Optimizing Supply Chain Management with Retail Analysis Technique

Shereen Zaki [1], Mahmoud M. Ismail [2], Mahmoud M. Ibrahim [*3]
[1,2,3] Faculty of computers and Informatics, Zagazig University, Zagazig, 44519, Egypt
Emails: SZSoliman@fci.zu.edu.eg; mmsabe@zu.edu.eg; mmsba@zu.edu.eg
*Correspondence: mmsba@zu.edu.eg

Abstract

The success of any business organization depends on efficient supply chain management (SCM) that ensures the timely delivery of products to customers. The retail analysis technique (RAT) has emerged as a powerful tool for optimizing SCM by enabling companies to identify and analyze sales trends, inventory levels, and customer demand patterns. This paper explores the use of retail analysis techniques in SCM and how they can be leveraged to enhance the efficiency and effectiveness of supply chain operations. We then present a unified RAT framework that integrates data mining, predictive analytics, and machine learning for optimizing SCM. Additionally, the paper presents a case study of a company that has successfully implemented retail analysis techniques in its SCM practices. The case study highlights the specific strategies and tactics that the company used to optimize its supply chain operations and achieve significant improvements in efficiency and performance. Our analysis shows that RAT is an effective technique for optimizing SCM, and companies that implement it can gain a competitive advantage in the marketplace.

Keywords: Supply Chain; Retail Analysis; Optimization

1. Introduction

The supply chain is a critical component of any business, as it involves managing the flow of goods and services from suppliers to consumers. In recent years, the increasing complexity of supply chains and the rise of e-commerce have made it more challenging to optimize supply chain management (SCM). One approach that has gained popularity is the use of retail analysis techniques to gain insights into customer behavior and preferences, which can improve supply chain efficiency and customer satisfaction [1-2].

Retail analysis involves analyzing large amounts of data from various sources, such as sales transactions, customer feedback, and social media, to identify patterns and trends in customer behavior. This information can be used to optimize various aspects of the supply chain, from inventory management to logistics and delivery. Business intelligence (BI) is a powerful tool for retailers seeking to gain valuable insights from their data to make informed business decisions. Retail analysis using BI involves the collection, integration, analysis, and visualization of data from various sources such as sales transactions, customer behavior, and inventory levels. BI tools can generate detailed reports, dashboards, and visualizations that provide retailers with a comprehensive view of their operations, including sales trends, customer preferences, and product performance [3-4]. By using BI for retail analysis, retailers can gain a deeper understanding of their customer's needs and preferences, optimize inventory levels, reduce costs, and increase sales. Moreover, BI can help retailers to identify new opportunities for growth and innovation, allowing them to stay competitive in an ever-changing market [5-6].

In this paper, we explore the use of retail analysis techniques to optimize SCM. We propose a unified RAT framework that acts as a comprehensive approach to SCM optimization through leveraging data mining, predictive analytics, and ML, to empower businesses to make informed decisions and gain a competitive advantage in the market. Next, we present a case study of a company, on which we successfully implement our retail analysis framework to optimize its supply chain. We describe the specific challenges the company faced and how our framework solved it, including the

DOI: https://doi.org/10.54216/AJBOR.020202
Received: March 20, 2021 Accepted: July 05, 2021

use of predictive analytics to optimize inventory management and the use of social media analysis to improve customer engagement.

2. Related Works

In this section, we provide a comprehensive overview of the existing literature related to optimizing SCM with retail analysis techniques. For example, the authors of [7] provided a comprehensive review of the literature on sustainable supply chain network design from an optimization perspective. They presented a detailed analysis of the key themes, trends, and challenges in sustainable supply chain design and identifies the optimization techniques and models used to address them. They also proposed a framework for sustainable supply chain network design that integrates economic, environmental, and social considerations. The authors of [8] examined the trends and challenges of supply chain digitization and the integration of knowledge management in this context. They provided a comprehensive review of the literature on supply chain digitization and knowledge management and propose a conceptual framework that integrates these two domains. They highlighted the benefits and challenges of digital technologies in SCM and discussed the implications for knowledge management. The authors argued that effective knowledge management is essential for successful supply chain digitization and proposed strategies for integrating knowledge management into the digitization process. The authors of [9] presented a comprehensive review of the literature on biomass-for-bioenergy supply chains by reviewing the key challenges in designing and managing biomass-for-bioenergy supply chains and reviewing the methods and models used to address these challenges. They discussed the importance of considering sustainability and environmental factors in the design and management of biomass-for-bioenergy supply chains and highlighted the need for multidisciplinary approaches. They also identify the gaps and limitations in the existing literature and propose future research directions. The authors of [10] presented an innovative approach for using lean Six Sigma methodologies in global SCM. They discussed the benefits of integrating Industry 4.0 and Industrial Internet of Things (IIoT) technologies in SCM and proposed a framework for implementing Lean Six Sigma in this context. They provided a detailed explanation of the methodology, including the use of data analytics, process improvement techniques, and continuous monitoring and feedback.

