applied sciences

Article

Dynamic Lead-Time Forecasting Using Machine Learning in a Make-to-Order Supply Chain

Mohammed Alnahhal [1,*], Diane Ahrens [2] and Bashir Salah [3]

[1] Mechanical and Industrial Engineering Department, American University of Ras Al Khaimah, Ras Al Khaimah P.O. Box 10021, United Arab Emirates
[2] Technology Campus Grafenau, Deggendorf Institute of Technology, 94469 Deggendorf, Germany; diane.ahrens@th-deg.de
[3] Industrial Engineering Department, College of Engineering, King Saud University, P.O. Box 800, Riyadh 11421, Saudi Arabia; bsalah@ksu.edu.sa
* Correspondence: mohammed.alnahhal@aurak.ac.ae; Tel.: +971-7-246-8748

Abstract: This paper investigates the dynamic forecasting of lead-time, which can be performed by a logistics company for optimizing temporal shipment consolidation. Shipment consolidation is usually utilized to reduce outbound shipments costs, but it can increase the lead time. Forecasting in this paper is performed in a make-to-order supply chain using real data, where the logistics company does not know the internal production data of manufacturers. Forecasting was performed in several steps using machine-learning methods such as linear regression and logistic regression. The last step checks if the order will come in the next delivery week or not. Forecasting is evaluated after each shipment delivery to check the possibility of delaying the current arriving orders for a certain customer until the next week or making the delivery to the customer immediately. The results showed reasonable accuracy expressed in different ways, and one of them depends on a type I error with an average value of 0.07. This is the first paper that performs dynamic forecasting for the purpose of shipment temporal consolidation optimization in the consolidation center.

Keywords: freight consolidation; lead-time forecasting; make-to-order; machine learning; supply chain

check for updates

Citation: Alnahhal, M.; Ahrens, D.; Salah, B. Dynamic Lead-Time Forecasting Using Machine Learning in a Make-to-Order Supply Chain. *Appl. Sci.* **2021**, *11*, 10105. https://doi.org/10.3390/app112110105

Academic Editor: Paweł Sitek

Received: 9 September 2021
Accepted: 25 October 2021
Published: 28 October 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

In temporal freight (shipment) consolidation, small orders are aggregated over time in the consolidation center to make bigger shipments to the customer (retailer) [1]. This is in addition to product-based consolidation in which products from several suppliers are aggregated to be delivered for the same customer [2]. A sustainable supply chain is about the achievement of an organization's social, environmental, and economic goals. These goals are supportive for the organization and the community. Reducing the number of shipments and using larger shipments contributes to decreasing the traffic jam as one of the social benefits. The environmental effect is attained by reducing the CO_2 emissions by reducing the number of shipments [3]. The economic goal is accomplished by reducing transportation costs. This is achieved as long as the savings are larger than the increase in inventory holding costs and as long as the lead time is not strongly affected. Moreover, consolidation reduces the number of vehicles and the size of the workforce needed. This is important in the time of the COVID-19 pandemic, where a shortage in the workforce in logistics is a big problem. Moreover, it will be useful for customers because they will have a lower number of shipments, and therefore, lower handling costs. All these three goals of sustainability and the response to COVID-19 are some of the motivations for this study. In addition, this study is the first one that considers temporal consolidation as the objective for forecasting.

However, temporal consolidation is only performed if there is a good chance that other orders for the same customer will come in the few coming days, where the waiting period is much smaller than the average lead time (time elapsed between making the order by the customer and receiving the order on the warehouse) [4]. Therefore, accurate forecasting of orders lead time is needed. Lead-time forecasting is an extremely challenging task [5]. The variability in production lead time can be due to factors such as machine breakdown and maintenance [6]. Predicting longer lead times is more difficult and tends to be less accurate because uncertainty leads to an increase in the variance of estimation [7]. Instead of making traditional static forecasting for each order, dynamic forecasting is performed in this paper, where the evaluation is recalculated after each delivery or each demand [8].

This research is inspired by a case study of a furniture design company, which makes designs and sells products. The company outsources all the manufacturing activities to external suppliers in the east of Europe, where customers of the company are mainly in the west of Europe. The company follows this strategy to utilize the low manufacturing costs in the east of Europe. There are more than 1000 customers (retailers) spread in several countries in the west of Europe. Most of the demand by each customer is usually one or a few pallets. One order by a customer might include items from several suppliers at the same time. To aggregate the small orders together, a consolidation center (large warehouse) is needed in the west of Europe to put the needed items for each group of customers in the same small zone on the same truck to reduce transportation costs. This strategy is called a milk run. The warehouse is managed by a third-party logistics provider (3PL). This case study is similar to those described in Alnahhal et al. [4,9]. However, the objectives in this paper are different.

The general aim of this paper is to make forecasting that satisfies the company's special needs of reducing the total costs by temporal consolidation. That means that forecasting is an input for the temporal consolidation plan, in which aggregating larger shipments is performed. To plan that, a special type of forecasting with binary results (come, not come) is needed. Instead of evaluating the exact lead time for each order, the manager is interested more in the probability of the arrival of the order in the next delivery week. Not all current orders will come in the next delivery week because the lead time is relatively long. If the probability that a certain order will come in the next delivery is high, then the currently available order can wait until the next delivery, and the two shipments can be aggregated together to reduce the shipment costs. Such a probability and a binary decision (come next week, not come) is usually investigated by logistic regression, which is the last and the most important type of forecasting in this study. To prepare for this stage, linear regression is needed to estimate the total number of coming orders in the next week. However, inventory costs, when temporal aggregation is used, will be higher because of the extra waiting. However, such waiting also depends on the period for which the current items available in the warehouse are waiting so far; if this period is not too long, waiting can occur. If the acceptable lead time is, for example, 4 weeks, and the available items have an order preparation time (OPT) of less than 3 weeks, then consolidation is possible. Consolidation can be for all the items from all suppliers for the same customer. However, if some of the current items have an OPT of more than three weeks, then all materials should be shipped immediately to the customer. That means that not all customer demand needs to be forecasted.

Such a type of forecasting for temporal consolidation purposes is performed for the first time in this paper. Most of the studies in the literature about forecasting focus on forecasting for other purposes such as inventory management and shop scheduling; the forecasting process is performed by the manufacturing company itself based on large datasets about production processes. However, in this study, the dynamic weekly forecasting process for each customer order is performed from the point of view of the 3PL, depending on current and historical data of demand and lead time. The forecasting process in this paper is performed for an environment that is overlooked by the previous studies, where the make-to-order supply chain contains four different parties, namely, the design company, many suppliers, the 3PL, and more than 1000 customers (retailers).

