point of sales and consumer behavior data is used to improve inventory management.

6. Consumers’ sentiment data on social media can be used to obtain a rich insight into the interaction between enterprise and consumers, and risk management can be used for fraud detection and to provide credit rating to define payment term offered to mitigate the risk.

In the future, the development of automated production at factory, autonomous truck, automated warehouse, shipping rerouting, drone technologies, and AI ecosystems will lift SCM to the next level of management and definition. Automated production machine will provide constant feedback on production capacity based on sales information and consumer behavior. Driverless trucks with live transit location via satellite get in touch with good movements. Operations in warehouses such as picking and transporting good and information flow of status will be handled by machines. Customers will be able to automatically reroute their delivery address to new information on their smartphones. Drones will deliver and pick the goods to the border areas. AI will perform an international network analysis, where, related to law, international transactions, international contracts or agreements, legal systems or blockchain systems, and everything humans perform currently are predicted being replaced by AI in the future. Figure 9 illustrates the previous, the current, and the future of supply chain concept.

### 3.5 Challenges of BDA Implementation

The result of this study also brought up the challenges of big data implementation. As big data starts to grow and evolve, BDA will continue to grow and become important in academic and enterprise areas. Despite the great potential and opportunity of BDA, there are also some great challenges and risks. When the BDA challenges are addressed in a proper manner, the success rate of implementing big data solutions automatically increases. As big data makes its way into academics and particularly enterprises around the world, addressing these challenges are extremely important to unlock the full potential of BDA. Below are some of the main challenges of BDA implementation:

1. Dealing with data growth and amount of accessible data. As big data is continuously growing and expanding, new technologies are being developed every day. The combination of data explosion and technology developments become a big challenge for enterprises to be answered without the introduction of new risks and problems. In addition, the limitless amount of accessible information can become an
obstacle for enterprises to deal with because the scale and variety of the data can be too overwhelming to be analyzed and create waste that negatively affects big data usefulness.

2. Synchronization among data source. A lot of different types of software systems from different areas of daily functional activities in enterprises, which are quite challenging to incorporate into analytical insight and meaningful messages, exist.

3. Organizational resistances and culture. Some organizations are still conservative. Most business firms, government institutions, and other organizations may resist big alterations. The biggest impedance to the adoption of BDA is related to cultural change such as organizational alignment, lack of understanding, and managerial change. In an interview, a director of supply chain in automotive industry mentioned, “Koreans, in general, are not familiar with making decision based on decision making quantitative data. Rather, they tend to make important decisions based on their past experience or opinion of someone they trust. This emotional attribute may be an obstacle” (Richey et al., 2016). Attempt to change the doctrine and the culture requires a lot of patience and great persistence.

4. Security and privacy of the data. Soon after the business enterprises discover how to utilize BDA, it brings them a wide range of possibilities and opportunities. However, confidential data, such as unit price data, are important for enterprises, and there are a lot of risks when enterprises share these data to be analyzed. If these kinds of data leak outside the organization, for example, when the company makes a $15 profit from the $25 product, the client will push the company to down the price. Therefore, some confidential information will be difficult to get into big data analysis.

5. Real-time analysis. The challenges of getting important insights through the use of real-time BDA can be an obstacle since the velocity of the data is growing swiftly time by time.

4. CONCLUSION

As the result of this exploratory paper of the current situation related to the implementation of big data techniques within logistics and supply chain journals, there were almost 2,000 scientific articles that used big data techniques within the journals of logistics and supply chain. Furthermore, there were 43 scientific articles that mentioned the context of their scientific research studies within the big data context (where there was a massive gap between big data scientific articles compared with the other topics in the journals). This indicated that the utilization of big data techniques and analysis was sparsely published in the logistics and supply chain journals (within the context of big data). However, a trend analysis showed that the topics related to big data published in logistics and supply chain journals have increased every year (from 2013 to 2018). This also showed that BDA and SCM have received more attention from academics and become a potential topic to be continued into further research, which will be beneficial to many sectors, such as academics and other industries.

Furthermore, BDA in the digital era means that big analytics turn that data into real insight. It is an art that must be understood and implemented. Thus, the principal focus in BDA is not what is being analyzed but how the analysis is being implemented within the context of big data. Based on the reviews and findings of this paper, the following are some points to be noted:

1. There is a lack topic in big data research. There were no adequate research studies in implementing big data techniques in the supply chain and logistics field from January 1, 2011, to December 30, 2018, and the most common finding across all areas is that the interest in big data research is only recent.

2. There is a need to improve and implement research on big data. Based on the statistical result about the current situation and trend in implementing BDA in this research, it is recommended that the future research should focus on the implementation of BDA within specific industries and use real data to reduce the existing gap now. All aspects of SCM must also be studied with real data, especially big data analysis for planning, forecasting, procurement, and material management, as well as big data analysis that supports the alignment, agility, and sustainability of supply chain with integrated systems.

3. Enterprises in ASEAN must assimilate BDA or risk themselves of being left behind. In developed countries, new digital technologies are changing the traditional ways of SCM. Using a digital foundation in place, now enterprises can capture, analyze, integrate, easily access, and interpret high-quality data in real time, which support enterprises to make better decisions. Meanwhile, enterprises in developing countries (i.e., ASEAN) should consider and realize the importance of big data and technology to support and change their business process or face the risk of being left behind.

