Applying Big Data Analytics to Retail for Improved Supply Chain Visibility

Alshaimaa A. Tantawy[1], Zenat Ahmed[2], Mahmoud M. Ali[*3]

[1,2,3] Faculty of computers and Informatics, Zagazig University, Zagazig, 44519, Egypt
Emails: Eatantawi @ fci.zu.edu.eg ; zenatahmed@zu.edu.eg; mmsabe@zu.edu.eg
*Correspondence: mmsabe@zu.edu.eg

Abstract

Retail supply chains generate huge volumes of data that can provide valuable insights if analyzed effectively. This paper explores how retailers can leverage Big Data analytics techniques on supply chain data to gain enhanced visibility into their operations. We examine three use cases of data-driven supply chain visibility: (1) predictive replenishment to anticipate future demand and optimize inventory levels; (2) personalized assortment optimization to tailor product selections for local customer segments; and (3) optimized order fulfillment to improve delivery times and reduce transportation costs. We analyze how retailers can apply machine learning algorithms and statistical analysis on point-of-sale data, inventory data, customer data and external data sources to uncover hidden patterns and drive data-driven decisions in these areas. The results include reduced excess inventory, fewer stock-outs, higher in-store product availability, lower fulfillment costs and improved customer experience. Data-driven supply chain visibility allows retailers to transition from a reactive, speculative business model to a predictive, personalized model that enhances competitiveness.

Keywords: Big Data; Supply Chain; Data Analytics; Optimization

1. Introduction

Retail companies today operate in a highly competitive environment with constant pressure to improve sales, reduce costs, and increase customer satisfaction. One of the keyways for retailers to gain a competitive edge is to optimize their supply chain management (SCM). With more digital data being generated from sources like point-of-sale systems, e-commerce platforms, and customer loyalty programs, retailers have new opportunities to gain valuable insights into their supply chain operations [1-3].

Big Data analytics techniques can help retailers achieve "supply chain visibility" - a comprehensive understanding of the flow of products from suppliers to customers. With improved visibility into their supply chains, retailers can make better decisions around inventory management, demand forecasting, product placement, and logistics optimization. This can help reduce excess inventory, anticipate, and respond to stock-outs, optimize product availability across the network, and enhance the overall customer experience [4-5].

In this paper, we explore how retailers can leverage Big Data analytics on multiple data sources to gain enhanced visibility into their supply chains, uncover hidden insights, and drive operational improvements. Specifically, we look at use cases around predictive replenishment, personalized assortment optimization, and optimized order fulfillment. The benefits of data-driven supply chain visibility are significant, especially as customers increasingly expect omni-channel shopping experiences with seamless product availability. Retailers that can successfully leverage analytics to transform their supply chain practices will be poised for sustainable success in the digital commerce era.

DOI: https://doi.org/10.54216/AJBOR.040104
Received: January 15, 2021 Accepted: August 14, 2021

The paper has four more sections. The related works are discussed in section 2 and the methodology is described in section 3. The results and discussions are explained in section 4 and the conclusion is summarized at section5.