2. Literature Review

The use of machine learning in supply chains is known in the literature. The main focus of this paper is the forecasting of lead time. Customers usually require an accurate estimation of lead time to ensure their production and delivery due date [10]. Usually, companies make forecasting based on full knowledge of their production line conditions, and this type of forecasting is called manufacturing lead-time forecasting (MLTF). However, sometimes, the forecasting of supplier companies might not be available for customers or 3PLs, at least not for each order. Therefore, the customers or 3PLs might need to make their own forecasting of lead time based on historical data about their suppliers. We call this type customer lead-time forecasting (CLTF). Based on that, the definition of lead time is different in both cases. The first one is the time that elapses between the release of an order and its completion (arrival in the finished goods inventory) [11]. In the second definition, transportation time until reaching the customer or the 3PL is added. Although MLTF is well known in the literature, little was published on CLTF. In this study, we assume that the exact situation in the supplier factory is a black box. For example, the production and lead time for other customers are unknown, so historical data are only available for the customers of the furniture design company under consideration.

The problem in this study is related to other decisions. In this study, forecasting performed by the 3PL is for the purpose of shipment consolidation. One of the studies that investigated this problem is the one by Hanbazazah et al. [12], who focused on transshipping multiple products from multiple suppliers to a single end customer. The 3PL consolidates the inbound shipments so as to reduce costs. Consolidation sometimes leads to aggregating less-than-pallet-size shipments from different suppliers. Therefore, a pallet loading problem is needed. Such a problem was investigated in the study by Aylak et al. [13] using machine learning. This means that different decision problems in the supply chain are interconnected. The disadvantage of shipment consolidation is that it increases procurement interdependencies. Therefore, Kosasih and Brintrup [14] used Graph Neural Networks (GNN) to predict hidden links in the supply chain and gain visibility to prepare contingency plans. Another direction is to increase cooperation between different supply chain parties. Such cooperation can enhance innovation activities and therefore competitiveness [15].

Most of the research was performed on make-to-stock environments to select the best inventory control policy. However, there are some studies that investigated engineer-to-order or make-to-order environments. Most of the studies used machine learning and data mining with its different techniques for forecasting [16,17]. However, there are other studies, which used simulation or operations research [18–21]. Moreover, Mean Absolute Percentage Error (MAPE) is usually used to estimate the performance of forecasting methods [6]. Azadeh et al. [7] performed a comprehensive procedure for comparing several machine-learning techniques for manufacturing lead-time estimation in the case of failure of the machines. The best model is selected in a study by Asadzadeh et al. [6] based on MAPE. They based their study on weekly lead times for a large complex electric-motor assembly line, where data are taken for 70 weeks.

One of the make-to-stock studies is performed by Dosdoğru et al. [22], who used artificial intelligence (AI)-based simulation to predict the lead time of supply chain members. They used an Artificial Neural Network (ANN) together with a Genetic Algorithm (GA). The objective was to design inventory and routing in the best way to assure adequate inventory level and also provide an efficient route. This is part of using the correct inventory control policy. The prediction of lead time was necessary for planned order releases and shop scheduling in a study by Askin and Hanumantha [23]. Therefore, they investigated the prediction of lead times under dynamic conditions. They assumed variable product mixes and demands with equipment or staffing changes. On the other hand, Indrianti et al. [20] determined the delivery time of a set of customer orders in a repetitive make-to-order manufacturing company that utilizes recycled waste products as raw materials. A simulation approach was applied to find the minimum total flow time or makespan to determine the delivery time of each order. Moreover, Mourtzis et al. [24] proposed a methodology, which has been developed into a software tool, for complex engineered-to-order products. This is achieved through the examination of the characteristics of the product. Öztürk et al. [16] used regression trees for lead-time forecasting in make-to-order manufacturing. Furthermore, Burggräf et al. [18] reviewed studies for the prediction of lead times in engineer-to-order environments with a typically large number of individual parts and complex production processes. In such studies, operations research and machine learning are usually used. In a study by Li et al. [19] in a produce-to-order environment, researchers quote a reliable lead time for a new job (or order) upon its arrival to a manufacturing system.

Usually, MLTF forecasting is performed by the company itself or at least after knowing enough details of the company production processes. For example, Gyulai et al. [25] investigated manufacturing lead time in the optics industry using analytical and machine learning techniques for customized products. Lingitz et al. [26] investigated manufacturing lead time using supervised machine learning approaches depending on historical production data from the manufacturing execution system (MES). Moreover, lead-time forecasting was performed based on ANN, fuzzy regression (FR), and conventional regression (CR) in a study by Asadzadeh et al. [6]. Their study estimates the weekly lead times of an actual assembly shop. Another study that depends on MES data is the one by Pfeiffer et al. [27], which combines simulation and statistical learning methods.

On the other hand, few studies focused on CLTF, which is the focus of this paper. Forecasting of purchasing lead time was investigated in a study by de Oliveira et al. [5] in the context of the pharmaceutical supply chain by machine learning regression algorithms. The support vector machines approach was found to have the best performance. Their work was motivated by the activity of a logistic company that receives the pharmaceutical products from the suppliers and then organizes the shipping to healthcare facilities. Moreover, Kar and Jha [28] investigated material management and the effect of lead-time estimation based on the identification of the factors that influence it, such as the price of materials. They collected a large sample of procurement data from 16 building construction projects, and cluster analysis and regression analysis were used. Furthermore, Yamini and Marathe [29] considered determining a realistic delivery time in the supply chain by reducing bias.

Lead-time estimation was expressed in the literature in different ways such as the study by Gacek [30], who used ANN to estimate due date assignment in a small batch and multi-assortment make-to-order production company. The study made a comparison between different due date assignment methods. Moreover, Schuh et al. [31] investigated the lead time prediction in a customer-individualized products' environment, where planning complexity grew significantly in such a case. They investigated the prediction of transition times (TT). As up to 99% of lead time consists of TT, which is the time between two processing steps, they used a data mining methodology for order-specific TTs. Some studies concentrated on special conditions such as Zhong et al. [10], who investigated an RFID-enabled real-time manufacturing shopfloor environment, where a large number of factors may greatly affect its precision. Actual processing time is hard to estimate due to the dynamic manufacturing environment and uncertain disturbances. Their paper proposed a data mining approach. Moreover, Berlec et al. [32] investigated the lead-time forecasting of production orders in SMEs. Their work is based on the actual lead times of operational and assembly orders processed in the company's workplaces in the past.

In many cases, the forecasting lead time process was static, however, Ioannou and Dimitriou [8] investigated the problem of dynamically updating the manufacturing lead times estimates used in MRP systems. The estimate is performed when an order enters the make-to-order manufacturing system. Moreover, Schneckenreither et al. [11] presented a flow time estimation procedure to set lead times dynamically using an ANN in a make-to-order flow-shop environment. The timing of the order release decision (start of production) is based on a planned lead time. To the best of the authors' knowledge, this is the first paper that investigates dynamic forecasting to manage shipment consolidation in a make-to-order supply chain.

