



HAL
open science

Goods and Activities Tracking Through Supply Chain Network Using Machine Learning Models

Lahcen Tamym, Ahmed Nait Sidi Moh, Lyes Benyoucef, Moulay Driss El Ouadghiri

► **To cite this version:**

Lahcen Tamym, Ahmed Nait Sidi Moh, Lyes Benyoucef, Moulay Driss El Ouadghiri. Goods and Activities Tracking Through Supply Chain Network Using Machine Learning Models. IFIP International Conference on Advances in Production Management Systems, Sep 2021, Nantes, France. pp.3-12, 10.1007/978-3-030-85874-2_1 . hal-03526656

HAL Id: hal-03526656

<https://hal.science/hal-03526656v1>

Submitted on 15 Feb 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



This document is the original author manuscript of a paper submitted to an IFIP conference proceedings or other IFIP publication by Springer Nature. As such, there may be some differences in the official published version of the paper. Such differences, if any, are usually due to reformatting during preparation for publication or minor corrections made by the author(s) during final proofreading of the publication manuscript.

Goods and Activities Tracking Through Supply Chain Network Using Machine Learning Models

Lahcen Tamym¹, Ahmed Nait Sidi Moh², Lyes Benyoucef³, and Moulay Driss El Ouadghiri¹

¹ Moulay Ismail University, IA Laboratory, Meknes, Maroc
l.tamym@edu.umi.ac.ma, d.elouadghiri@umi.ac.ma

² Picardie Jules Verne University, IT Laboratory (LTI), Saint Quentin, France
ahmed.nait@u-picardie.fr

³ Aix-Marseille University, University of Toulon, CNRS, LIS, Marseille, France
lyes.benyoucef@lis-lab.fr

Abstract. End-consumers satisfaction with the higher efficiency and reliability of the products and services provided by the enterprises is a highly important factor in their competitiveness. However, providing efficient tracking and tracing of shipped products enhance customer loyalty and the enterprise image. Satisfied customers are one of the enterprise's greatest assets. In doing so, we are mainly interested in detection of fraudulent transactions and late delivery of orders, as well as tracking commodities and related supply chain costs over different countries. Two datasets are used for model training and validation: DataCo Supply Chain Dataset and SCMS Delivery History Dataset. A case study is worked out, and the finding results are compared to some related works in the literature. The obtained results show the added value of our proposed models.

Keywords: Artificial Intelligence · Goods Tracking · Machine Learning · Supply Chain Network.

1 Context and Motivations

Managing and predicting the future performances of supply chain network (SCN) operations are the most challenging tasks facing many enterprises in today's business. It is the primary reason to leverage more relevant modeling approaches. Artificial intelligence (AI) and machine learning (ML) are the two tools that we will use in this study. Moreover, tracking every SCN activity from procurement of raw materials to final product delivery to end-customers remains an essential process for networked enterprises. Hence, predictive analytics enable real-time monitoring of shipment and predicting delays and risks related to the goods flow on roads, plants, warehouses, distribution centers, etc. Furthermore, this leads researchers and professionals to ask this question: What are more relevant tools for SCN management?. Accordingly, enterprises start to pay more attention to machine learning models and what they can do to improve their SCNs. Also, one

asks how this powerful technology can address the operation problems linked to their network's activities. In this research work, we are interested in studying ML algorithms for tracking and monitoring goods flow through SCNs. The reason behind this interest is that end-consumer satisfaction with a higher quality of products and services are the main decisive elements within enterprises [1]. Our objective is to develop flexible, scalable, robust, and data-driven models for goods tracking within large-scale and dynamic SCNs. Often, these networks are exposed to various levels of uncertainty [2] which complicate their management and future predictions. The developed models are based on two datasets: DataCo Supply Chain Dataset and SCMS Delivery History Dataset. A case study is worked out, and developed models are analyzed and discussed. Findings results are compared to some developed research works in the literature and show that the proposed models in this work are promising.

The remainder of the paper is organized as follows: Section 2 reviews the importance of ML in SCN Tracking. The adopted research methodology is addressed in Section 3. Section 4 illustrates two case studies and discusses finding results. Finally, Section 5 concludes the paper and gives some challenges and future work outlines.