The authors of [11] provided a comprehensive review of the literature on sustainable SCM (SSCM) over a 15-year period. They reviewed of the key themes, trends, and challenges in SSCM and identify the theoretical frameworks and empirical studies used to address them. They discussed the importance of sustainability in SCM and highlighted the need for integrated approaches that consider economic, social, and environmental factors. The authors of [12] presented a review of the challenges, opportunities, and future perspectives of using big data in SCM. They provided an overview of the key concepts and applications of big data analytics in the service and manufacturing sectors and discussed the challenges and opportunities associated with its use. The authors of [13] presented an optimization model for green SCM (GSCM) using a big data analytic approach. They proposed a mathematical model that considers economic, social, and environmental factors to optimize the design and operation of GSCM systems. They also described the big data analytic approach used to collect, process, and analyze the data needed to support the optimization model. They argued that the proposed approach can help companies achieve a more sustainable and efficient supply chain while reducing environmental impacts. The authors of [14] reviewed the literature on competitive supply chain network design (CSCND) and discussed the solution techniques and applications of this approach. They argued that CSCND can help companies achieve a competitive advantage by optimizing their supply chain network design and operation. The authors of [15] provided a comprehensive review of the literature on hydrogen supply chain network design (HSCND) from an optimization perspective. They highlighted the potential benefits of HSCND, such as reducing greenhouse gas emissions and improving energy security. They argued that optimization models can help companies and policymakers make informed decisions about the design and operation of HSCND systems.

3. Case Study

In this section, we present a real-world example of a company that implemented our proposed solution for optimizing SCM. The Big Mart Sales dataset is used as a popular case study in the field of SCM and retail analytics. The dataset contains sales data for over 1500 products sold in 10 different stores of a retail chain, Big Mart. The objective of the case study is to develop a model that can predict the sales of different products based on various factors such as store location, product visibility, and promotional offers. The Big Mart Sales case study is used in our work as an illustrative example to demonstrate the effectiveness of our proposed solution. We can use the dataset to develop and test our optimization models for SCM and retail analytics. For instance, we can use the dataset to predict the demand for different products at different stores, optimize the inventory levels for each product, and determine the optimal

DOI: https://doi.org/10.54216/AJBOR.020202
Received: March 20, 2021 Accepted: July 05, 2021

allocation of products to different stores based on their sales potential. Moreover, we can also use the dataset to evaluate the performance of our models and compare them with existing models in the literature. This would help us demonstrate the superiority of our proposed solution and its potential for improving supply chain efficiency and profitability. The Big Mart Sales case study provides a useful platform for us to demonstrate the effectiveness of our proposed solution in the context of SCM and retail analytics.

4. Proposed Technique

In this section, we present our solution for optimizing SCM using retail analysis techniques. Our proposed solution is designed to leverage the power of retail analysis to enhance the efficiency and effectiveness of supply chain operations. We will explain the key components of our proposed solution, including the analytical techniques, and visualization tools used to optimize supply chain performance.

1. Prepare Sales Data

Data preparation is applied here as a crucial step in any data analysis project, and it is particularly important when working with retail data such as the Big Mart Sales dataset. To ensure that the data is of high quality and that it can be effectively analyzed, several preprocessing steps need to be applied. This step involves removing any invalid or missing data, such as incorrect product names, null values, or outliers [16]. Then, we transform the data into a format that is more suitable for analysis. We convert categorical variables such as product categories or store locations into numerical values or scale numerical variables such as prices or weights to a common range. In our retail case study, data normalization is applied to the sales data to ensure that all the products are evaluated on a common scale.

$$x_i^* = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \tag{1}$$

The pseudocode of our data preparation procedure is given as follows:

Algorithm 1. The applied preprocessing for the big mart data.
1: Read the Big Mart Sales data from a CSV file.
2: Remove any missing values from the data.
3: Convert any categorical variables to numerical values, such as converting "Low Fat" to 0 and "Regular" to 1.
4: Create any new variables that may be useful for analysis, i.e., mean item visibility or the age of the outlet.
5: Remove any unnecessary variables from the data, such as the item and outlet identifiers.
6: Standardize numerical variables so that they have zero mean and unit variance.
7: Encode any remaining categorical variables using one-hot encoding, which creates new binary variables for each category.
8: Split the data into training and testing sets if necessary.