3. Materials and Methods

The main aim of the paper is to predict using machine learning which items already ordered will come in the next week. This forecasting is performed just after receiving the items in the consolidation center in the current week. This is for the purpose of temporal consolidation. Therefore, when there are no previous items already waiting in the warehouse for a certain customer, there is no need to make forecasting for the items of that customer. Usually, the supplier makes delivery every week, most probably on Wednesday. However, sometimes, the shipment comes to the warehouse one day later or earlier. In some special cases, the delay is one week, and that means that the delivery comes in two weeks. The warehouse has a notice from the supplier for that delay. The decision makers in the warehouse do not know which item will come in the next delivery. This is because the lead time is usually very long (some weeks), and any open order might not come in the next delivery. For example, only 30% of the orders might come in the next week. What makes it more difficult to forecast is that many orders do not come based on the first come first served (FCFS) principle, because the supplier has limited capacity for different production lines, and demand from other customers is not known.

There are several reasons why the lead time is longer than usual, which are as follows:

- The nature of the supply chain, which is make-to-order. In many cases, the design of furniture is unique. Therefore, it needs more time.
- Customer demand can occur every day, but the supplier delivery is performed every week and sometimes every two weeks. The supplier does this because of the shipment consolidation inside the supplier factory to reduce transportation costs [4].
- Some of the products do not come based on FCFS, as mentioned before.
- Other customers' orders compete for the same production lines of the suppliers.

In this study, the main affecting factors on forecasting are the size of demand in the last few weeks, the number of orders not yet satisfied, and the period since the order is still opened. Other factors such as customer, product type, and model types were investigated in a primary stage, and it was found that their effects are negligible. Therefore, they were neglected in the analysis in this paper.

In forecasting, it is important to distinguish two types of errors:

- Type I error, in which the forecasting result indicates delivery in the next week, but delivery does not happen.
- Type II error, in which the forecasting results indicate no delivery in the next week, but delivery is actually completed.

In this study, type I error is the most important one because it is not good for the customer order in the consolidation center to wait the whole week, and then the next order does not come, and therefore, larger inventory holding costs occur. Therefore, the manager might set a value of 0.05 as an acceptable level for such an error. Type I error has a close relationship to the objective of the study. The manager wants to let some shipments wait, only if there is a high chance that another shipment is coming in the next week. The probability of expecting that a shipment will not come in the next delivery, while it will not come, does not have the same importance as type I error. This differentiation of types of errors cannot be measured using common indices such as RMSA and MSE.

The available data are for the period from December 2014 until August 2016. In this study, one of the suppliers will be investigated. The approach of this study can be applied to any supplier. Forecasting is repeated after each delivery, which occurs almost every week. The dynamic forecasting in the last step will use logistic regression since it is appropriate for the purpose of the study. The results will be 1 or 0, where 1 means delivery of a certain order is expected to occur, and 0 means delivery does not occur. R software was used to make the calculations. Generally, the popular algorithms that can be used for binary classification include: logistic regression, k-nearest neighbors (KNN), decision trees, support vector machine (SVM), and naïve Bayes. Logistic regression is easier to implement, interpret, and very efficient to train. Moreover, it can interpret model coefficients as indicators of feature importance. KNN is a non-parametric model, where logistic regression is a parametric model. Moreover, KNN is comparatively slower than logistic regression. In decision trees, you can have two or more decisions (for example, option 1, option 2, and option 3). However, in logistic regression, there are only two options (for example, come and not come). SVM works well with unstructured and semi-structured data such as text and images, while logistic regression works with already identified independent variables. However, in many cases, both of them can be used. Naïve Bayes has a naive assumption that the algorithm expects the features to be independent which not always is the case.

In the last step of forecasting (logistic regression), the *glm* function is used. The continuous results from zero to one are first obtained. Then, they are rounded to 0 or 1. For the optimal level of overall accuracy, normal rounding is used, where any value from 0.5 until 1 is 1, and the rest of the values are 0. However, to obtain results with a type I error that is less, larger thresholds can be used. For example, 0.55 can be used. The study is performed using the following steps:

1. The demand during the last three months. This step shows the pattern of demand over the study period. Most of the demand is satisfied during six weeks. However, sometimes, some orders need more than two months to be satisfied. This is why it might be better to investigate the demand during the last three months.
2. The percentage of satisfied demand occurred in the last three months (PSDTM). The moving average can be used with $n = 3$.
3. The demand for the next week (DNW). It is given, and it is simply the waiting demand (the demand not satisfied so far). Such demand does not include the orders coming after the current delivery date. Any order that comes after the current delivery date will not be satisfied in the next delivery time, because of the long lead time. Therefore, DNW will be used to forecast if a certain order will come in the next week or not, in the final step.
4. The total number of orders delivered in the next delivery (expected satisfied demand in the next week) (SDNW). Moving average can represent a fast and practical solution since the changes in demand are relatively slow over time. However, simple linear regression can also be used. The independent variable is the DNW. It is multiplied by the estimated percentage of satisfied demand next delivery week (PSD), and this percentage is found using moving average or using the linear regression. This percentage is different from the one in step 2 for the last three months. Forecasting PSD, and then SDNW from it, was found to have better accuracy than forecasting SDNW directly.
5. The probability that a certain order will come in the next delivery week (PONW). This is performed using logistic regression, and it is the most important step in this research.

Figure 1 shows the methodology of the study. The results of forecasting are that there are orders that are expected to come in the next week and others that are not coming in the next week. For those coming in the next week, the question is if there are orders already waiting in the warehouse for the same customer (come in the current week). If yes, then the question is if waiting for one more week is possible without exceeding the allowable lead time. If yes, then temporal aggregation is possible. It is clear that moving average, linear regression, and logistic regression are used in the previous steps. A moving average captures the average change in a data series over time. Linear regression is used to predict the continuous dependent variable using a given set of independent variables.

Logistic regression tries to find the optimal decision boundary that best separates the classes. Logistic regression is not good when the training data size is small relative to the number of features. Logistic regression is able to handle categorical and continuous input variables. A key difference from linear regression is that the output value is a binary value (0 or 1) rather than a numeric value. The reader might refer to the study by Walker and Duncan [33] for more information about logistics regression. The forecasting process in this paper is dynamic. This means that the forecasting process is completed every week. The accuracy of the forecasted results is measured using the real available data in the next week. The accuracy from one week to another can be different. Logistic regression is a probabilistic classifier that makes use of supervised machine learning, so it trains the model by providing it with pairs of input–output examples from which it can learn. Training a logistic regression model means learning a mapping between the input variables and the expected output. Training the logistic regression model and using it for predictions is very simple, fast, and needs low computational requirements.