2. Related Works

The application of Big Data Analytics in SCM has become a topic of significant interest in recent years. A literature review of studies on this topic is presented in this section to reveal that the use of Big Data Analytics can improve supply chain visibility, which can lead to better decision-making, and reduced costs. Waller and Fawcett [5] studied the transformative potential of data science, predictive analytics, and Big Data in the field of SCM. They argued that these technologies can revolutionize supply chain design and management by providing real-time visibility, enabling data-driven decision-making, and facilitating collaboration among stakeholders. They provided a set of examples on how these technologies are being applied in various industries, including retail, healthcare, and logistics. Barratt and Oke [6] examined the antecedents of supply chain visibility in retail supply chains from a resource-based theory perspective. They argued that supply chain visibility is a critical resource for organizations and can provide strategic advantages such as improved responsiveness, reduced costs, and enhanced customer service. They identified the key antecedents of supply chain visibility, including information technology capabilities, supplier relationships, and internal processes. Dubey et al [8] argued that these technologies could help organizations to better understand and address sustainability issues by providing real-time data, identifying patterns and trends, and predicting future outcomes. They provided examples of how big data and predictive analytics are being applied in various industries, including energy, transportation, and agriculture, to enhance sustainability. In [12], Zhong et al contributes to the discipline of SCM by exploring the challenges meeting the big data in SCM, and argued how it can provide significant benefits in the service and manufacturing sectors, including improved visibility, enhanced decision-making, and increased efficiency. Wamba et al [13] argued that big data analytics and supply chain ambidexterity can improve supply chain performance by enhancing responsiveness, reducing costs, and increasing efficiency. They provided experimental confirmation to support the positive effects of big data analytics and supply chain ambidexterity on supply chain performance, and the moderating effect of environmental dynamism on these relationships. Chen et al [17] studied how big data analytics created a value in SCM by providing real-time insights, enhancing decision-making, and improving collaboration among supply chain partners. They provided empirical evidence to support the positive effects of big data analytics on value creation in SCM, and identifies the key factors that influence these effects, including data quality, data integration, and analytical capabilities. They also discussed the implications of their findings for practice and provided recommendations for organizations seeking to leverage big data analytics to enhance the value they create in their supply chains. Gunasekaran et al. [18] provided empirical evidence to support the positive effects of big data and predictive analytics on supply chain and organizational performance, and identifies the key factors that influence these effects, including data quality, analytical capabilities, and organizational culture. They also discussed the challenges and opportunities associated with the adoption of these technologies and provided recommendations for organizations seeking to leverage them effectively. Bag et al [19] explored the potential of big data analytics as an operational excellence approach to enhance sustainable supply chain performance. They argued that big data analytics can provide significant benefits in sustainable SCM, including enhanced visibility, improved decision-making, and increased efficiency, which can lead to improved environmental and social performance. They provided empirical evidence to support the positive effects of big data analytics on sustainable supply chain performance, and identifies the key factors that influence these effects, including data quality, analytical capabilities, and stakeholder collaboration.

3. Methodology

This section provides a detailed description of the research design, data collection methods, data analysis techniques, and the tools used in the study. This section provides insights into how Big Data Analytics was applied in retail to improve supply chain visibility, the data sources used, and the analytical tools used to measure the impact of the approach.

A. Case Study

BigMart Sale Data is a well-known and widely used case study in the field of Big Data Analytics. The case study involves the analysis of sales data from a fictional retail chain called BigMart. The dataset contains information on various product attributes such as weight, visibility, and type, as well as store attributes such as size and location. The objective of the case study is to use Big Data Analytics techniques to analyze the data and gain insights into factors that affect sales and profitability [6-10]. The BigMart Sale Data case study is relevant to our study as it

DOI: https://doi.org/10.54216/AJBOR.040104
Received: January 15, 2021 Accepted: August 14, 2021

provides a practical example of how Big Data Analytics can be applied in the retail industry to improve supply chain visibility. The case study highlights the importance of having access to high-quality data and the need for