2. Predictive modeling

XGBoost (Extreme Gradient Boosting) is a popular machine learning algorithm that we investigate for optimizing SCM. It is a supervised learning algorithm that is capable of handling both regression and classification problems. XGBoost works by building a decision tree ensemble, where each tree is trained to predict the residual errors of the previous tree. The XGBoost algorithm works by iteratively adding decision trees to an ensemble model in a way that minimizes a specific loss function [17]. At each iteration, the algorithm calculates the gradient and the second-order derivative of the loss function with respect to the current ensemble predictions. These values are then used to fit a decision tree that predicts the negative gradient, which is the residual error of the current model. The tree is then added to the ensemble, and the process is repeated until the desired number of trees is reached.

The pseudocode of the XGBoost-based SCM optimization procedure is given below:

DOI: https://doi.org/10.54216/AJBOR.020202
Received: March 20, 2021 Accepted: July 05, 2021

Algorithm 2. Pseudocode of the XGBoost-based SCM optimization

Inputs: train set $\{(x_j, y_j)\}_{j=1}^{M}$, number of weak learners N, objective function $L(y, F(x))$, learning rate (α)

1: Do Initialization $F_i^{(0)}(x) = \underset{\vartheta}{\mathrm{argmin}} \sum_{j=1}^{M} L(y_j, \vartheta)$

2: **for** k=1 to N

3: Calculate hessians: $h_k(x_j) = \left[\frac{\partial^2 L(y_j, f(x_j))}{\partial f(x_j)^2}\right]_{f(x)=f_{k-1}(x)}$

4: Calculate gradients: $g_k(x_j) = \left[\frac{\partial L(y_j, f(x_j))}{\partial f(x_j)}\right]_{f(x)=f_{k-1}(x)}$

5: Fitting tree on the training set $\left\{\left(x_j, -\frac{g_k(x_j)}{h_k(x_j)}\right)\right\}_{j=1}^{M}$ thru optimizing $f_k(x) = \alpha\phi_k(x)$

6: where $\phi_k = \underset{\phi \in \Phi}{\mathrm{argmin}} \sum_{j=1}^{M} \frac{1}{2} h_k(x_j)\left[-\frac{g_k(x_j)}{h_k(x_j)} - \phi(x_j)\right]^2$

7: $F_i^{(k)}(x) = F_i^{(k-1)}(x) + f_k(x)$

8: **end**

9: **Return** $F_i^{(N)} = \sum_{k=0}^{N} f_k(x)$

One of the main advantages of XGBoost is its ability to handle large and complex datasets with high-dimensional features, which is often the case in SCM applications.

5. Results

In this section, we will analyze the data collected and evaluate the effectiveness of our proposed solution in optimizing SCM using retail analysis techniques. The results section is critical as it provides insights into the practical application and impact of the proposed solution.

Statistical analysis is conducted to explore and understand the Big Mart Sales dataset, as shown in Table 1, which involves calculating summary statistics for each variable in the dataset, such as mean, median, mode, standard deviation, and range. These statistics are made to identify any outliers or unusual values in the data, as well as provide a general overview of the distribution of the variables.

Table 1: Descriptive statistics for BIG-MART sales.

	Item_Weight	Item_Visibility	Item_MRP	Outlet_Establishment_Year	Item_Outlet_Sales
count	7060.0	8523.0	8523.0	8523.0	8523.0
mean	12.9	0.1	141.0	1997.8	2181.3
std	4.6	0.1	62.3	8.4	1706.5
min	4.6	0.0	31.3	1985.0	33.3
25%	8.8	0.0	93.8	1987.0	834.2
50%	12.6	0.1	143.0	1999.0	1794.3
75%	16.9	0.1	185.6	2004.0	3101.3
max	21.4	0.3	266.9	2009.0	13087.0

DOI: https://doi.org/10.54216/AJBOR.020202
Received: March 20, 2021 Accepted: July 05, 2021

Box plots are visualized in Figure 1 as a useful visualization tool to visualize the distribution of sales for different