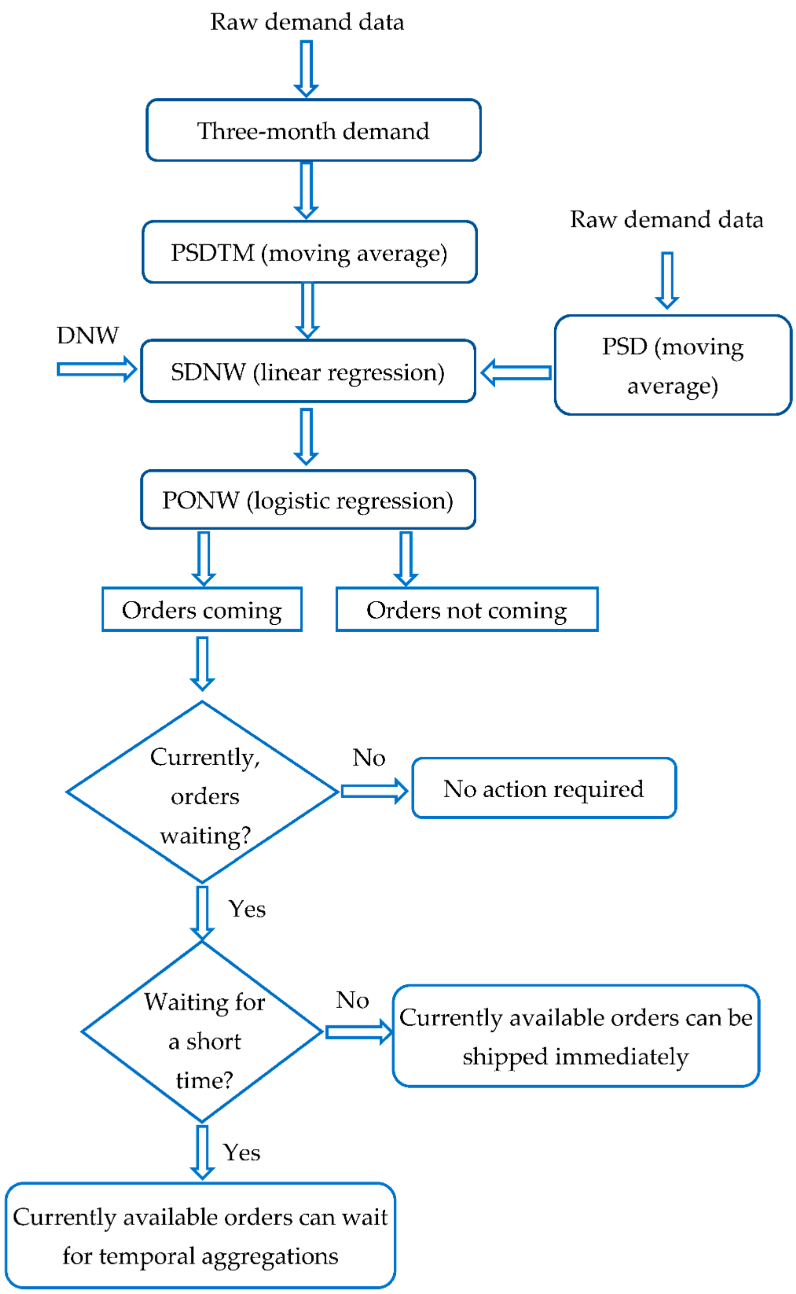

Figure 1. Methodology of the study.

A certain order achieves the satisfied order value (SO_i) of 1 if it is satisfied in the next week and 0 otherwise:

$$SO_i = \begin{cases} 1, & if\ the\ order\ i\ is\ satisfied\ in\ the\ next\ week \\ 0, & otherwise \end{cases}$$

Percent of satisfied demand in the delivery week d (PSD_d) can be found using the following equation:

$$PSD_d = \frac{\sum_{i=1}^{\mathrm{DNW}_d} SO_i}{DNW_d} \times 100\% \tag{1}$$

The next equation shows the moving average method to find $SDNW_d$ in step 4 when $n = 3$:

$$SDNW_d = PSD_d \times DNW_d = \frac{(PSD_{d-1} + PSD_{d-2} + PSD_{d-3})}{3} \times DNW_d \tag{2}$$

where d is the index for delivery week. The last type of forecasting depends mainly on two variables:

1. Index for locating the order among the other orders (ILO). It is an indication of the position of a certain order compared to other orders and the expected number of satisfied orders in the next week. This index is found from the following equation:

$$ILO_i = \frac{(SDNW_d - i)}{SDNW_d} \tag{3}$$

   where i is the order position, and it means the row number when data are arranged from the oldest to the newest orders. ILO_i starts with a value that is very close to 1 for the oldest not satisfied order and then starts to decrease until it is 0 for the order, which has a position equal to $SDNW_d$, and after that, recent orders have negative ILO_i values.
2. The order preparation time so far for order i ($OPTSF_i$), which is the difference between the current delivery date (CDD_i) and the order date (OD_i). So, the equation will be as follows:

$$OPTSF_i = CDD_i - OD_i \tag{4}$$

The previous two measures are updated every delivery, where the new available data is added to the original data. Forecasting will depend mainly on these two values plus the $SDNW_d$ value found in Equation (2). The expected satisfied order (ESO_i) is written in the same way as SO_i, and it is found using the logistic regression model. This can be written as follows:

$$ESO_i = \begin{cases} 1, & if\ the\ order\ is\ expected\ to\ be\ satisfied\ in\ the\ next\ week \\ 0, & otherwise \end{cases}$$

The variable A_i is used to check if the expected and real values are the same:

$$A_i = \begin{cases} 1, & if\ SO_i = ESO_i \\ 0, & otherwise \end{cases}$$

Forecasting accuracy for the delivery week d (FA_d) can be found using the following equation:

$$FA_d = \frac{\sum_{i=1}^{\mathrm{DNW}_d} A_i}{DNW_d} \times 100\% \tag{5}$$

Total satisfied demand next week (TSD_d) can be found as follows:

$$TSD_d = \sum_{i=1}^{DNW_d} SO_i \tag{6}$$

A_i is order expectation accuracy, which means that expectation about the order i is correct if A_i = 1. FA is found by finding the percentage of orders that meet the expectations. Adding the new data to the old data means that the same order might be in the old data with a value of SO = 0, and in the new data, SO = 1. Therefore, the same order might be repeated several times. All of them are with SO = 0, except the last one, with SO = 1. The function *glm* in R software was used. The initial results for ESO are numbers from 0 to 1. Usually, numbers greater than or equal to 0.5 are set to be 1. The other numbers less than 0.5 are set to be 0s. Therefore, the threshold value is usually 0.5 to optimize the accuracy value.