4. The ability of big data analysis to minimize the existing gap now must be improved. Big data and technology have played important roles in this new era. Since the skills needed for these new roles are not readily available today, the biggest challenge for researchers is to fill those critical roles.

5. The future of SCM is AI. It is clear that the future of SCM is the technology that optimally manages end-to-end workflows and requires little human intervention.

In addition, the limitation of this research was that each of the big data techniques (as previously stated) has overlapped with one another (for example, a combination of statistics and regression analysis constitutes the part of predictive analytics, etc.). Thus, it affects the accuracy of the search results in several aspects. Another limitation was that the sample size examined in the time series forecasting analysis was quite small. It could affect the result of forecasting accuracy in this research.

REFERENCES


APPENDIX 1: AUTHORS LIST

AUSTRALIA
Author Name Affiliation Publication(s) SINGAPORE
Author Name Affiliation Publication(s)
Adrian Ellison University of Sydney 1 Indu Niranjan S.P. Jain School of Global Management 1
Anta Geodir University of Sydney 1 Indu Niranjan S.P. Jain School of Global Management 1
Dayne Simpson Monash University 1 Indu Niranjan S.P. Jain School of Global Management 1
John Hopkins Swinburne University of Technology 1 Indu Niranjan S.P. Jain School of Global Management 1
Jyothymma Bhadhanyara University of Sydney 2 Indu Niranjan S.P. Jain School of Global Management 1
Kacil Manuel Quezoz Swinburne University of Technology 1 Indu Niranjan S.P. Jain School of Global Management 1
Nageeth Shikle University of Technology Sydney 2 Indu Niranjan S.P. Jain School of Global Management 1
Paul Hawking Victoria University 1 Indu Niranjan S.P. Jain School of Global Management 1
Renato Telles University of South Wales 1 Indu Niranjan S.P. Jain School of Global Management 1
Shahriar Aitler University of Wollongong 1 Indu Niranjan S.P. Jain School of Global Management 1
Soami Tripathi University of Sydney 1 Indu Niranjan S.P. Jain School of Global Management 1
Vincenzo Panza University of New South Wales 1 Indu Niranjan S.P. Jain School of Global Management 1

BRAZIL
Author Name Affiliation Publication(s) BRAZIL
Author Name Affiliation Publication(s)
Alberto de la Calle Vicente University of Ouro Preto 1 Marcelo Bronzo Ladeira Federal University of Minas Gerais 1
Bruno Franco University of Minas Gerais 1 Marcelo Bronzo Ladeira Federal University of Minas Gerais 1
M. P. Valadare de Oliveira Federal University of Espirito Santo 1 Marcelo Bronzo Ladeira Federal University of Minas Gerais 1
Renato Telles Universidade Federal Fluminense 1 Marcelo Bronzo Ladeira Federal University of Minas Gerais 1

CHILE
Author Name Affiliation Publication(s) CHILE
Author Name Affiliation Publication(s)
Roberto Chavez Universidad Diego Portales 1 Patrik Jonsson Chalmers University of Technology 1

CHINA
Author Name Affiliation Publication(s) CHINA
Author Name Affiliation Publication(s)
Helen Sun Harbin Institute of Technology 1 Emily R. H. de St. Croix Deloitte Consulting AG 1
Jian Ran Harbin Institute of Technology 1 Emily R. H. de St. Croix Deloitte Consulting AG 1
Yuanxian Li Harbin Institute of Technology 1 Emily R. H. de St. Croix Deloitte Consulting AG 1
Zongwei Luo Southern University of Science and Technology 1 Emily R. H. de St. Croix Deloitte Consulting AG 1

DENMARK
Author Name Affiliation Publication(s) DENMARK
Author Name Affiliation Publication(s)
Andreas Weidig Copenhagen Business School 1 Ninh Phat S.P. Jain School of Global Management 1
Christopher Rajkumar University of Southern Denmark 1 Ninh Phat S.P. Jain School of Global Management 1
Jan Stensfeld University of Southern Denmark 1 Ninh Phat S.P. Jain School of Global Management 1
Jesper Kromborg Jansen ENERGUSK 1 Ninh Phat S.P. Jain School of Global Management 1
Kamran Gioviani University of Southern Denmark 1 Ninh Phat S.P. Jain School of Global Management 1
Martin Blich University of Southern Denmark 1 Ninh Phat S.P. Jain School of Global Management 1

FRANCE
Author Name Affiliation Publication(s) FRANCE
Author Name Affiliation Publication(s)
David Raultman ESSEC Business School 1 Despale Anruchaml University of Sheffield 1
Frank Chen ESSEC Business School 1 Despale Anruchaml University of Sheffield 1
Ramezehr Dube University of Southern Denmark 1 Despale Anruchaml University of Sheffield 1
Samuel Fossan Wambis Toxouche Business School 1 Despale Anruchaml University of Sheffield 1