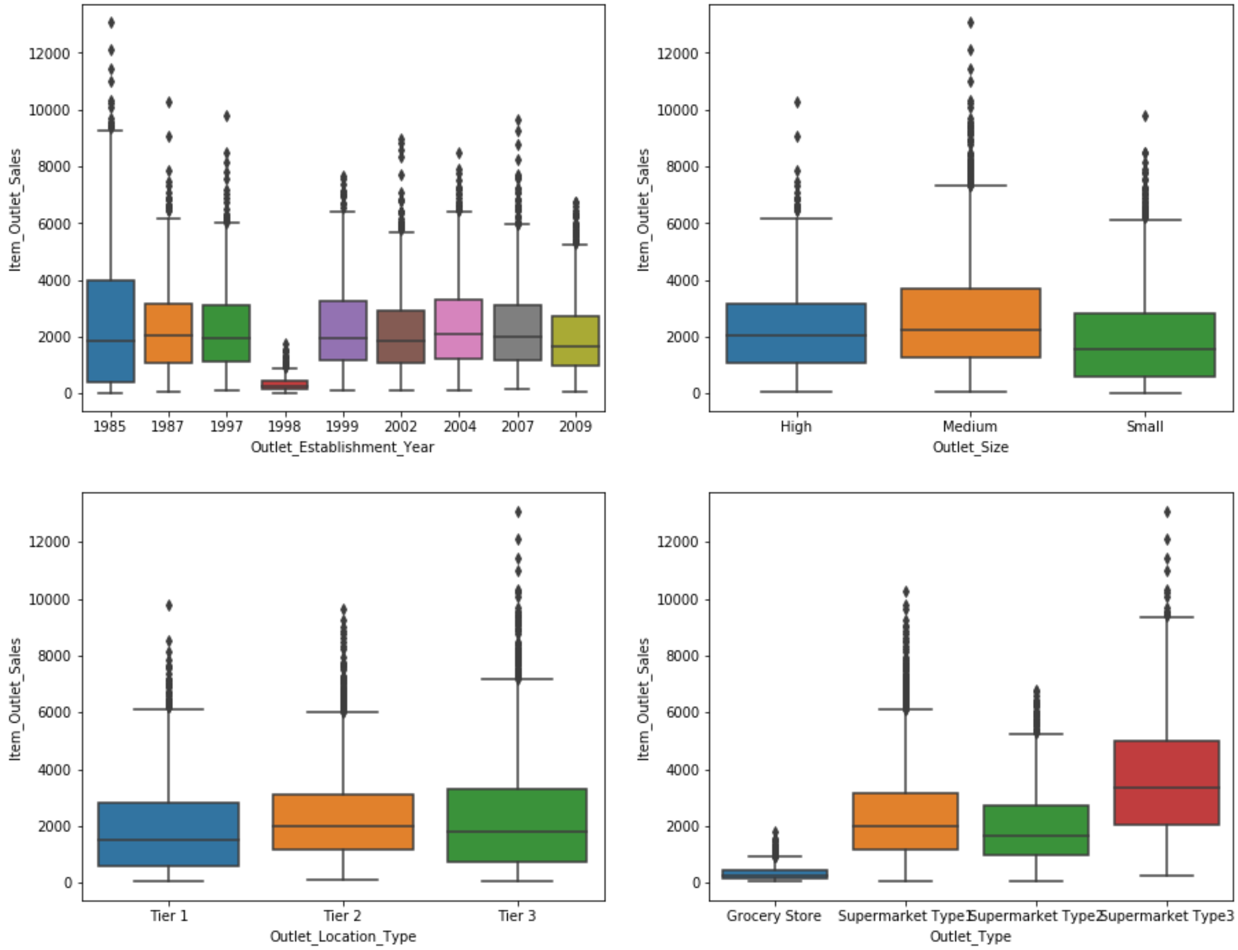

Figure 1: Visualization of box-blots of samples distribution in our case study.

products, categories, and stores. As shown, the data is divided into quartiles. The bottom of the box represents the first quartile, or the 25th percentile, while the top of the box represents the third quartile or the 75th percentile. The median, or the 50th percentile, is represented by a line in the box. The whiskers extend from the top and bottom of the box to the minimum and maximum values that are not considered outliers. These Box plots help us to identify outliers, skewness, and variability in the data. We also can compare the sales distribution of different outlet location types, outlet types, outlet sizes, and years. This can provide insights into which entities are performing well or poorly. The box plot for an entity that shows many outliers, or a wide range of values may indicate that the entity is not performing well and needs to be further analyzed. The box plot that shows a tight distribution with no outliers indicates that the product is performing well and can be further promoted.

