ISSN 2167-1528
v15 n5735

Big Data Analytics in Supply Chain Management: A Review and Recommendations

Huiyun Ouyang
ouyan1h@cmich.edu

Xiao Shen Cham
cham1x@cmich.edu

Xiaofang Chen
chen9x@cmich.edu

Mark Hwang
hwang1m@cmich.edu

BIS Department
Central Michigan University
Mt. Pleasant, MI 48859, USA

Abstract

Massive data has become a critical resource for companies to achieve significant performance improvement and sustainable competitive advantage. However, traditional data analytics techniques are deficient in absorbing and making sense of the deluge of big data. In response, big data analytics (BDA) has generated pronounced interest in recent years in diverse fields. BDA has also garnered keen attention from academics and practitioners of supply chain management (SCM). This research explores ways to achieve supply chain performance improvements from the effective use of big data analytics. Big data technologies and security issues are investigated. Four types of analytics--descriptive, diagnostic, predictive, and prescriptive analytics--and their application as BDA in SCM are discussed. The pros and cons of each analytics type are detailed, along with developments that facilitate the progression to the next-generation supply chain: the digital or thinking supply chain. Technological and organizational factors that are prerequisites for the success of the digital supply chain are elucidated.

Keywords: Big data, big data analytics, supply chain management, digital supply chain

1. INTRODUCTION

The supply chain is a network system that covers a wide range of businesses, from raw material suppliers to end consumers. The traditional supply chain includes many activities connecting the supplier, manufacturer, wholesaler, retailer, and customer in a row. A company's ability to provide a positive customer experience is adversely affected if its supply chain has limited visibility and lacks real-time updates. Because the supply chain plays a critical role in business success, companies look for opportunities to improve supply chain management (SCM) by making it easier and faster to execute processes with lower costs. Supply chain analytics has become critical in managing the ever-increasingly complex system, especially with the swelling number of international partners (McCue, 2020). Supply chain analytics allows data to be analyzed in real-time and alerts decision-makers with potential problems before they snowball into more serious issues, allowing quicker adjustments and more effective tactical decisions. New technologies, including IoT devices and cloud solutions, manage the supply chain more effectively, but at the same time, they generate much more data than previously. Studies show that data generated in 2020 from an average supply chain was 50 times more than that of 2015 (Ellis, 2020). In recent years, big data analytics (BDA) has emerged to handle enormous datasets using advanced analytic techniques in various fields. BDA applications in SCM have achieved many benefits (Anitha & Patil, 2018). However, significant challenges remain. This paper reviews the literature to highlight issues related to the effective use of BDA in SCM. An in-depth look at big data, including its technologies and security issues, is presented in the next section. A discussion of the current practice of BDA in descriptive, diagnostic, predictive, and prescriptive analytics follows. Next, it compares the pros and cons of each analytics. The paper concludes with future development directions and recommendations for building the next-generation supply chains. These recommendations provide guidelines to SCM practitioners to help with their BDA initiatives. The guidelines in turn identify future research directions.

2. CURRENT PRACTICE

Big Data

Big data can be defined as enormous and complex datasets that exceed the capabilities of traditional systems used for storing, handling, overseeing, deciphering, and visualizing (Darvazeh et al., 2020). Big data per se offers little value if not properly managed. In fact, with large amounts of data pouring in from diverse sources, the probabilities of having data errors are higher. Initially characterized by three V's – volume, variety, and velocity of data, big data has been expanded to five or more V's in recent years. In SCM, big data shares five common characteristics – velocity, veracity, variety, value, and volume (Darvazeh et al., 2020; Maheshwari et al., 2021). Other researchers discuss different characteristics, including variability, visualization, virality, viscosity, and volatility (Anitha & Patil, 2018).

A systematic approach to data analysis using statistical and mathematical models is essential to combat big data. Known as big data analytics (BDA), this new approach to data analysis enables organizations to turn low-value data into high-value information (Lee & Mangalaraj, 2022). The result is insights that can help businesses make better decisions to solve complex problems (Darvazeh et al., 2020). While still a relatively new concept, BDA and its technologies are beginning to play a significant role in various industries such as retail, healthcare, education, government, manufacturing, and services to boost organizational performance (Wang et al., 2020). Businesses that are open to investments in BDA, if properly implemented, could find its applications highly rewarding.