In addition to the accuracy measure, which is FA_d, the two types of errors (ε_{1d} and ε_{2d}) can also be estimated according to the following equation:

$$\varepsilon_{1d} + \varepsilon_{2d} = 1 - \frac{FA_d}{100} \tag{7}$$

where,

$$\varepsilon_{1d} = \frac{\sum_{i=1}^{\mathrm{DNW}_d} E1_i}{DNW_d} \tag{8}$$

$$\varepsilon_{2d} = \frac{\sum_{i=1}^{\mathrm{DNW}_d} E2_i}{DNW_d} \tag{9}$$

where,

$$E1_i = \begin{cases} 1 & if\ SO_i < ESO_i \\ 0 & otherwise \end{cases}$$

$$E2_i = \begin{cases} 1 & if\ SO_i > ESO_i \\ 0 & otherwise \end{cases}$$

When $SO_i < ESO_i$, SO_i = 0 and ESO_i = 1. In addition, when $SO_i > ESO_i$, SO_i = 1 and ESO_i = 0. The estimated demand using the logistic regression ($LRED_d$) is different from the total estimated demand using moving average or simple regression ($SDNW_d$). In other words:

$$SDNW_d > LRED_d = \sum_{i=1}^{\mathrm{DNW}_d} ESO_i \tag{10}$$

This is because most of the current orders will not be satisfied in the next week. On average, only 30% will be satisfied. Therefore, reducing $SDNW_d$ is necessary to reduce type 1 error. Another accuracy measure is the accuracy of the total demand for the next delivery week ($ADNW$), which can be found using the following equation:

$$ADNW_d = \frac{1 - |TSD_d - SDNW_d|}{\frac{\sum_{d=1}^{N} TSD_d}{N}} \times 100\% = \frac{N(1 - |TSD_d - SDNW_d|)}{\sum_{d=1}^{N} TSD_d} \times 100\% \tag{11}$$

Lead time is usually larger than the time between deliveries, which is usually one week. In this case, there is a short lead time in which there is an order after the last delivery and served during the next delivery. It will not be considered because it is not helpful for temporal aggregation, and it is considered an outlier.

The effect of orders not coming according to $FCFS$ is measured (% $FCFS$) in this paper using the following method: For the previously known data, find the forecasting accuracy assuming that the total satisfied demand is already known, and then orders are sorted from the oldest to the newest. Then, assign the value of 1 for the first $SDNW_d$ orders. If the $FCFS$ principle is fully applied, this accuracy should be 100%.

4. Results and Discussions

The demand of the last three months over time is shown in Figure 2. In the first few weeks, the previous three months' data are not complete and therefore the values are lower than usual. This is why these few weeks are excluded from the calculations in this paper. The last few weeks were also excluded. Generally, the demand follows a way in which it increases in the first few months of the year as shown. However, the percent of satisfied demand is more stable, except when the demand is more than 2500 orders, as shown in Figure 3. The average percent of satisfied demand, when the three-month demand is less than 2500, is 73.4%. However, when demand is higher than 2500, this percentage is reduced to be 68.1%. That means higher demand cannot be satisfied with the same level of lower demand. This fact can be useful to predict this percentage. Since there is no trend in the data, the seasonal effect can be found by finding the demand with the same week number from the previous year. The results were obtained from February until May, and the MAPE was about 9%. However, the peak at the beginning of 2015 was flatter than that for 2016. Therefore, another way for better forecasting accuracy is the moving average (with $n = 3$). In this case, MAPE is about 4%, which is more accurate. This is because the rate of changes in demand from one week to another is relatively slow. The average percent of satisfied demand in the year 2016 was about 68%. To forecast it, a moving average was used with $n = 3$. In this case, the average mean absolute deviation (MAD) value was about 0.02 for the months from February to May. If we take all the data for the last three months, then about 70% of the orders are usually satisfied. The older orders are usually satisfied. The average number of orders, which are not satisfied, is 759. For those, which are not satisfied until today, about 30% will be satisfied in the next week.

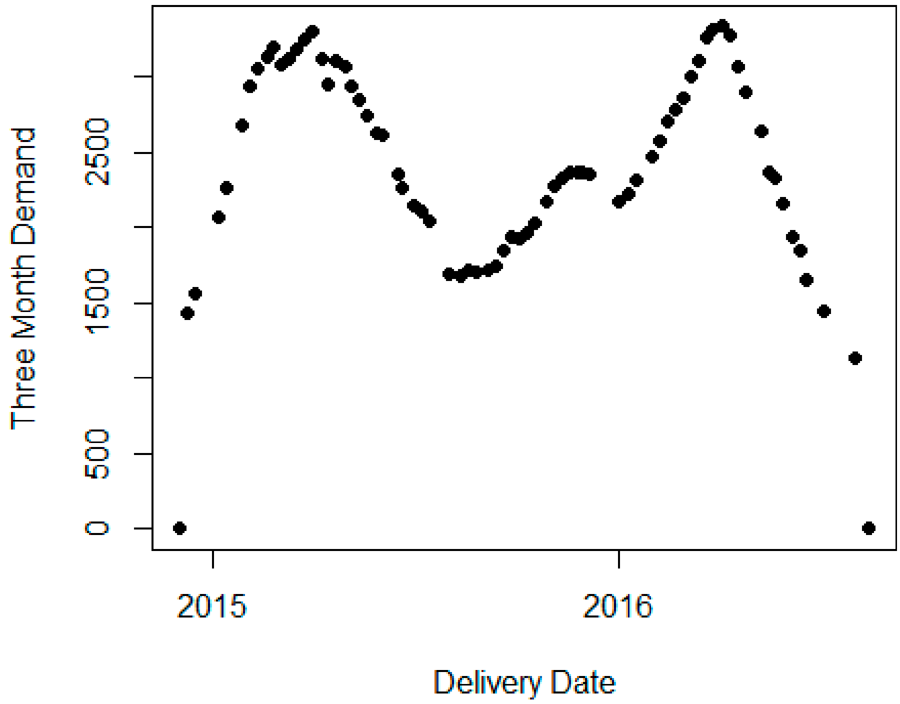

Figure 2. Total demand in the last three months.

For SDNW, the weekly size of arriving orders is rapidly changing, and it can be from 50 to 400. However, it can be predicted with some reasonable accuracy with two methods, namely, the moving average and simple regression, where the independent variable is DNW. Figure 4 shows the results of the two methods. The MAD for both of them is almost the same for the supply chain under consideration.

