Supply Chain Analytics 1 (2023) 100003

Contents lists available at ScienceDirect

Supply Chain Analytics

journal homepage: www.sciencedirect.com/journal/supply-chain-analytics

A novel machine learning model for predicting late supplier deliveries of low-volume-high-variety products with application in a German machinery industry

Fabian Steinberg[a,*,1], Peter Burggräf[a,2], Johannes Wagner[a,3], Benjamin Heinbach[a,4], Till Saßmannshausen[a,5], Alexandra Brintrup[b,6]

[a] Department of Mechanical Engineering, Chair of International Production Engineering and Management (IPEM), University of Siegen, Siegen, Germany
[b] Institute for Manufacturing, Department of Engineering, University of Cambridge, CB3 0FS Cambridge, UK

ARTICLE INFO

Keywords:
Supply chain management
Prediction methods
Machine learning
Supervised Learning
Regression analysis
Data analysis

ABSTRACT

Although Machine Learning (ML) in supply chain management (SCM) has become a popular topic, predictive uses of ML in SCM remain an understudied area. A specific area that needs further attention is the prediction of late deliveries by suppliers. Recent approaches showed promising results but remained limited in their use of classification algorithms and struggled with the curse of dimensionality, making them less applicable to low-volume-high-variety production settings. In this paper, we show that a prediction model using a regression algorithm is capable to predict the severity of late deliveries of suppliers in a representative case study of a low-volume-high-variety machinery manufacturer. Here, a detailed understanding of the manufacturer's procurement process is built, relevant features are identified, and different ML algorithms are compared. In detail, our approach provides three key contributions: First, we develop an ML-based regression model predicting the severity of late deliveries by suppliers. Second, we demonstrate that prediction within the earlier phases of the purchasing process is possible. Third, we show that there is no need to reduce the dimensionality of high-dimensional input features. Nevertheless, our approach has scope for improvement. The inclusion of information such as component identifiers may improve the prediction quality.

1. Introduction

Today, companies source their goods from all over the world, with multi-modal transport chains delivering everything from simple components to highly complex products. Efficient procurement of goods is largely dependent on supplier performance. Goods that are delivered with insufficient quality, quantity, or with a time delay, lead to disruptions in manufacturing companies that need these goods in their production. Those disruptions are especially serious for manufacturers whose value stream is critically dependent on the assembly process, where several material flows from different suppliers and the in-house manufacturing converge [1]. Here, only a single missing component can impede the timely start of the entire assembly process involving up to several thousand components [2]. Thus, to ensure a timely start of the assembly process, and consequently, to meet their delivery dates, for manufacturing companies it would be helpful to predict potential delays in their upstream supply chain.

With the advances in data analytics utilizing machine learning (ML) and the availability of large-scale, unstructured data sets, ML-based prediction models are becoming more and more established. Despite these advances, in supply chain management (SCM) recent review articles have identified a predominance of descriptive analytics rather than predictive analytics, except for demand forecasting [3–5]. As the complexity of supply chains is continuously increasing, predicting

* Corresponding author.
E-mail address: fabian.steinberg@uni-siegen.de (F. Steinberg).
[1] **ORCID**: 0000–0002-9842–5343
[2] **ORCID**: 0000–0001-9018–9959
[3] **ORCID**: 0000–0002-3226–2457
[4] **ORCID:** 0000–0002-2552–2855
[5] **ORCID:** 0000–0002-1061–8686
[6] **ORCID**: 0000–0002-4189–2434

https://doi.org/10.1016/j.sca.2023.100003
Received 9 February 2023; Received in revised form 2 March 2023; Accepted 3 March 2023

supply chain disruptions before they occur is becoming increasingly important as well [6]. Recent developments in digitalization technologies, such as the internet of things or artificial intelligence, present novel opportunities for predicting disruptions in SCM [7,8]. However, contributions presenting such prediction models are limited in SCM literature [3–5].