Technologies and Security Issues

The traditional information system and architecture used in most businesses today would not suffice for big data applications. In terms of data storage, conventional components such as SQL databases can no longer accommodate this new wealth of data. According to Leveling et al. (2014), big data is structured too individually to fit into relational databases, and therefore, databases with minimal schema restrictions such as NoSQL databases are often preferred. NoSQL graph databases are especially suitable for supply chain modeling because their structure supports describing applications, including schedule optimization, navigation systems, and social networks (Leveling et al., 2014). Other NoSQL databases, such as key-value pair and document databases, offer speedy processing of unstructured data to facilitate customer segmentation or sentiment analysis that is hard to undertake using a relational database.

Irregular arriving data streams add to extant large data sets and can present another challenge to supply chain processes. Often in the form of real-time data, data streams require lambda architecture for data processing and analytics (Leveling et al., 2014). Big data integrated with modern technologies, such as complex event processing (CEP), RFID, blockchain, IoT, and wireless sensor networks (WSN), can boost supply chain performance. CEP processes data streams and is used to integrate and process events from RFID devices, sensors, and barcode readers (Leveling et al., 2014). In the shipping and logistics sector, big data for air quality has been collected with WSN technologies for analytical decision-making (Lee & Mangalaraj, 2022). RFID has become an integral component in all stages of SCM to ensure quick and secure identification of various items particularly (Anitha & Patil, 2018). Lee & Mangalaraj (2022) note that KPIs can be tracked by equipping production floors with RFID systems and wireless networks for logistical decision-making. Businesses must analyze their own organizational needs and invest resources in appropriate technologies to make the most of big data.

Despite the increasing importance of big data in SCM, the issue of data privacy and security remains underexplored for the most part in areas concerning BDA (Melnyk et al., 2022). With emerging technologies such as IoT, blockchain, and machine learning gaining traction in the supply chain industry, robust security considerations are essential for the seamless flow of data and analytics capabilities (Lee & Mangalaraj, 2022). As enterprises shore up their cybersecurity defenses, perpetrators are increasingly targeting suppliers or third-party vendors in supply chain attacks (Crowstrike, 2021). One well-known example is the SolarWinds hack uncovered in 2020, which affected over 18,000 organizations including numerous US government agencies and Fortune 500 companies (Vaughan-Nichols, 2021). At the same time, current security policies make integrating different data sources difficult despite new solutions developed for greater supply chain visibility. Leveling et al. (2014) expound that unstructured large data files are being transformed into structured data records for conventional databases, which is time-intensive and hard to process later by other systems due to access restrictions. Corporate security policies likewise limit data integration with external systems. Therefore, revamping security policies and procedures is a prerequisite to achieving big-data integration.

3. BIG DATA APPLICATIONS IN SCM

The application of big data is widespread in SCM. Companies ranging from Amazon, Microsoft, to Procter & Gamble and Rolls Royce are using BDA to make their supply chain management more effective (Maruti Techlabs, N.D.). Darvazeh et al. (2020) provide a good summary of the supply chain areas transformed by BDA, including supplier relationship management, supply chain network design, product design and development, demand planning, procurement management, customized production, inventory management, logistics, agile supply chain, and sustainable supply chain.

Researchers discuss different types of analytics and commonly distinguish descriptive, predictive, and prescriptive analytics (e.g., see Darvazeh et al., 2020; Hallikas et al., 2021). A more granular classification includes diagnostic analytics, which along with descriptive analytics, are focused on the past and are the domain of data analysts. On the other hand, predictive and prescriptive analytics are future-oriented and are more suitable for data scientists (Maheshwari et al., 2021). The four types of analytics also differ in their complexity and potential value with descriptive being the lowest and prescriptive being the highest (Kalsbeek, 2020). Table 1 shows the four types of analytics and their characteristics.