2 Literature Review

Modern and international (global) networked enterprises generate vast amounts of complex data. ML can analyze and process these data and use the extracted information to enhance network management. Many applications based on ML to support this kind of network have been developed in the literature. Learning modelling to generate goods tracking solutions for networked enterprises is a critical process. Besides, ML improves the robustness of solutions.

2.1 SCN Management

SCN management involves integrating networked enterprises' activities such as suppliers, manufacturers, distributors, third-party service providers, customers, etc. This integration needs coordination, collaboration, and enduring alliances between all entities of the network. It leads to the enhancement of the visibility of management processes involved in SCN activities such as: sourcing, procurement, demand forecasting, materials requisition, order processing and fulfillment, transportation services, invoicing and payment processing, and goods flow tracking through the network [3]. Finding the right strategy to track and manage each stage of the network is challenging for these linked enterprises. The management of SCN provides, with more efficiency, the opportunity for the distribution of goods across the network [4]. Hence, this helps to achieve a high degree of satisfaction with requirements at a minimum cost [5]. Moreover, goods and activities tracking improves customer service level, optimizes the production cycle, reduces warehouse inventory, improves enterprise productivity and profitability, controls the production process, etc. [6].

2.2 Goods Tracking Within a SCN

Quickly developing information technologies, optimization algorithms, internet of things (IoT), AI methods optimize logistic processes across networked enterprises. Relevant solutions offered by these technologies combined with Big-Data (BD) analytics (e.g., GPS vehicle tracking data, information of customers and their locations, changing the information on traffic volume on route sections, organization/enterprise specific historical data, etc.) are required [7]. Furthermore, the efficiency of goods distribution and services through SCNs is improved by machine learning algorithms based on the flow of input data [8]. Using GPS technology coupled with radio frequency identification (RFID) traces offers a complementary source of data used to identify logistics facilities in urban areas using density-based clustering [9]. Also, this will enable efficient tracking and tracing of trucks on the roads or locomotives on rails or ships in oceans. Using these location technologies, tracing delivery status for customers often involves many parties to check the conditions and status of the cargo during shipments [3]. In addition, it helps to eliminate customers’ doubts by answering their frequent queries, such as, "Has my cargo arrived?", "When is my cargo arrived?", and "How is the current state of my cargo?". Therefore, enterprises are investing considerable amounts of money in providing tracking services to their customers. Table 1 summarises some goods and activities tracking applications through SCN in the literature.

Table 1: Goods and activities tracking applications

Tracking applications	Models, methods, and platforms	References
Determining the position of the trains and their expected time of arrival at destination. As well as, recording the evolution of critical parameters inside the cars during transportation.	GRailChem: Identification of Software Specifications through Quality Function Deployment.	[10]
Collaborative platform based on advanced technologies for positioning, identification, communication, tracking and data sharing about logistic flows within supply chains.	IoT, Cloud computing, GPS/GPRS and RFID	[11]
Decision-making agents in a blood supply chain network via Machine Learning in order to solve large-scale optimization problems.	ANN, CART, RF, k-NN	[12]
Predicting fraudulent transactions, and mitigating future dangers within smart supply chains.	Random Forest, Rpart	[13]

3 Research Methodology

In this study, we aim to use machine learning models to track goods and activities through a SCN. Figure 1 shows the global framework of ML applications in SCNs tracking. This framework is composed of five major components: 1) SCNs data sources in which each single SC provides its data in different formats. 2) Data preprocessing in which the data are being processed and prepared as input of ML models. 3) Use of ML models. 4) Predictive analytics. 5) The outputs of predictive analytics which are goods and activities tracking along with the network.