DOI: https://doi.org/10.54216/AJBOR.020202
Received: March 20, 2021 Accepted: July 05, 2021

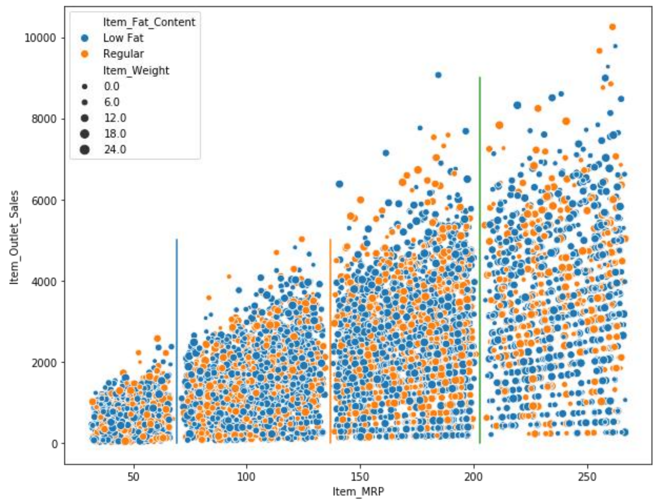

Figure 2: cluster visualization for ITEM_MRP column in our case study

Figure 2 provides a visual exploration of the ITEM_MRP column in our case study which contains prices that fall into clusters. Hence, we visualize these prices in bins for further processing. The indicated perpendicular lines are then used to divide data into proper bins as valued in the x-axis.

To visualize the fitting line plots for the predictions from XGBoost, Figure 3 plots the actual sales values against the predicted sales values, which gives us an idea of how well our model is performing in terms of predicting sales. We start by selecting a subset of data, such as a single product and store combination, and plotting the actual sales values on the y-axis and the predicted sales values on the x-axis. We then add a diagonal line to represent perfect prediction, where the actual and predicted sales values are equal. We can infer that our model is performing well, since we see a tight clustering of points around the diagonal line, indicating that the predicted sales values are close to the actual sales values.

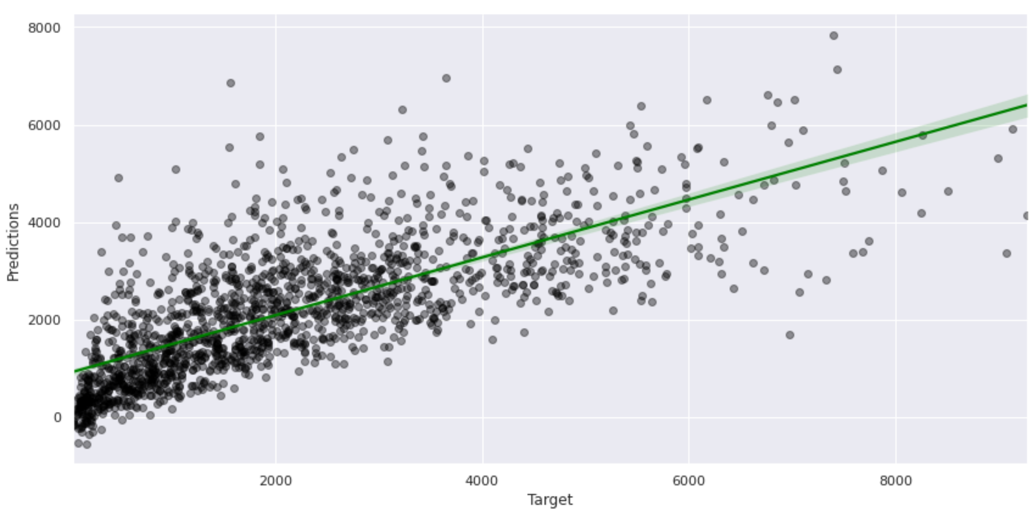

Figure 3: visualization of fitting line of our model predictions.