Type	Focus	User	Complexity & Value
Descriptive	Past	Data analysts	Low
Diagnostic	Past	Data analysts	Medium
Predictive	Future	Data scientists	High
Prescriptive	Future	Data scientists	Highest

Table1: Four Types of Analytics

Descriptive Analytics

Through descriptive analytics, problems and opportunities can be identified by analyzing historical data. According to Darvazeh et al. (2020), sensor networks and instruments installed on the production floor generate large amounts of data. Companies can improve distribution efficiency and process control using big data through tighter analysis and database integration (Darvazeh et al., 2020). Big data can also help with supply chain optimization needs stemming from rapid globalization. In demand planning, cluster-based association rule mining can be applied to past data to uncover customer purchase patterns and anticipate future purchases (Lee & Mangalaraj, 2022). Concerning the physical flow of goods, trend analysis of historical data can achieve single shipping route optimization and appropriate resource allocations per trip (Anitha & Patil, 2018).

Applying descriptive analytics can help more easily uncover data patterns; however, this method rarely reveals the complete picture of the underlying raw data. Common approaches include data queries, dashboards, and data mining methods (Camm et al., 2020). In this scenario, anyone with limited analytical skills can generate routine reports such as “data preparation, automated dashboard, and data visualization” based on pertinent historical information (Xu et al., 2021). On the one hand, descriptive analytics is a relatively straightforward practice for day-to-day work because it utilizes several standard procedures to find patterns and trends in past business activities. On the other hand, descriptive analytics is not appropriate for performing in-depth analyses of the relationships between specific variables in the summarized reports (Camm et al., 2020) because it only reflects what happened in the past. It cannot disclose the root cause of poor performance or predict future occurrences. In sum, descriptive analytics is easy and helpful to figure out trends in historical data without providing an in-depth examination.

Diagnostic Analytics

By understanding the impact of past supply chain incidences, diagnostic analytics with big-data integration proactively tackle global supply chain disruptions. Brintrup et al. (2019) describe an expert-proposed conceptual framework involving natural language processing to pinpoint potential supply chain risks from unstructured data sources such as news outlets. Once the risks are identified, a simulation engine executes to predict their possible consequences on company performance (Brintrup et al., 2019). Recently, the global supply chain disruptions caused by the COVID-19 pandemic unearthed more diagnostic capabilities of BDA. The impact of the pandemic on consumers was thoroughly analyzed using statistics and text mining to identify significant patterns and correlations in consumer data and online sentiments (Lee & Mangalaraj, 2022). With diagnostic analytics in big data, companies can improve supply chain operations and respond decisively to future disturbances.

The advantages of diagnostic analytics include gaining tailored solutions and conducting analytical reviews based on solid data (Maheshwari et al., 2021). Diagnostic analytics allows a business to investigate the mistakes in designated fields and develop customized solutions in specific areas. The company can find out the root cause of problems through diagnostic analytics. In addition, because these solutions are generated from the actual historical data within the organization, they present valuable experiences for business improvement in the future. On the other hand, the disadvantage of diagnostic analytics is that this approach studies previous activities and does not offer insight into future operations (Maheshwari et al., 2021). In this regard, diagnostic analytics provides the business with a better understanding of events that happened in the past but does not provide clues on how to solve future problems.

Predictive Analytics

One evolving area of big-data utilization is predictive machine learning. Predictive analytics can find hidden patterns within big data using statistical and mathematical models. Predictive machine learning offers substantial benefits in SCM despite its early stage of development. For instance, decision models can embed predictive machine learning to foster continuous learning to enhance decision-making over time as algorithms become increasingly optimized (Brintrup et al., 2019). In addition, trained machine learning algorithms can convincingly predict impending machine failures through real-time supply chain data analysis (Darvazeh et al., 2020). A deeper understanding of automation within the production process can form a more reliable and intelligent supply chain system. In cybersecurity, likewise, machine learning techniques have been applied to predict cyberattack patterns on cyber supply chain systems (Lee & Mangalaraj, 2022). Despite its encouraging benefits, Brintrup et al. (2019) assert that machine learning usage in supply chains remains underexplored, as most organizations have limited its use only to sales and customer behavior forecasting.

Predictive analytics can support a business to better respond to uncertainties and potential risks in future operations, which is critical to optimize firm performance and enhance business value (Camm et al., 2020). Shamout (2019) discusses how forecasting as a predictive analytics tool is essential to secure funding and planning for the long-term development of an enterprise. However, the predictions can be inaccurate due to inconsistent, missing, ambiguous, or unknown information (Camm et al., 2020). Also, since predictive analytics is built on historical events, the conclusion may not be applicable in future circumstances if there is a drastic change in either the internal or external environment. Consequently, it is vital to note that predictive analytics can arm a business with the knowledge to respond more effectively to future incidents. The outcome can be suboptimal if the data used to build the predictive model is no longer valid or relevant.