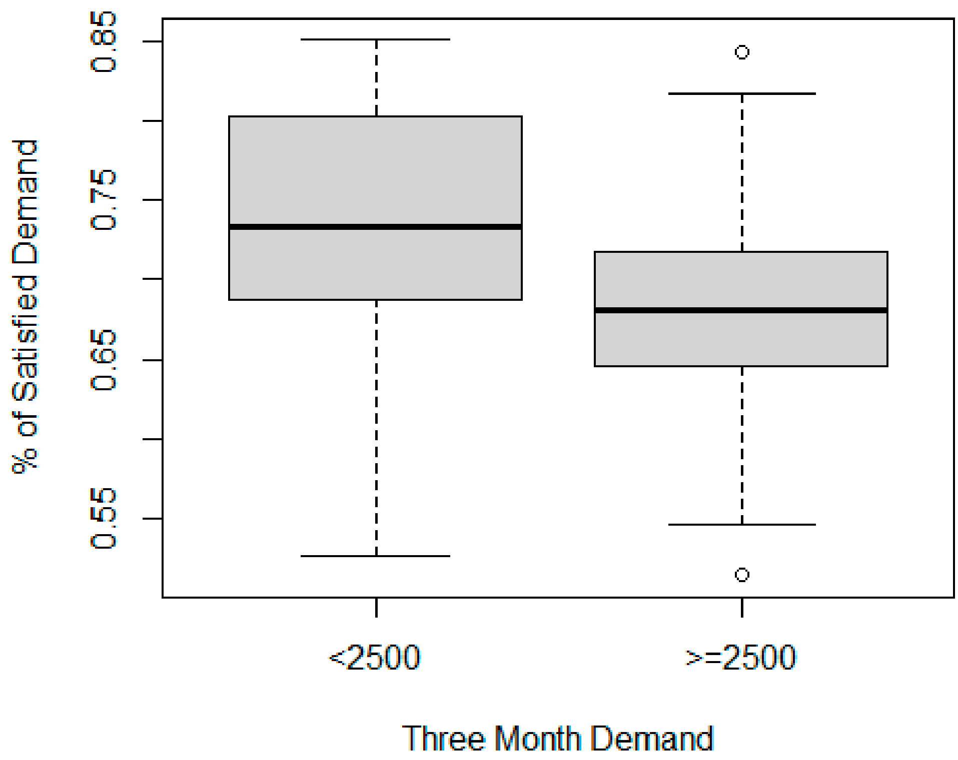

Figure 3. The effect of demand on the percentage of satisfied orders in the three-month period.

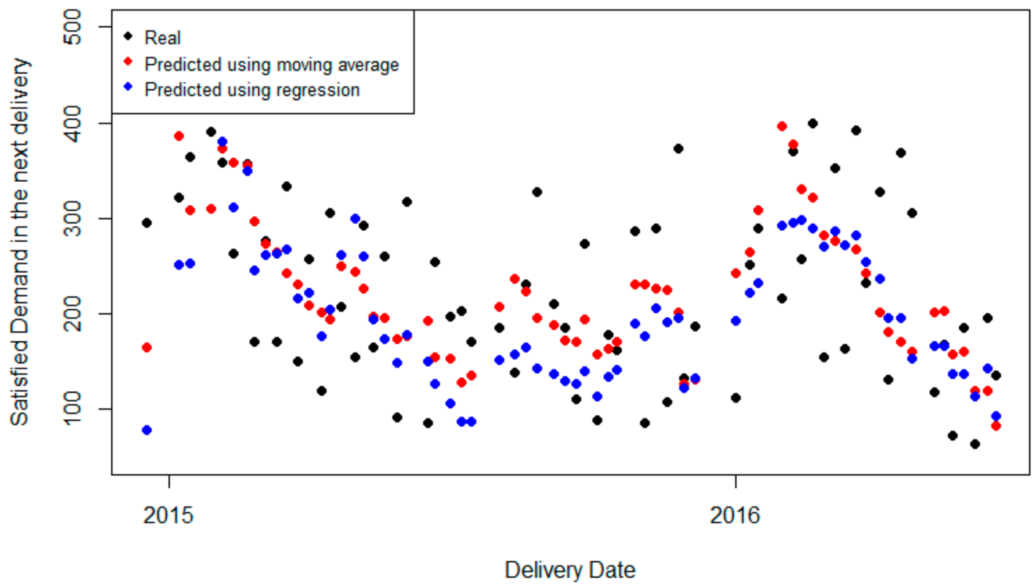

Figure 4. Prediction of satisfied demand in the next delivery (SDNW).

In Figure 5, the numbers inside the circles are the number of orders, which are in that date and with that specified lead time. Lead time follows a line with a slope that decreases with time. This is because customer orders are every day, but the delivery is usually every week. There are some parallel hypothetical lines these points follow, where each line represents a new week delivery. For example, on 12 February, 32 orders were triggered. Only 22 orders from them were satisfied with a lead time of 20 days. After one week, three of these orders were satisfied. The rest of the orders have to wait two more weeks. It is very clear that the exact time of the order is not possible to be predicted. However, if dynamic forecasting is to be used, the probability that a certain order will come in the next week can be obtained with reasonable accuracy. Most of the orders are satisfied within three or four successive weeks. Four of the orders coming on 15 February arrived earlier than usual.

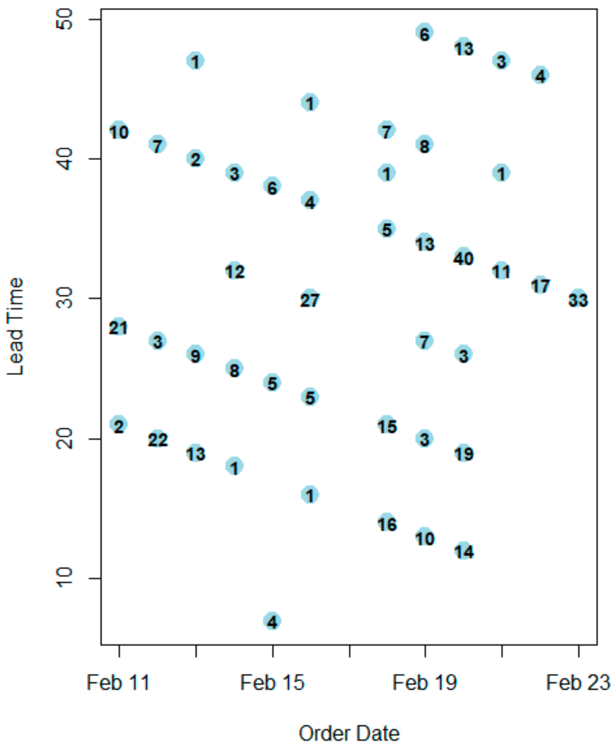

Figure 5. Lead time for orders triggered in one month.

Table 1 shows the main results of the study. It shows 10 weeks' data and also the grand average. Usually, the time between different deliveries is seven days, and sometimes one or two days earlier or later. The assumption here is that such a difference will not affect the data significantly. DNW's and FA's grand average values are significantly greater than the average values for the 10 weeks data. Therefore, the consideration will be for the grand averages. The general accuracy of forecasting is somewhat not very high, which is 72.41%. However, what really matters in this study is the type I error value, which is 0.07. This means that the accuracy in this regard is 93%. In other words, when it is expected that a certain order will come the next week, there is a chance that it comes 93% of the time. If better forecasting accuracy is needed, then a higher threshold value, such as 0.55 can be used instead of 0.5 for logistic regression. Type II error is indeed larger, with an average of 0.2, but this kind of error is not the critical one in this paper.

Table 1. The main results for 10 weeks plus the grand average results.