GERMANY
Author Name Affiliation Publication(s) GERMANY
Author Name Affiliation Publication(s)
Arkadius Schir FrÃ¶hlich-Institut fur Materialfluss und Logistik IMI 1 Stephen J. Childe Plymouth University 1
Elisabeth Nederland Hungarian Academy of Sciences 1 Stephen J. Childe Plymouth University 1
Georgia Pruksa University of Miskat University 1 Stephen J. Childe Plymouth University 1
Katharina Bokota University of Miskat University 1 Stephen J. Childe Plymouth University 1
Lea Zwilling Hungarian Academy of Sciences 1 Stephen J. Childe Plymouth University 1

HONG KONG
Author Name Affiliation Publication(s) HONG KONG
Author Name Affiliation Publication(s)
Eric Ngai Hong Kong Polytechnic University 1 A. Michael Kneemeyer The Ohio State University 1
T.C. Cheng Hong Kong Polytechnic University 1 A. Michael Kneemeyer The Ohio State University 1
Tsang-Ming Choi Hong Kong Polytechnic University 1 A. Michael Kneemeyer The Ohio State University 1

HUNGARY
Author Name Affiliation Publication(s) HUNGARY
Author Name Affiliation Publication(s)
Angela Rinkacs Budapest University of Technology and Economics 1 George G. Adams St. Louis University 1
Gabor Bohacs Budapest University of Technology and Economics 1 George G. Adams St. Louis University 1
Laszlo Monczo Hungarian Academy of Sciences 1 George G. Adams St. Louis University 1
Zoltan Kemery Hungarian Academy of Sciences 1 George G. Adams St. Louis University 1

INDIA
Author Name Affiliation Publication(s) INDIA
Author Name Affiliation Publication(s)
Aatish Singh Indian Institute of Technology Delhi 1 A. Michael Kneemeyer The Ohio State University 1
Anand Prajhap National Institute of Construction Management Research 1 A. Michael Kneemeyer The Ohio State University 1
Surya Prakash Singh Indian Institute of Technology Delhi 1 A. Michael Kneemeyer The Ohio State University 1

ITALY
Author Name Affiliation Publication(s) ITALY
Author Name Affiliation Publication(s)
Carlo Maria Persino University of Bologna 1 A. Michael Kneemeyer The Ohio State University 1
Sandro Barbonesi University of Bologna 1 A. Michael Kneemeyer The Ohio State University 1

MALAYSIA
Author Name Affiliation Publication(s) MALAYSIA
Author Name Affiliation Publication(s)
Nori Hobiah Abu Universiti Utara Malaysia 1 A. Michael Kneemeyer The Ohio State University 1

ROMANIA
Author Name Affiliation Publication(s) ROMANIA
Author Name Affiliation Publication(s)
Alexandru Armien Universitatea de Vest din Timisoara 1 A. Michael Kneemeyer The Ohio State University 1

SOUTH AFRICA
Author Name Affiliation Publication(s) SOUTH AFRICA
Author Name Affiliation Publication(s)
Jong-Min Park Korea Research Institute of Ships Ocean Engineering KRISO 1

SWEDEN
Author Name Affiliation Publication(s) SWEDEN
Author Name Affiliation Publication(s)
Yu-Chung Tasso National Taiwan University of Science Technology 1 A. Michael Kneemeyer The Ohio State University 1

UNITED ARAB EMIRATES
Author Name Affiliation Publication(s) UNITED ARAB EMIRATES
Author Name Affiliation Publication(s)

UNITED KINGDOM
Author Name Affiliation Publication(s) UNITED KINGDOM
Author Name Affiliation Publication(s)

UNITED STATES
Author Name Affiliation Publication(s) UNITED STATES
Author Name Affiliation Publication(s)

© 2020
Wecka Imam Yudhistyra is a university student in the Civil, Transportation, and Geo-Resource Engineering at Suranaree University of Technology in Nakhon Ratchasima, Thailand. His research interests focus on implementation of big data analytics in logistics and supply chain management. He enjoys playing music, reading books, and learning how to take advantage of information technology and systems so they can add value to their businesses, products and services.

Evri Marta Risal is a university student in the Project and Enterprise Management school at University of College London in London, United Kingdom. His research interests focus on project management for enterprise. He enjoys traveling around the world and taking artistic photos as a hobby.

I-soon Raungratanaamporn completed his bachelor’s in architecture and planning studies, and master’s degree in urban planning study from Thammasat University in Bangkok, Thailand. Dr. Raungratanaamporn also has a Ph.D. in Policy Science from Ritsumeikan University in Kyoto, Japan and teaches regularly at Suranaree University of Technology. He also speaks frequently at both international and domestic conferences. In addition, he lives in Bangkok, Thailand with his wife and enjoys playing music and writing creatively as a hobby.

Vatanavongs Ratanavaraha is a professor at Suranaree University of Technology. He specializes in supply chain management, transportation safety, and tourism management. He obtained his PhD in transportation studies from Vanderbilt University in Tennessee, United States. He has published over 13 international scientific articles in journals and speaks frequently at both international and domestic conferences. In addition, he and his wife live in Nakhon Ratchasima, Thailand and enjoy preserving fish as a hobby.