Prescriptive Analytics

Big data can also be analyzed using prescriptive analytics to enhance customer relationships. Understanding customer preferences over-time is crucial in maintaining customer relations, which has received a tremendous boost from simulations and optimization techniques employed in prescriptive analytics. According to Darvazeh et al. (2020), BDA has been used for product design by monitoring up-to-date customer behavior and informing the customers' expectations and opinions. Once data is collected and analyzed, engineering design transforms customer needs into design specifications. Wang et al. (2020) describe a growing use of big data for customer segmentation. Big data techniques such as logistics regression, data mining, and neural networks have been used to categorize and cluster customers for targeted marketing strategies (Wang et al., 2020). Similarly, distribution network optimization can significantly lower distribution costs by picking the appropriate warehouses and production plants (Anitha & Patil, 2018). The results translate into customer cost savings and better overall customer experiences.

A business can exploit prescriptive analytics to improve efficiency and reduce risks with specific actions (Camm et al., 2020). As an example, ConAgra Foods realized a 100% return on its investment in analytics in less than three months by incorporating the inherent uncertainty in commodities pricing in its capacity utilization model (Camm et al., 2020, p. 12). At the top level of analytics techniques, prescriptive analytics integrates the relevant parties inside and outside the organization and advises on the best possible resolutions for future development (Xu et al., 2021). From this perspective, the business can capture a competitive advantage with a plan of action based on the business condition (Camm et al., 2020). The disadvantage of prescriptive analytics is that this approach is complex and costly, and the output can be useless if there is improper input (Camm et al., 2020). This outcome can happen because the optimization model and problem formulation of prescriptive analytics are complicated, and the development of algorithms requires a knowledgeable data scientist with experience and judgment (Maheshwari et al., 2021). In addition, if the decision variables are incorrectly set up, or the historical data used to build the model is no longer relevant, the results would be uninterpretable. Thus, prescriptive analytics can be a powerful tool to guide a business in optimizing its performance; however, the implementation is challenging, and the firm can go down the wrong path if applied inappropriately.

Future Developments

While digital technology has been applied to individual processes in the supply chain for many years, the "digital supply chain" (DSC) dictates a holistic approach to solving supply chain problems (Kurz & Anandarajan, 2021). Currently, big data is pushing changes in every field of the business. In addition, it is hard to manage all the contingencies using the traditional linear process. A transparent network that connects all systems, both internal and external, to share big data and detect and respond to any changes automatically is a DSC or a "thinking supply chain" (Calatayud et al., 2019; Ellis, 2020; Kurz & Anandarajan, 2021; McCue, 2020). Big data analytics is a key to transforming business data and external information into valuable insights, conclusions, and practical actions (Maheshwari et al., 2021). It will play a similar role in a DSC, especially analytics using artificial intelligence, machine learning, and deep learning, also known as cognitive analytics (Ellis, 2020; McCue, 2020). An end-to-end integrated DSC is characterized by five C's: connected, collaborative, cyberaware, cognitively enabled, and comprehensive (Ellis, 2020). Cyberaware addresses a common weakness in traditional SCM. As more data and systems are integrated into a DSC, expanding protections against cyberattacks has become increasingly critical. An effective cybersecurity strategy is needed to build security into every level of a corporation and every endpoint of a DSC (Kache & Seuring, 2017). A well-implemented digital supply chain solution can improve firm performance in three aspects: less need to make tradeoffs between conflicting goals, e.g., cost reduction versus customer satisfaction, new business models that generate new revenues, and sustained competitive advantage (Kurz & Anandarajan, 2021).