Delivery Date	P *	DNW	TSD	SDNW	ADNW	FA	Type II Error	Type I Error	% FCFS	PSD	LRED	Average OPSF
03/04/2015	7	993	276	273	98.7	79.2	0.2	0.0	76.4	27.8	12.8	17.6
11/03/2015	7	993	171	265	58.4	81.7	0.1	0.1	80.3	17.2	11.8	14.9
18/03/2015	7	1016	333	243	60.2	73.3	0.3	0.0	82.7	32.8	8.5	16.7
25/03/2015	7	906	150	231	64.2	77.9	0.1	0.1	73.5	16.6	9.7	14.6
01/04/2015	7	929	257	209	78.8	74.6	0.2	0.0	81.5	27.7	11.1	17.5
09/04/2015	8	825	119	201	63.7	82.8	0.1	0.1	80.1	14.4	11.8	18.4
15/04/2015	6	893	305	194	50.9	72.2	0.3	0.0	84.1	34.2	11.8	19.0
22/04/2015	7	994	208	250	81.4	78.4	0.2	0.1	77.9	20.9	12.2	12.8
01/05/2015	9	1071	154	244	60.2	85.1	0.1	0.1	83.9	14.4	12.1	15.2
06/05/2015	5	1013	293	226	70.4	76.9	0.2	0.0	78.7	28.9	12.7	16.6
13/05/2015	7	875	165	197	85.8	83.0	0.1	0.0	83.5	18.9	10.7	17.4
Average	7	955.27	221.00	230.27	70.23	78.64	0.17	0.05	80.24	23.06	11.38	16.41
Grand Average	7.8	770.69	218.80	224.21	70.27	72.41	0.20	0.07	74.56	29.72	16.81	15.72

* Period from the previous delivery (P).

TSD is the real demand, but SDNW is the expected one. The average difference between them is small. A metric to measure the accuracy of using SDNW to predict TSD is ADNW. The "% FCFS" term represents the effect of the FCFS rule. If the rule is completely followed by suppliers, then it should be 100%. However, it is only followed 74.56% of the time. This is why FA is not so high. The average value of PSD is 29.72%, which means that most of the current demand will not be satisfied in the next week because the lead time is long. However, logistic regression gives a percentage of 16.81 for LRED, which is almost half of PSD. This is to reduce the type I error. The average order-processing period so far (OPSF) is 15.72. This waiting, so far, is for the not satisfied demand in the current week. That means that all these orders must wait at least until the next week.

The general results show a reasonable accuracy, especially when it comes to the critical error, which is the type I error. This paper provides the decision maker with a way to know if the current order by a customer should wait for the next week for possible consolidation or not. This is, of course, if that order belongs to a customer having other orders. In case that a certain customer has only one order that comes in the current delivery, then there is no need for that order to wait. The forecasting process should be performed for all strategic suppliers and all orders in process. Since many customers are asking for orders from different suppliers, even with a lower level of accuracy, the model can provide an advantage. This is because if there are, for example, seven orders of one customer needed by supplies suppliers with an average accuracy of only 70%, then the accuracy for the probability that at least one of the seven orders will be satisfied in the next week will be much larger than 70%. Actually, it will be almost 100%. However, decision-makers should be aware of the impact of that on the service level. For example, if the due date is exceeded if consolidation is performed, then it should not be performed. Therefore, in the following cases, consolidation should not be performed, and therefore, there is no need for forecasting (for consolidation purposes) for that particular order:

1. If the due date is exceeded;
2. If there is only one order in process by the customer;
3. If inventory holding cost for one more week will cost more than the savings of reducing the number of shipments;
4. If the size of the current shipment is large enough to transship it without aggregating it with any further shipments.

5. Conclusions

This paper investigates the forecasting problem for a make-to-order supply chain based on real data, where the purpose is to predict if a certain order will be delivered to the consolidation center in the next delivery week or not. The lead time is usually long, and therefore, accurate forecasting based on little information is very challenging. Forecasting was performed in several steps. A reasonable accuracy level was obtained after applying weighted average, linear regression, and logistic regression (93% for logistic regression). Factors such as the size of demand for the last several weeks, the number of unsatisfied orders before the forecasted order, and the period so far in which the order is opened, are the independent variables. Freight consolidation is valuable to reduce the number of shipments and, therefore, the CO_2 emissions. Decision makers should weigh the advantages versus the disadvantages of temporal consolidation, and use it only after careful consideration about the other types of costs and the effect of increasing the lead time. Combining the forecasting process with temporal consolidation is performed in this paper for the first time. Future research can focus on comparing the forecasting done by supplier companies and 3PLs. There are some limitations in this study. For example, the lead time is assumed to be long and it takes usually some weeks, as in the case study. Moreover, if accurate data about the delivery times are provided by suppliers, then there is no need for forecasting by the 3PL. Suppliers usually have more information on the ground, and therefore can provide better forecasting. If inventory holding costs are high and transportation costs are low, then temporal consolidation might not be attractive.