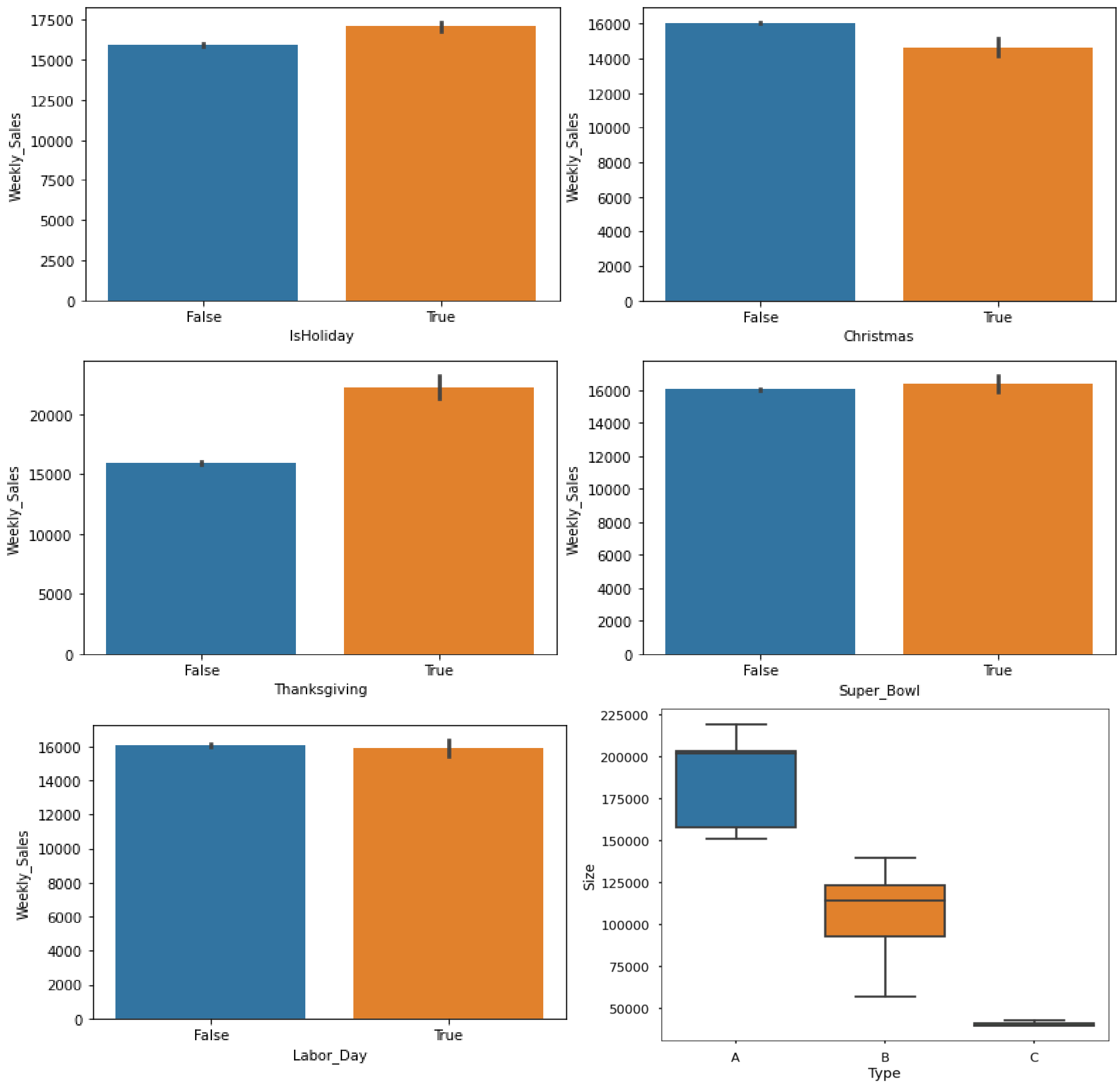

Figure 1: Visualization of retail data distribution according to different criteria.

appropriate analytical tools and techniques to extract insights from the data.

Additionally, the case study demonstrates how the application of Big Data Analytics can help organizations to make data-driven decisions, optimize their supply chain operations, and improve customer satisfaction. Our study uses the BigMart Sale Data case study as a reference point to illustrate the potential of Big Data Analytics in retail SCM. By analyzing the data and applying appropriate analytical techniques, we could identify patterns and trends in the data that could inform decision-making and improve supply chain visibility. Additionally, we could use the case study to highlight the challenges and opportunities associated with the adoption of Big Data Analytics in retail SCM and provide recommendations for organizations seeking to leverage these technologies effectively.