4. RECOMMENDATIONS

The use of BDA in SCM is promising and can result in the next-generation supply chain, the DSC. However, to fulfill the promise, a company must invest heavily in organizational and technology dimensions (Kurz & Anandarajan, 2021). The technology dimension consists of two components: infrastructure and analytics. Infrastructure includes technologies dealing with big data. The starting point is an integrated internal system, implemented as an ERP system in many organizations that can process structured transaction data spawned from a supply chain. However, unstructured data generated from IoT devices, social media posts, tweets, etc., needs to be captured and processed too. Investments in NoSQL databases will be the next logical step. Once (big) data is captured, cleansed, and stored, sharing it among partners on the supply chain is needed. The solution often requires commercial cloud services to enable multi-enterprise collaboration (Ellis, 2020). Another critical technology is blockchain. The immutable ledger stored on a blockchain allows trading partners to record and share data in real-time in a trusted manner. Companies of all sizes can follow the lead of pioneers such as Walmart to upgrade their supply chains (Chapman, 2020).

The second technology component is analytics, the software used to analyze big data. As discussed previously, BDA can be applied to different types of analytics, ranging from descriptive to prescriptive. Using cognitive analytics individually or in combination offers tremendous potential for making better decisions in SCM (Wang et al., 2020). Analytics, however, does not run itself. It takes a nurturing environment to deploy and flourish, determined by organizational factors. As Hallikas et al. (2021) put it, "various competencies, processes, and organizational leadership are critical enablers for realizing the potential of the digital tools." (p.634)

Kurz & Anandarajan (2021) also discuss the importance of developing analytics capabilities related to employee and organizational factors. Companies need talents proficient in data analytics and new approaches to application development for systems with high degrees of collaboration and integration. Employees who can spot new threats or opportunities in the DSC are critical to reanalyzing significant supply chain performance improvements. McCue (2020) similarly discusses the training needs for both developers and users of BDA as a chief challenge. A company needs to invest in "analytics capabilities" to transform raw data into meaningful insights to increase its supply chain productivity and strengthen its competitive advantage (Hallikas et al., 2021). These capabilities include leading-edge analytics hardware, software, and training for both developers and users, all of which require good management vision and leadership (Kurz & Anandarajan, 2021).

5. CONCLUSION

In the digital technology era, BDA provides vital support to decision-makers with a comprehensive understanding of the supply chain processes and the market environment. Analytic applications include tools for retrieving past and current data from various channels, including social media platforms, third-party cloud platforms, IoT devices, and business transactions. They can forecast market trends and flag the risks at various points in the supply chains (Darvazeh et al., 2020). The resulting insights enable different levels of management to address operational problems and take corrective measures promptly to raise enterprise performance levels significantly. A company can leapfrog its rivals and generate a sustained competitive edge with effective supply chain analytics.

BDA can be applied in SCM by analyzing both historical and real-time data using different analytics: descriptive, diagnostic, predictive, and prescriptive analytics. Descriptive and diagnostic analytics utilize historical data to identify patterns and understand errors or problems in designated areas. Predictive analytics has wide-ranging applications in SCM, such as predictive maintenance and prediction of cyberattacks on the supply chain. Prescriptive analytics improves efficiency, reduces risk, and recommends what should be done next to achieve the desired outcome. The digital supply chain aims to make the next-generation supply chain smarter by using artificial intelligence to allow thinking or self-learning. While investments in technologies are expensive and critical, equally important are organizational agility in embracing new business models and developing human capital that spans top management, BDA developers, and end-users.

6. REFERENCES

Anitha, P., & Patil, M. M. (2018). A review on data analytics for supply chain management: A case study. International Journal of Information Engineering and Electronic Business (IJIEEB), 10(5), 30-39. https://doi.org/10.5815/ijieeb.2018.05.05

Brintrup, A. M., Woodall, P., Pearce, T., Mcfarlane, D., Pak, J., Ratiney, D., & Wichmann, P. (2019). Supply chain data analytics for predicting supplier disruptions: A case study in complex asset manufacturing. International Journal of Production Research. 58(11), 3330-3341. https://doi.org/10.1080/00207543.2019.1685705

Camm, J. D., Cochran, J. J., Fry, M. J., & Ohlmann, J. W. (2020). Business Analytics (4th ed.). Cengage Learning, Inc.

Calatayud, A., Mangan, J. and Christopher, M. (2019). The self-thinking supply chain.

Supply Chain Management: An International Journal, 24(1), 22-38.