In the body of literature of the specific area of identifying and quantifying late deliveries of suppliers, currently, there are only two approaches available: First, the research of Brintrup et al. [9], and second, the research of Baryannis et al. [10]. Both approaches are limited in their use of classification algorithms and struggle with the curse of dimensionality, making them less applicable to low volume high variety manufacturing settings. In addition, the prediction models are based on information that is available after the order has been placed. In both approaches, we see the following three shortcomings, which we want to address in our manuscript: First, unlike a classification model, a regression model can make predictions of delivery delays in calendar days, thus assessing the severity of a delay and providing a more valuable prediction. In real-world manufacturing applications, the evaluation of the severity of a delay is essential to select and prioritize appropriate countermeasures. Nevertheless, a regression model predicting delivery delays of orders placed at suppliers of low volume high variety manufacturers is currently not available in the body of literature. Second, the time of a prediction is essential for the implementation of potential countermeasures - in our work, we define the time of prediction as the point of time in the purchasing process when the prediction model is applied, and the prediction is made; this could be, for example, the time of creating an order request, the time of placing an order, or time of receiving a delivery confirmation. An earlier prediction than at the time of placing an order as used by Brintrup et al. [9] and Baryannis et al. [10], such as after the creation of an internal purchasing request, would provide more opportunities for countermeasures in case of a predicted delay. However, typically, less information is available at earlier time points, which can cause prediction models to have lower quality. But, in the specific area of predicting late deliveries of suppliers, there has been no study of the influence of the time of the prediction on the model quality. Third, both available approaches limit the scope – e.g., they exclude components that are ordered less than five times – because they struggle with the curse of dimensionality. Especially in low volume high variety environments, however, it is typical that a large proportion of the components are designed individually for each customer need and are therefore procured only once. Such a restriction in the scope would then lead to a large proportion of the required components no longer being considered in the prediction model, which in turn limits the practical applicability of the model.

Reflecting on the shortcomings of the above-mentioned approaches, in this work we focus on the following three research questions (RQ):

- RQ1: Are regression algorithms capable to predict delivery delays of orders placed at suppliers of a low volume high variety manufacturer?
- RQ2: What is the impact of the time of the prediction on the model quality?
- RQ3: Is the curse of dimensionality an issue when setting up a regression model predicting delivery delays of orders placed at suppliers of a low volume high variety manufacturer?

To answer these research questions, we conduct a case study at a machinery manufacturer focusing on the prediction of potential delays of components delivery dates ordered at suppliers. A case-based research approach is an objective, detailed investigation of a current phenomenon where the researcher has little control over real events [11]. One motivation for the case-based research approach is to gain insights for real needs of manufacturing companies, rather than to develop theories without practical relevance [12]. Furthermore, a case-based research approach has already been successfully applied in the area of predicting delivery delays [5,9,10]. Accordingly, a case-based research approach is an appropriate method to answer the research question and to investigate the working hypothesis.

In detail, we set up a two-stage experimental plan for this purpose. In the first stage, the quality of ML-based regression models predicting delivery delays is compared at different times of the prediction and thus the influence of the time of the prediction on the regression model quality is quantified. At each time of the prediction, a range of standard ML algorithms is applied to an identical data set to allow comparability between prediction time points and algorithms. With this comparison, we can also evaluate if regressions algorithms are capable to predict delivery delays. Thus, in the first stage of the experimental plan, we will answer RQ1 and RQ2. In the second stage, we then investigate the impact of reducing the dimensionality of high dimensional input features on the model quality by comparing different exclusion criteria and thus answering RQ3. The setup of all ML models within the experimental plan follows the established procedure model Cross Industry Standard Process for Data Mining (CRISP-DM) [13,14] consisting of the six phases business understanding, data understanding, data preparation, modelling, evaluation, and deployment. To ensure comparability across our experimental plan, the datasets, features, and ML algorithms utilized remain equal in both stages of the experimental plan.

The evaluation of the case study provides the following three main contributions:

We show that an ML-based regression model can predict delivery delays of orders placed at suppliers of machinery manufacturing in calendar days.

We demonstrate that an early prediction within the purchasing process based on information available after creating the internal order request is possible and only slightly worse than a later prediction.

We show that there is no need to reduce the dimensionality of high dimensional input features.

This paper is structured as follows. First, Section 2 introduces the state of the art in predictive data analytics in SCM with a focus on available approaches predicting delays of orders. Section 3 is structured according to the CRISP-DM framework giving a description of the case study dataset and details about the feature selection and engineering as well as the setup of the experimental plan, the ML models and the results. Further, a comparison of our model performance with recent approaches is conducted. Subsequently, Section 4 critically reviews the limitations of our approach and the results obtained. Furthermore, the implications for further research are derived. Finally, a summary is given in the last section.