DOI: https://doi.org/10.54216/AJBOR.040104
Received: January 15, 2021 Accepted: August 14, 2021

B. Preprocessing

Preprocessing techniques are discussed in this section because they are essential in the context of Big Data Analytics to ensure that the data is of high quality and consistency, and to prepare it for analysis. For our case study, the following preprocessing techniques can be applied to improve supply chain visibility. First, data cleaning is applied through removing or correcting any errors, inconsistencies, or missing data in the sales data. For example, incorrect or missing product codes, incorrect pricing information, or missing customer data can be corrected or imputed. Second, data integration is applied, and it involves combining data from different sources to create a single, unified dataset. In the context of retail sales data, this may involve integrating data from different store locations, different sales channels, or different suppliers. Third, data transformation is then applied, and it involves converting data into a more suitable format for analysis [10-15]. By applying these preprocessing techniques to retail sales data, organizations can improve the quality and consistency of the data and prepare it for analysis using Big Data Analytics techniques. This can help organizations to gain insights into supply chain operations, identify areas for improvement, and optimize their supply chain to improve visibility and customer satisfaction [16-17].

C. Predictive Analysis

ARIMA (Autoregressive Integrated Moving Average) is a time series forecasting method that is usually applied during the Big Data Analytic to predict future sales and demand. Hence, it is widely used in the retail industry to forecast sales and demand for products, which can help organizations optimize their inventory levels, improve supply chain efficiency, and enhance customer satisfaction. ARIMA models work by analyzing the historical sales data to identify patterns and trends, and then using this information to make predictions about future sales. The model consists of three components: the autoregressive (AR) component, the integrated (I) component, and the moving average (MA) component. The AR component considers the impact of past sales on future sales, the MA component considers the impact of past errors on future sales, and the I component considers any trends or seasonality in the data. The main procedure of training ARIMA is shown in Algorithm 1.

Algorithm 1. Training process for ARIMA.

Input: Historical Retail Sales. D
Output: Prediction Model
1: D_{series}= series (D)
2: size=$int(len(D_{series}) * 0.8)$
3: D_{train}, D_{val}= D_{series}[0: size], D_{series}[size: len (D_{series})] ←split into training and Validation set
4: p−value← Dickey-Fuller test (95%)
5:auto_arima(D_{train}, start_p=0, start_q=0, seasonal=True, trace=True, d=1, D=1, error_action='warn', suppress_warnings=True, n_fits=30)← fit ARIMA method
6: AIC← compute Akaike's Information Criterion
7: RMSE← evaluate arima (D_{val})

In our case study, ARIMA is applied to large volumes of sales data to make predictions at a granular level, such as by store location, product relations, or sales channel. This allows organizations to optimize their supply chain operations and inventory management based on detailed sales forecasts. However, there are some limitations to using ARIMA models for predicting retail sales data. For example, ARIMA models assume that the data is stationary, meaning that the statistical properties of the data do not change over time. Sales data can be influenced by various factors such as seasonality, trends, and external events, which can affect the accuracy of the forecasts. Additionally, ARIMA models may not be suitable for predicting sales data for new or innovative products, as there may not be enough historical data to provide accurate forecasts.

Exponential smoothing is another popular predictive big data tool to predict future sales and demand. Exponential smoothing works by calculating a weighted average of past sales data, with more recent data being given more weight than older data. The method assumes that the most recent sales data is the most relevant for predicting future sales. In our case study, we apply exponential smoothing to large volumes of sales data to make predictions at a granular level, as with ARIMA. One advantage of exponential smoothing over other forecasting methods such as ARIMA is that it is relatively simple to implement and can be easily automated. This makes it a useful tool for

DOI: https://doi.org/10.54216/AJBOR.040104
Received: January 15, 2021 Accepted: August 14, 2021

organizations that want to make quick and accurate sales forecasts without the need for extensive statistical knowledge or expertise. However, like ARIMA, exponential smoothing assumes that the underlying sales data is stationary and does not consider external factors that may influence sales, such as seasonality, trends, or changes in consumer behavior. Additionally, exponential smoothing may not be suitable for predicting sales data for new or innovative products, as there may not be enough historical data to provide accurate forecasts.