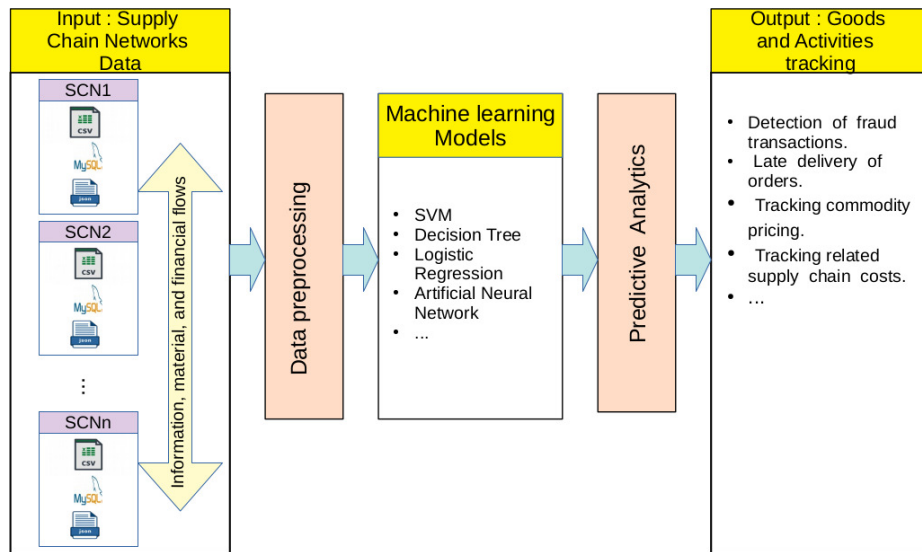


Fig. 1: Global framework of ML applications in SCN tracking

We adopt the case study methodology to validate our approach. In doing so, we conduct our research on two real datasets described below in section 4.1. The best model based on their performance in the simulation study is selected to be used in practice. In the following, we briefly review the background of the candidate ML models before their implementation for the case studies.

3.1 Artificial Neural Network (ANN)

ANN is a general function approximator, which learns the relationship between independent and dependent variables after performing a learning process from training data [12]. In addition, due to its nonlinear characteristics, ANN performs well when modeling complex data patterns with nonlinear functions. ANNs are composed of processing units called neurons. According to their architecture,

and how neurons interact in the network and their structure, they can generally be divided into three classifications: multilayer feedforward network, recurrent network and mesh network [14].

3.2 Support Vector Machine (SVM)

SVM is a set of related supervised learning methods for classification and regression and belongs to the family of generalized linear classifiers. In other words, SVM is a classification and regression prediction tool that uses machine learning theory to maximize prediction accuracy while automatically avoiding overfitting the data. In addition, SVMs are systems that use a hypothesis space of linear functions in a high-dimensional feature space and are trained using a learning algorithm from optimization theory, which implements learning bias from statistical learning theory.

3.3 Decision Tree (DT)

DT is a supervised learning technique that can be used for both classification and regression problems. It is a tree-structured classifier, where internal nodes represent the features of a dataset, branches represent the decision rules, and each leaf node represents the outcome. It is a graphical representation for getting all the possible solutions to a problem/decision based on given conditions. It is called a decision tree because similar to a tree, it starts with the root node, which expands on different branches and constructs a tree-like structure. In order to build a tree, we use the CART algorithm, which stands for Classification and Regression Tree algorithm.

3.4 Logistic Regression (LR)

Logistic regression or logit model is a binomial regression model. As for all binomial regression models, it aims to model a simple mathematical model and possible with numerous real observations. In other words, to associate to a vector of random variables (x_1, \dots, x_K) a binomial random variable generically noted y . Logistic regression is a special case of the generalized linear model.

4 Goods Tracking: Case Studies

In order to show how ML technology offers many solutions to any SC's enterprise, we conducted our research work on two datasets, "SCMS Delivery History" and "DataCo Supply Chain" by leveraging different ML models and evaluating the performance of each model by the accuracy of prediction.

4.1 Used Datasets

SCMS Delivery History Dataset [15]. This dataset provides transportation modes and pricing data of health commodities in the supply chain. It identified antiretroviral (ARV) and HIV laboratories shipped to supporting countries. Also, it provides the commodity pricing and related supply chain costs necessary to move the commodity to the countries for use. After analyzing these data, we came out by the flowing statistics; the top country for pack price is Nigeria - 25,620.72, top Shipping Mode: Air, the Max Air Shipment Mode is: 1000, top Manufacturing Site: Aurobindo Unit III, India - 3172, top Air Manufacturing Site: Aurobindo Unit III, India - 1694. Further analysis of this dataset is given in Fig.2