6. Conclusion

This paper has presented a solution to optimize SCM in the retail industry using retail analysis techniques. The proposed solution involves using XGBoost as a method for optimizing the SCM and involves data preprocessing, normalization, and statistical analysis. The case study of Big Mart Sales has shown the effectiveness of the proposed

DOI: https://doi.org/10.54216/AJBOR.020202
Received: March 20, 2021 Accepted: July 05, 2021

solution in improving the supply chain performance of a retail business. The results demonstrate the potential benefits of using retail analysis techniques for optimizing SCM in the retail industry. This paper provides a valuable contribution to the field of SCM and provides insights for retailers seeking to improve their supply chain performance.

Reference

[1]. Beamon, Benita M. "Supply chain design and analysis:: Models and methods." *International journal of production economics* 55, no. 3 (1998): 281-294.

[2]. Huan, Samuel H., Sunil K. Sheoran, and Ge Wang. "A review and analysis of supply chain operations reference (SCOR) model." *Supply chain management: An international Journal* 9, no. 1 (2004): 23-29.

[3]. Wang, Gang, Angappa Gunasekaran, Eric WT Ngai, and Thanos Papadopoulos. "Big data analytics in logistics and supply chain management: Certain investigations for research and applications." *International journal of production economics* 176 (2016): 98-110.

[4]. Golroudbary, Saeed Rahimpour, and Seyed Mojib Zahraee. "System dynamics model for optimizing the recycling and collection of waste material in a closed-loop supply chain." *Simulation modelling practice and theory* 53 (2015): 88-102.

[5]. Tiwari, Sunil, Hui-Ming Wee, and Yosef Daryanto. "Big data analytics in supply chain management between 2010 and 2016: Insights to industries." *Computers & Industrial Engineering* 115 (2018): 319-330.

[6]. Anand, Neeraj, and Neha Grover. "Measuring retail supply chain performance: Theoretical model using key performance indicators (KPIs)." *Benchmarking: An international journal* (2015).

[7]. Eskandarpour, Majid, Pierre Dejax, Joe Miemczyk, and Olivier Péton. "Sustainable supply chain network design: An optimization-oriented review." *Omega* 54 (2015): 11-32.

[8]. Schniederjans, Dara G., Carla Curado, and Mehrnaz Khalajhedayati. "Supply chain digitisation trends: An integration of knowledge management." *International Journal of Production Economics* 220 (2020): 107439.

[9]. De Meyer, Annelies, Dirk Cattrysse, Jussi Rasinmäki, and Jos Van Orshoven. "Methods to optimise the design and management of biomass-for-bioenergy supply chains: A review." *Renewable and sustainable energy reviews* 31 (2014): 657-670.

[10]. Jayaram, Athul. "Lean six sigma approach for global supply chain management using industry 4.0 and IIoT." In *2016 2nd international conference on contemporary computing and informatics (IC3I)*, pp. 89-94. IEEE, 2016.

[11]. Ansari, Zulfiquar N., and Ravi Kant. "A state-of-art literature review reflecting 15 years of focus on sustainable supply chain management." *Journal of cleaner production* 142 (2017): 2524-2543.

[12]. Zhong, Ray Y., Stephen T. Newman, George Q. Huang, and Shulin Lan. "Big Data for supply chain management in the service and manufacturing sectors: Challenges, opportunities, and future perspectives." *Computers & Industrial Engineering* 101 (2016): 572-591.

[13]. Zhao, Rui, Yiyun Liu, Ning Zhang, and Tao Huang. "An optimization model for green supply chain management by using a big data analytic approach." *Journal of Cleaner Production* 142 (2017): 1085-1097.

[14]. Farahani, Reza Zanjirani, Shabnam Rezapour, Tammy Drezner, and Samira Fallah. "Competitive supply chain network design: An overview of classifications, models, solution techniques and applications." *Omega* 45 (2014): 92-118.

[15]. Li, Lei, Hervé Manier, and Marie-Ange Manier. "Hydrogen supply chain network design: An optimization-oriented review." *Renewable and Sustainable Energy Reviews* 103 (2019): 342-360.

[16]. Ba, Birome Holo, Christian Prins, and Caroline Prodhon. "Models for optimization and performance evaluation of biomass supply chains: An Operations Research perspective." *Renewable Energy* 87 (2016): 977-989.

[17]. Nguyen, T., Li, Z.H.O.U., Spiegler, V., Ieromonachou, P. and Lin, Y., 2018. Big data analytics in supply chain management: A state-of-the-art literature review. *Computers & Operations Research*, *98*, pp.254-264.

DOI: https://doi.org/10.54216/AJBOR.020202
Received: March 20, 2021 Accepted: July 05, 2021