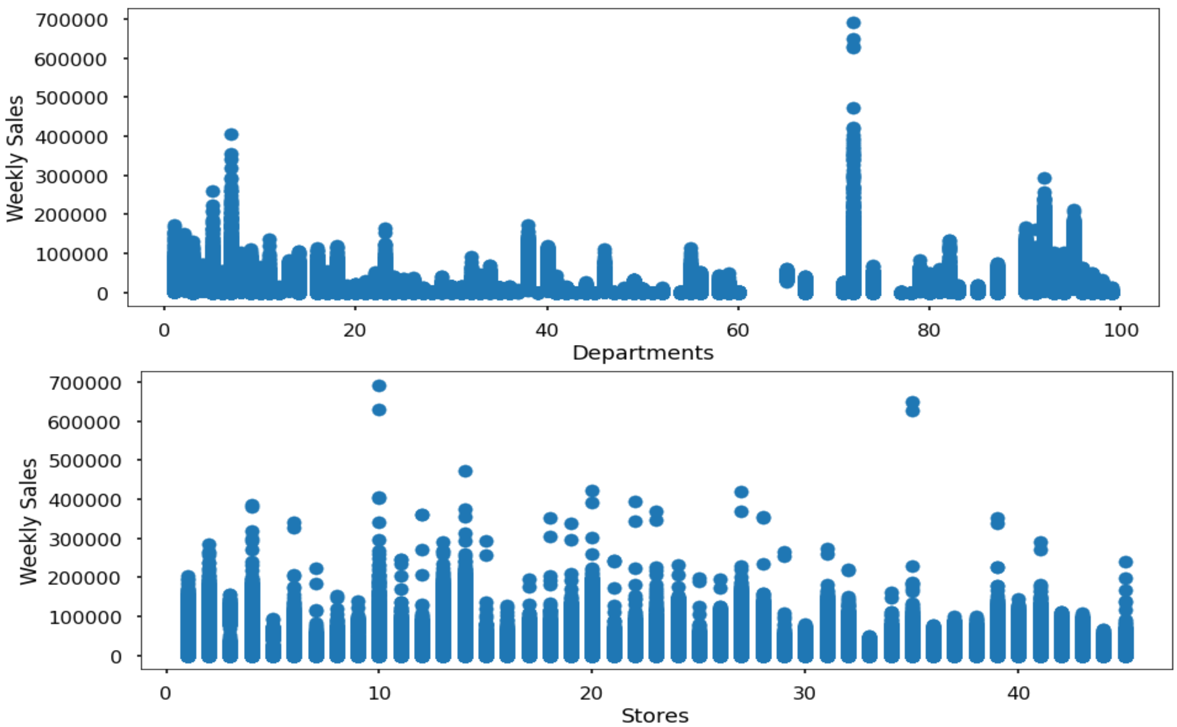

Figure 2. visualization of retail data distribution across departments and stores.

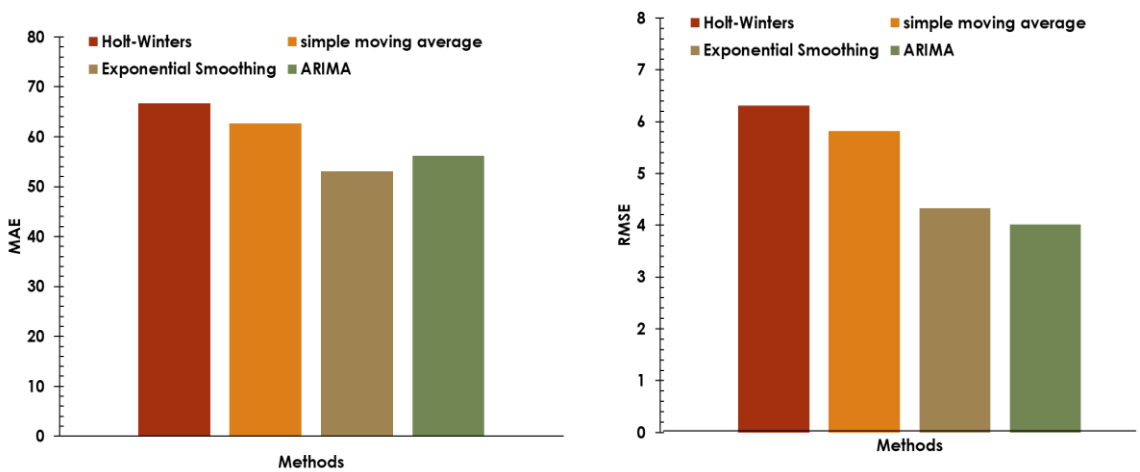

Figure 3: visualization of prediction performance on different predictors on our case study.