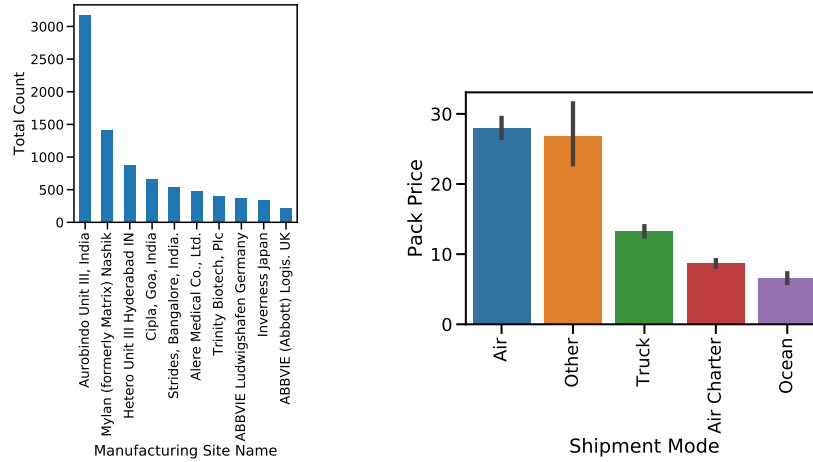


Fig. 2: (1) Top 10 manufacturing sites (left). (2) Shipment mode per pack price (right)

DataCo Supply Chain Dataset[16] which is used by the enterprise DataCo Global for the analysis, and consists of roughly 180k transactions from supply chains for 3 years. It's a dataset of Supply Chain, which allows the use of Machine Learning Algorithms. These data offer areas of important registered activities such as, provisioning, production, sales, commercial distribution. In this dataset, we are interested in detection of fraud transactions and late delivery of orders.

On the one hand, finding which payment method is used to conduct frauds can be useful to prevent fraud from happening in the future. Figure 3 shows which payment method is preferred the most by people in different regions. While, Figure 4 indicates which region and what product is being suspected to the fraud the most.

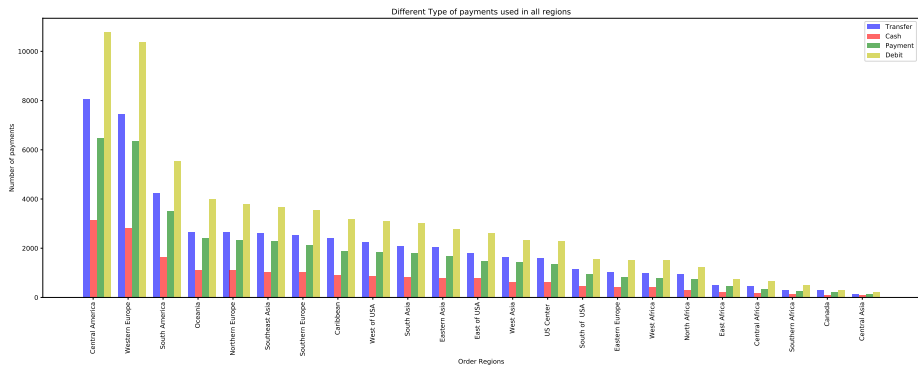


Fig. 3: Different preferred payment methods in different regions

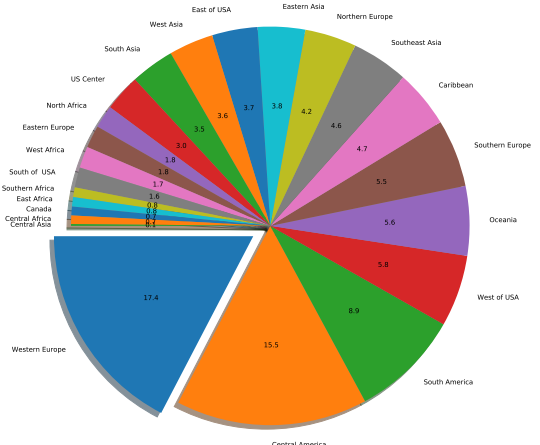


Fig. 4: Different regions that are suspected to the fraud the most

On the other hand, when products are not delivered on time to customers, this will be a source of doubts and decreases in satisfaction. Consequently, this leads to trust loss. Thus, solving the late delivery problem is vital to any supply chain enterprise within the SCN. From the performed analysis, Figure 5 shows what category of products are being delivered late the most.