Author Contributions: Conceptualization, M.A.; methodology, M.A.; software, M.A.; validation, M.A.; formal analysis, M.A.; investigation, M.A.; resources, B.S. and D.A.; data curation, M.A. and D.A.; writing—original draft preparation, M.A.; writing—review and editing, M.A. and B.S.; visualization, M.A. and B.S.; supervision, D.A.; project administration, B.S. and D.A.; funding acquisition, B.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by King Saud University, Riyadh, Saudi Arabia, under researchers supporting project number RSP-2021-145.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors would like to thank King Saud University, Riyadh, Saudi Arabia, with researchers supporting project number RSP-2021-145.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Çapar, I. Joint shipment consolidation and inventory decisions in a two-stage distribution system. *Comput. Ind. Eng.* **2013**, *66*, 1025–1035. [CrossRef]
2. Aboutalib, Z.; Agard, B. Improvement of freight consolidation with a data mining technique. In Proceedings of the Eighth International Conference on Information Systems, Logistics and Supply Chain (ILS Conference), Austin, TX, USA, 22–24 April 2020.
3. Kang, K.; Hong, K.-S.; Kim, K.H.; Lee, C. Shipment Consolidation Policy under Uncertainty of Customer Order for Sustainable Supply Chain Management. *Sustainability* **2017**, *9*, 1675. [CrossRef]
4. Alnahhal, M.; Ahrens, D.; Salah, B. Modeling Freight Consolidation in a Make-to-Order Supply Chain: A Simulation Approach. *Processes* **2021**, *9*, 1554. [CrossRef]
5. de Oliveira, M.B.; Zucchi, G.; Lippi, M.; Cordeiro, D.; da Silva, N.R.; Iori, M. Lead Time Forecasting with Machine Learning Techniques for a Pharmaceutical Supply Chain. In Proceedings of the 23rd International Conference on Enterprise Information Systems (ICEIS), Online Streaming, 26–28 April 2021; Volume 1, pp. 634–641. [CrossRef]
6. Asadzadeh, S.M.; Azadeh, A.; Ziaeifar, A. A Neuro-Fuzzy-Regression Algorithm for Improved Prediction of Manufacturing Lead Time with Machine Breakdowns. *Concurr. Eng.* **2011**, *19*, 269–281. [CrossRef]
7. Azadeh, A.; Ziaeifar, A.; Pichka, K.; Asadzadeh, S.M. An intelligent algorithm for optimum forecasting of manufacturing lead times in fuzzy and crisp environments. *Int. J. Logist. Syst. Manag.* **2013**, *16*, 186. [CrossRef]
8. Ioannou, G.; Dimitriou, S. Lead time estimation in MRP/ERP for make-to-order manufacturing systems. *Int. J. Prod. Econ.* **2012**, *139*, 551–563. [CrossRef]
9. Alnahhal, M.; Tabash, M.I.; Ahrens, D. Optimal selection of third-party logistics providers using integer programming: A case study of a furniture company storage and distribution. *Ann. Oper. Res.* **2021**, *302*, 1–22. [CrossRef]
10. Zhong, R.Y.; Huang, G.Q.; Dai, Q.-Y.; Zhang, T. Estimation of Lead Time in the RFID-Enabled Real-Time Shopfloor Production with a Data Mining Model. In Proceedings of the 19th International Conference on Industrial Engineering and Engineering Management, Changsha, China, 27–29 October 2012.
11. Schneckenreither, M.; Haeussler, S.; Gerhold, C. Order release planning with predictive lead times: A machine learning approach. *Int. J. Prod. Res.* **2021**, *59*, 3285–3303. [CrossRef]
12. Hanbazazah, A.S.; Castro, L.E.; Erkoc, M.; Shaikh, N.I. In-transit freight consolidation of indivisible shipments. *J. Oper. Res. Soc.* **2020**, *71*, 37–54. [CrossRef]
13. Aylak, B.L.; Ince, M.; Oral, O.; Süer, G.; Almasarwah, N.; Singh, M.; Salah, B. Application of Machine Learning Methods for Pallet Loading Problem. *Appl. Sci.* **2021**, *11*, 8304. [CrossRef]
14. Kosasih, E.E.; Brintrup, A. A machine learning approach for predicting hidden links in supply chain with graph neural networks. *Int. J. Prod. Res.* **2021**, 1–14. [CrossRef]
15. Lewandowska, A. Interactions between investments in innovation and SME competitiveness in the peripheral regions. *J. Int. Stud.* **2021**, *14*, 285–307. [CrossRef]
16. Öztürk, A.; Kayalıgil, S.; Özdemirel, N.E. Manufacturing lead time estimation using data mining. *Eur. J. Oper. Res.* **2006**, *173*, 683–700. [CrossRef]
17. Gyulai, D.; Pfeiffer, A.; Bergmann, J.; Gallina, V. Online lead time prediction supporting situation-aware production control. *Procedia CIRP* **2018**, *78*, 190–195. [CrossRef]
18. Burggräf, P.; Wagner, J.; Koke, B.; Steinberg, F. Approaches for the Prediction of Lead Times in an Engineer to Order Environment—A Systematic Review. *IEEE Access* **2020**, *8*, 142434–142445. [CrossRef]
19. Li, M.; Yang, F.; Wan, H.; Fowler, J.W. Simulation-based experimental design and statistical modeling for lead time quotation. *J. Manuf. Syst.* **2015**, *37*, 362–374. [CrossRef]

20. Indrianti, N.; Rakagandhi, B.; Wibawa, T. A Simulation Approach for Determining Delivery Time in a Waste Materials Based Make-To-Order Manufacturing Company. In Proceedings of the International Conference on Industrial Engineering and Operations Management, Bandung, Indonesia, 6–8 March 2018.
21. Sabeti, H. Lead-Time Quotation by Synergistically Modeling Real and Simulation Data. Master's Thesis, West Virginia University, Morgantown, WV, USA, 2018.
22. Dosdoğru, A.T.; İpek, A.B.; Göçken, M. A novel hybrid artificial intelligence-based decision support framework to predict lead time. *Int. J. Logist. Res. Appl.* **2021**, *24*, 261–279. [CrossRef]
23. Askin, R.G.; Hanumantha, G.J. Automated lead time estimation for manufacturing networks with dynamic demand. In Proceedings of the 2017 13th IEEE Conference on Automation Science and Engineering (CASE), Xi'an, China, 20–23 August 2017; pp. 994–999.
24. Mourtzis, D.; Doukas, M.; Fragou, K.; Efthymiou, K.; Matzorou, V. Knowledge-based Estimation of Manufacturing Lead Time for Complex Engineered-to-order Products. *Procedia CIRP* **2014**, *17*, 499–504. [CrossRef]
25. Gyulai, D.; Pfeiffer, A.; Nick, G.A.; Gallina, V.; Sihn, W.; Monostori, L. Lead time prediction in a flow-shop environment with analytical and machine learning approaches. *IFAC-PapersOnLine* **2018**, *51*, 1029–1034. [CrossRef]
26. Lingitz, L.; Gallina, V.; Ansari, F.; Gyulai, D.; Pfeiffer, A.; Sihn, W.; Monostori, L. Lead time prediction using machine learning algorithms: A case study by a semiconductor manufacturer. *Procedia CIRP* **2018**, *72*, 1051–1056. [CrossRef]
27. Pfeiffer, A.; Gyulai, D.; Kádár, B.; Monostori, L. Manufacturing Lead Time Estimation with the Combination of Simulation and Statistical Learning Methods. *Procedia CIRP* **2016**, *41*, 75–80. [CrossRef]
28. Kar, S.; Jha, K.N. Investigation into Lead Time of Construction Materials and Influencing Factors. *J. Constr. Eng. Manag.* **2021**, *147*, 04020177. [CrossRef]
29. Yamini, S.; Marathe, R.R. Mathematical model to mitigate planning fallacy and to determine realistic delivery time. *IIMB Manag. Rev.* **2018**, *30*, 242–257. [CrossRef]
30. Gacek, S. Due date assignment using neural networks for standard products in small batch and multi assortment make-to-order company. In Proceedings of the Carpathian Logistics Congress, Prague, Czech Republic, 3–5 December 2018.
31. Schuh, G.; Prote, J.-P.; Sauermann, F.; Franzkoch, B. Databased prediction of order-specific transition times. *CIRP Ann.* **2019**, *68*, 467–470. [CrossRef]
32. Berlec, T.; Potocnik, P.; Govekar, E.; Starbek, M. Forecasting lead times of production orders in SME'S. *Iran. J. Sci. Technol. Trans. A Sci.* **2010**, *34*, 521–538.
33. Walker, S.H.; Duncan, D.B. Estimation of the probability of an event as a function of several independent variables. *Biometrika* **1967**, *54*, 167–179. [CrossRef]