4. Results and Discussions

The performance of the above prediction tools is assessed with the following indicators:

DOI: https://doi.org/10.54216/AJBOR.040104
Received: January 15, 2021 Accepted: August 14, 2021

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (y_a - y_f)^2}{N}} \tag{1}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N} |y_a - y_f| \tag{2}$$

Exploratory data analysis (EDA) is a crucial step in analyzing retail sales data to gain insights into sales distribution and identify patterns and trends. In Figure 1, we perform EDA by visualizing sales distribution based on different criteria. It is noted that the data is well distributed for all criteria except for Type criterion. By visualizing sales distribution based on different criteria, organizations can gain insights into sales patterns and trends that may not be apparent from the raw data. This can help organizations to make data-driven decisions, optimize their supply chain operations, and improve customer satisfaction. There are several tools and techniques available for visualizing sales

Figure 2: Visualization of prediction plots for ARIMA (upper) and Exponential smoothing algorithm (down).

data, such as bar charts, line charts, scatter plots, and heat maps. Additionally, interactive dashboards and visualization tools can be used to provide real-time insights into sales data and enable organizations to make quick and informed decisions. Further, we show the distribution of retail sales across different departments and stores in Figure 2.

Experimental comparison is then conducted on our case study of retail data to analyze and select the most accurate and reliable forecasting methods. Figure 3 shows the comparison of different prediction methods, namely ARIMA, Exponential Smoothing, simple moving average, and holt's winter, aiming to help organizations to identify the most appropriate and effective method for their specific needs. It's important to note that the performance of prediction models may vary depending on the specific characteristics of the model and the context in which they are used. Therefore, it's recommended to experiment with multiple prediction methods and evaluate their accuracy on different sets of data to ensure that the chosen method is appropriate and effective for the specific needs of the organization.

DOI: https://doi.org/10.54216/AJBOR.040104
Received: January 15, 2021 Accepted: August 14, 2021

To further validate the above results, generate a prediction vs actual sales plot for different prediction models, as shown in Figure 4. The training set is used to fit different prediction models such as ARIMA, and Exponential Smoothing. Then, we use the fitted models to make predictions for the test data. Then, we plot the actual sales data and the predicted sales data for each model on the same graph to compare them visually.

5. Conclusion

This study investigates the role of big data analytics to improve supply chain visibility in the retail industry through providing significant benefits in sustainable SCM, including enhanced visibility, improved decision-making, and increased efficiency, which can lead to improved environmental and social performance. By applying appropriate preprocessing techniques and using prediction models such as ARIMA, Exponential Smoothing, and others, organizations can gain insights into sales patterns and trends, optimize inventory levels, and improve supply chain efficiency and customer satisfaction. Our findings recommend that organizations seeking to leverage Big Data Analytics in their SCM practices should prioritize data quality, analytical capabilities, and stakeholder collaboration. Finally, we found that visualizing sales distribution based on different criteria can be a useful tool in exploratory data analysis, enabling organizations to gain insights into sales patterns and trends that may not be apparent from the raw data.

References

[1] Leveling J, Edelbrock M, Otto B. Big data analytics for supply chain management. In2014 IEEE international conference on industrial engineering and engineering management 2014 Dec 9 (pp. 918-922). IEEE.

[2] Tiwari S, Wee HM, Daryanto Y. Big data analytics in supply chain management between 2010 and 2016: Insights to industries. Computers & Industrial Engineering. 2018 Jan 1;115:319-30.