4.2 Results

This section details the experimental results obtained from comparing popular ML classifiers and measuring their performances in terms of accuracy. The goal is to find out which ML model performs better on both datasets described above. For SCMS Delivery History Dataset, we are interested in transportation mode

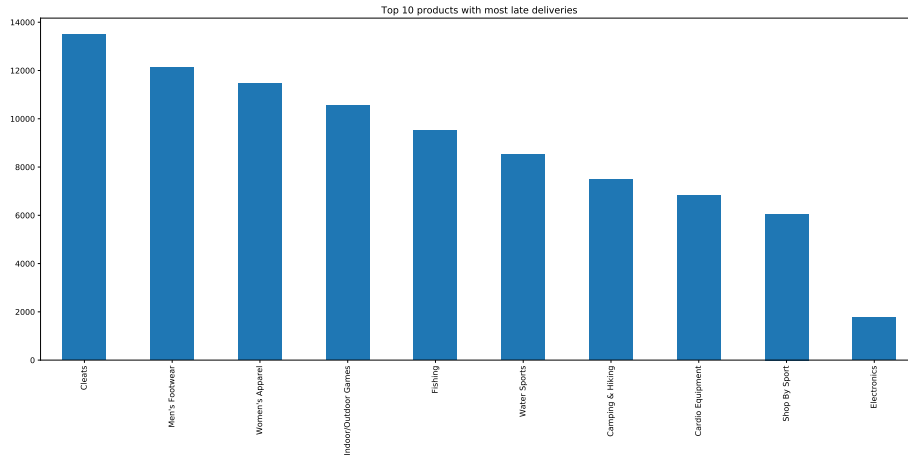


Fig. 5: Category of products that are being delivered late the most

prediction based on the scheduled delivery date, delivered to customer date, cargo weight, and freight cost, etc. In doing so, we studied and examined different ML models such as ANN, SVM, DT and LR. Hence, the obtained results exposed in Table 2 show that we were able to reach a significant accuracy of prediction. For DataCo Supply Chain Dataset, we are interested in two classi-

Table 2: Transportation mode prediction.

Dataset	Classification problem	Classes	Samples	Accuracy			
				ANN	SVM	Decision Tree	Logistic Regression
SCMS Delivery History	Transportation mode prediction	Air, Truck, Air Charter, Ocean, Other	10324	0.89	0.833	0.80	-

fication problems: late delivery of orders and detecting fraudulent transactions. All the orders that are delivered late every time are creating risks to the SC enterprises. In this dataset, most of the orders with Cleats, Men's Footwear, and Women's Apparel category products are causing late delivery, as well as these products, are suspected to fraud the most. Indeed, for predicting late delivery risk and detecting fraud, we applied the same ML models for the first dataset. The accuracy of each model are mentioned in Table 3. The finding results are compared to the obtained results in [13], Table 4. Indeed, we can see that the proposed models for fraud detection are outperformed the results found in [13].

Table 3: Late delivery of orders and detecting fraudulent transactions.

Dataset	Classification problem	Classes	Samples	Accuracy			
				ANN	SVM	Decision Tree	logistic regression
DataCo Supply Chain	Late delivery risk	0, 1	180519	0.9778	0.988	0.993	0.9884
	Frauds detection	0, 1	180519	0.98	0.9775	0.9907	0.978

Table 4: Detecting fraudulent transactions [13].

DataCo	Classification problem	Random Forest	Rpart
Supply Chain	Frauds detection	0.8155	0.761

After leveraging different ML models on both datasets, we have obtained the best results in prediction. These will allow networked enterprises to be able to track different activities and goods through the network, as well as, solving many problems that will occur in the future. It enables us to build a robust network and remain competitive in the global market.

5 Conclusion

In this research work, we showed interest in applying machine learning models in supply chain networks. Tracking every SC activity and goods through the network, from procurement of raw materials to final product delivery to end customers, remains an essential process for networked enterprises. Hence, machine learning predictive analytics enable real-time monitoring of shipment and predicting delays and risks related to the goods flows and related activities. We selected the four most well-known machine learning models to examine our research work. These models can be trained to be used in day-to-day operational decision-making through SCNs.

Consequently, the use of machine learning for this kind of tracking will arise many challenges such as data privacy, understanding of data related to SCN, which is a complex part of this application due to missing data and the scarcity of data. In addition, in the case of a real-time application, it required performing tools and sophisticated techniques.