Chapman, S. (2020). Walmart urges its suppliers to use IBM blockchain technology. Available at https://supplychaindigital.com/technology/walmart-urges-its-suppliers-use-ibm-blockchain-technology

Crowdstrike. (2021). What is a supply chain attack? Available at https://www.crowdstrike.com/cybersecurity-101/cyberattacks/supply-chain-attacks/

Darvazeh, S. S., Vanani, I. R., & Musolu, F. M. (2020). Big data analytics and its applications in supply chain management. In L. R.

Martínez, R. A. O. Rios, & M. D. Prieto (Eds.), New Trends in the Use of Artificial Intelligence for the Industry 4.0. IntechOpen. https://doi.org/10.5772/intechopen.89426

Ellis, S. (2020, August) IDC technology spotlight: The path to a thinking supply chain. IDC. Available at https://www.ibm.com/downloads/cas/PKQXXQJM

Fosso Wamba, S., Gunasekaran, A., Papadopoulos, T., and Ngai, E. (2018). Big data analytics in logistics and supply chain management. The International Journal of Logistics Management, 29(2), 478-484.

Hallikas, J., Immonen, M., & Brax, S. (2021). Digitalizing procurement: The impact of data analytics on supply chain performance. Supply Chain Management, 26(5), 629-646. http://dx.doi.org/10.1108/SCM-05-2020-0201

Kache, F., and Seuring, S. (2017). Challenges and opportunities of digital information at the intersection of big data analytics and supply chain management. International Journal of Operations & Production Management, 37(1), 10-36.

Kalsbeek, R. (2020). Where to start with the 4 types of analytics. Available at https://iterationinsights.com/article/where-to-start-with-the-4-types-of-analytics/

Kurz, D. B., & Anandarajan, M. (2021). Digital Supply Chain Leadership: Reshaping Talent and Organizations. Taylor & Francis Group.

Lee, I., & Mangalaraj, G. (2022). Big data analytics in supply chain management: A systematic literature review and research directions. Big Data and Cognitive Computing, 6(1), 17. https://doi.org/10.3390/bdcc6010017

Leveling, J., Edelbrock, M., Otto, B. (2014). Big data analytics for supply chain management. 2014 IEEE International Conference on Industrial Engineering and Engineering Management, 918–922. https://doi.org/10.1109/IEEM.2014.7058772

Maheshwari, S., Gautam, P., & Jaggi, C. K. (2021). Role of big data analytics in supply chain management: Current trends and future perspectives. Journal of Manufacturing Systems, 59(6), 1875–1900. https://doi.org/10.1080/00207543.2020.1793011

Maruti Techlabs. (N.D.). 9 ways machine learning can transform supply chain management. Available at https://marutitech.com/machine-learning-in-supply-chain/#Companies_Using_Machine_Learning_to_Improve_Their_Supply_Chain_Management

McCue, L. (2020, December 16). Supply chain analytics: What it is and why it matters. Oracle NetSuite. Available at https://www.netsuite.com/portal/resource/articles/erp/supply-chain-analytics.shtml

Melnyk, S.A., Schoenherr, T., Speier-Pero, C., Peters, C., Chang, J. F., & Friday, D. (2022). New challenges in supply chain management: Cybersecurity across the supply chain. International Journal of Production Research, 60(1), 162–183.

Shamout, M. D. (2019). Does supply chain analytics enhance supply chain innovation and robustness capability? Organizacija, 52(2), 95-106. https://doi.org/10.2478/orga-2019-0007

Vaughan-Nichols, S. (2021). SolarWinds: The more we learn, the worse it looks. Available at https://www.zdnet.com/article/solarwinds-the-more-we-learn-the-worse-it-looks/

Wang, S.-C., Tsai, Y.-T., & Ciou, Y.-S. (2020). A hybrid big data analytical approach for analyzing customer patterns through an integrated supply chain network. Journal of Industrial Information Integration, 20, 100177. https://doi.org/10.1016/j.jii.2020.100177

Xu, J., Pero, M. E. P., Ciccullo, F., & Sianesi, A. (2021). On relating big data analytics to supply chain planning: Towards a research agenda. International Journal of Physical Distribution & Logistics Management, 51(6), 656-682. http://dx.doi.org/10.1108/IJPDLM-04-2020-0129