[3] Kache F, Seuring S. Challenges and opportunities of digital information at the intersection of Big Data Analytics and supply chain management. International journal of operations & production management. 2017 Jan 3.

[4] Tan KH, Zhan Y, Ji G, Ye F, Chang C. Harvesting big data to enhance supply chain innovation capabilities: An analytic infrastructure based on deduction graph. International Journal of Production Economics. 2015 Jul 1;165:223-33.

[5] Waller MA, Fawcett SE. Data science, predictive analytics, and big data: a revolution that will transform supply chain design and management. Journal of Business Logistics. 2013 Jun;34(2):77-84.

[6] Barratt M, Oke A. Antecedents of supply chain visibility in retail supply chains: a resource-based theory perspective. Journal of operations management. 2007 Nov 1;25(6):1217-33.

[7] Arunachalam D, Kumar N, Kawalek JP. Understanding big data analytics capabilities in supply chain management: Unravelling the issues, challenges and implications for practice. Transportation Research Part E: Logistics and Transportation Review. 2018 Jun 1;114:416-36.

[8] Dubey R, Gunasekaran A, Childe SJ, Papadopoulos T, Luo Z, Wamba SF, Roubaud D. Can big data and predictive analytics improve social and environmental sustainability?. Technological Forecasting and Social Change. 2019 Jul 1;144:534-45.

[9] Kamble SS, Gunasekaran A, Gawankar SA. Achieving sustainable performance in a data-driven agriculture supply chain: A review for research and applications. International Journal of Production Economics. 2020 Jan 1;219:179-94.

[10] Hazen BT, Boone CA, Ezell JD, Jones-Farmer LA. Data quality for data science, predictive analytics, and big data in supply chain management: An introduction to the problem and suggestions for research and applications. International Journal of Production Economics. 2014 Aug 1;154:72-80.

[11] Delen D, Hardgrave BC, Sharda R. RFID for better supply-chain management through enhanced information visibility. Production and operations management. 2007 Sep 10;16(5):613-24.

[12] Zhong RY, Newman ST, Huang GQ, Lan S. Big Data for supply chain management in the service and manufacturing sectors: Challenges, opportunities, and future perspectives. Computers & Industrial Engineering. 2016 Nov 1;101:572-91.

[13] Wamba SF, Dubey R, Gunasekaran A, Akter S. The performance effects of big data analytics and supply chain ambidexterity: The moderating effect of environmental dynamism. International Journal of Production Economics. 2020 Apr 1;222:107498.

DOI: https://doi.org/10.54216/AJBOR.040104
Received: January 15, 2021 Accepted: August 14, 2021

[14] ur Rehman MH, Chang V, Batool A, Wah TY. Big data reduction framework for value creation in sustainable enterprises. International journal of information management. 2016 Dec 1;36(6):917-28.
[15] Dash R, McMurtrey M, Rebman C, Kar UK. Application of artificial intelligence in automation of supply chain management. Journal of Strategic Innovation and Sustainability. 2019;14(3):43-53.
[16] Akter S, Wamba SF. Big data analytics in E-commerce: a systematic review and agenda for future research. Electronic Markets. 2016 May;26:173-94.
[17] Chen DQ, Preston DS, Swink M. How the use of big data analytics affects value creation in supply chain management. Journal of management information systems. 2015 Oct 2;32(4):4-39.
[18] Gunasekaran A, Papadopoulos T, Dubey R, Wamba SF, Childe SJ, Hazen B, Akter S. Big data and predictive analytics for supply chain and organizational performance. Journal of Business Research. 2017 Jan 1;70:308-17.
[19] Bag S, Wood LC, Xu L, Dhamija P, Kayikci Y. Big data analytics as an operational excellence approach to enhance sustainable supply chain performance. Resources, Conservation and Recycling. 2020 Feb 1;153:104559.

DOI: https://doi.org/10.54216/AJBOR.040104
Received: January 15, 2021 Accepted: August 14, 2021