Finally, as future work, we expect to apply deep learning models for enhancing the solutions offered by tracking in SCN, especially for real-time tracking. Moreover, studying a real-life scenario of SCN will be one of our primary concerns in the near future.

References

1. America, S.: Track/trace solutions for the logistics supply chain. <https://www.supplychainmarket.com/doc/tracktrace-solutions-for-the-logistics-0001>. Accessed 07 June 2021.
2. Biller, B., Yi, J., Inc., S.I., NC, C.: Optimizing supply chain robustness through simulation and machine learning. <https://www.sas.com/content/dam/SAS/support/en/sas-global-forum-proceedings/2020/4535-2020.pdf>. Accessed 07 June 2021.
3. He, W., Tan, E., Lee, E., Li, T.: A solution for integrated track and trace in supply chain based on rfid and gps. pp. 1–6 (2009). <https://doi.org/10.1109/ETFA.2009.5347146>

4. Kadadevaramath, R.S., Chen, J.C., Latha Shankar, B., Rameshkumar, K.: Application of particle swarm intelligence algorithms in supply chain network architecture optimization. *Expert Systems with Applications* **39**(11), 10,160–10,176 (2012). <https://doi.org/10.1016/j.eswa.2012.02.116>
5. Boiko, A., Shendryk, V., Boiko, O.: Information systems for supply chain management: uncertainties, risks and cyber security. *Procedia Computer Science* **149**, 65–70 (2019). <https://doi.org/10.1016/j.procs.2019.01.108>
6. McKinsey, C.: Next generation supply chain : Supply chain 2020 (2020). Accessed 10 June 2020.
7. Tamym, L., Benyoucef, L., Nait Sidi Moh, A., El Ouadghiri, M.D.: A big data based architecture for collaborative networks: Supply chains mixed-network. *Computer Communications* **175**, 102–111 (2021). <https://doi.org/https://doi.org/10.1016/j.comcom.2021.05.008>
8. Tarapata, Z., Nowicki, T., Antkiewicz, R., Dudzinski, J., Janik, K.: Data-driven machine learning system for optimization of processes supporting the distribution of goods and services a case study. *Procedia Manufacturing* **44**, 60 – 67 (2020). <https://doi.org/10.1016/j.promfg.2020.02.205>
9. Trent, N.M., Joubert, J.W., Gidofalvi, G., Kordnejad, B.: A matching algorithm to study the evolution of logistics facilities extracted from gps traces. *Transportation Research Procedia* **46**, 237–244 (2020). <https://doi.org/10.1016/j.trpro.2020.03.186>
10. Vlad, R.C., Benyoucef, L., Vlad, S.: Identification of software specifications through quality function deployment. In: 2006 IEEE International Conference on Automation, Quality and Testing, Robotics, vol. 2, pp. 74–79 (2006). <https://doi.org/10.1109/AQTR.2006.254603>
11. Gnimpieba, Z.D.R., Nait-Sidi-Moh, A., Durand, D., Fortin, J.: Using internet of things technologies for a collaborative supply chain: Application to tracking of pallets and containers. *Procedia Computer Science* **56**, 550–557 (2015). <https://doi.org/10.1016/j.procs.2015.07.251>
12. Abbasi, B., Babaei, T., Hosseinifard, Z., Smith-Miles, K., Dehghani, M.: Predicting solutions of large-scale optimization problems via machine learning: A case study in blood supply chain management. *Computers and Operations Research* **119**, 104,941 (2020). <https://doi.org/10.1016/j.cor.2020.104941>
13. Constante-Nicolalde, F.V., Guerra-Terán, P., Pérez-Medina, J.L.: Fraud prediction in smart supply chains using machine learning techniques. In: *Applied Technologies*, pp. 145–159. Springer International Publishing, Cham (2020)
14. Zhang, Z.: *Multivariate Time Series Analysis in Climate and Environmental Research* (2018). <https://doi.org/10.1007/978-3-319-67340-0>
15. Ardeshana, D.: Supply chain shipment pricing data. <https://www.kaggle.com/divyeshardeshana/supply-chain-shipment-pricing-data>. Accessed 07 June 2021.
16. Constante, F., Silva, F., Pereira, A.: Dataco smart supply chain for big data analytics data. <https://data.mendeley.com/datasets/8gx2fvg2k6/5>. Accessed 07 June 2021.