2. State of the art: predictive analytics in supply chain management

To reduce the impact of a disruption, there are typically two options. First, to minimize its risk of occurrence, and second to strive for a resilient supply chain that quickly returns to its original state after a disruption [15–17]. These two options are covered by two individual domains in SCM, namely, supply chain risk management (SCRM) and supply chain resilience. In both and the superordinate field SCM as well, data analytics is one of the core tools used. Waller and Fawcett [18] define the term data analytics in SCM as '*the application of quantitative and qualitative methods from a variety of disciplines in combination with SCM theory to solve relevant SCM problems and predict outcomes, taking into account data quality and availability issues*'. They further classify predictive analytics as a subset of data analytics to improve supply chains and mitigate risks by forecasting what could probably happen in the future. In contrast, Wang et al. [3] and Nguyen et al. [4] give a wider differentiation of data analytics in SCM. They classify the available approaches in descriptive, predictive, and prescriptive analytics. Descriptive analytics in SCM focuses on what happened in the past (see, for example, [19–21]). Predictive analytics attempts to predict and explain events that will occur in the future (see, for example, [22–24]). Prescriptive analytics is using data and algorithms to find alternative decision options (see, for example, [25–27]). Out of these three categories, current research focuses mainly on prescriptive analytics rather than descriptive and predictive analytics [4]. Nevertheless, as typical for data analytics in general and not just in the area of SCM, the performance of prescriptive models relies on descriptive and predictive models [3,4]. Thus, the aforementioned review papers call for new research in descriptive and predictive analytics in SCM. Hence, we contribute to the body of literature with a case study focusing on predictive analytics in SCM.

Bienhaus and Haddud [28], Ivanov et al. [29], and Queiroz et al. [30] highlight the premise and use of big data and Artificial Intelligence in the digital transformation of the procurement process as one of the key factors to enhance the competitiveness, efficiency and profitability of companies' supply chains. With the continuously increasing availability of a wider volume, velocity, and variety of data, new opportunities arise to revolutionize the impact of data analytics approaches [18,31,32].

In SCM as a whole, ML and other data mining techniques are frequently considered for demand forecasting (see for example, [33–35]), determining retail prices in supply chains including the handling of financial flows (see, for example, [36–38]), or dealing with the effects of low-frequency high-impact disruptions on supply chains such as the COVID-19 pandemic (see, for example, [39–42]). In the specific subdomains of procurement and logistics current research mainly supports the selection of potential suppliers for specific products (see, for example, [43–45]) or deals with problems of vehicle routing [4,46], but missing material due to late deliveries is a neglected area of research [47]. Models predicting late deliveries of suppliers are still rare. To the best of our knowledge, we identified only two articles focusing on predicting late deliveries of suppliers based on real data sets using machine learning.

Baryannis et al. [10] proposed an ML-based approach predicting late deliveries of suppliers with a focus on their interpretability to be able to support decision-making following the prediction. Given a real data set of a multi-tier aerospace manufacturing supply chain consisting of product data such as the part number and price, order data such as due dates, quantities ordered and original delivery requests, and delivery data such as receipt date and quantity receipt they compare the performance and interpretability of support vector machines (SVM) with decision trees (DT). Prioritizing interpretability over performance they recommended DT as the ML algorithm of choice resulting in slightly worse performance metrics. While we agree with the need for more interpretable ML in SCM, we postulate that other algorithms such as ensemble algorithms or ANN which suffer from interpretability also need to be studied to be able to present the full range of options to the decision-maker.

Brintrup et al. [9] presented a case study at an original equipment manufacturer (OEM) predicting delivery delays of Tier 1 suppliers also based on historical product data such as product description and product type, order data such as order date and supplier ID, and delivery records such as the received date. Comparing five ML algorithms they identified a random forest algorithm (RF) outperforming SVM, logistic regression, linear regression, and k-nearest neighbour algorithm. Similar to Baryannis et al. [10], more complex ML algorithms such as ANN might have performed better but remained uncovered.

Further, one of their biggest challenges was the curse of dimensionality due to high variability in their categorical features leading to a high number of variables in their feature space. To reduce the dimensionality, they excluded data with less than five samples for each categorical attribute. Thus, they excluded for instance suppliers who delivered less than five times or components that were ordered less than five times. This restriction in the variability of the input data might be a limitation when transferring the approach to industries that focus on low volume high variety customized production, where components ordered at suppliers may vary strongly. Thus, we postulate that the impact of such a limitation on models predicting delivery delays in low volume high variety production needs to be investigated.

In summary, there are several approaches available focusing on data analytics in supply chain management. However predictive analytics in SCM remains an understudied topic. A specific area that needs further attention is the identification and quantification of late deliveries. Extant approaches are limited in their use of classification algorithms and struggle with the curse of dimensionality, making them less applicable to low volume high variety settings.

Hence, we contribute to the body of literature with a novel case study in predictive analytics in supply chain management using machine learning – specifically in predicting delivery delays of suppliers with a supervised learning approach using a real data set from a machinery manufacturer. Here, we compare simple ML algorithms such as DT, RF, or SVM with more sophisticated approaches such as ANN. Furthermore, we analyse the impact of reducing the dimensionality of high dimensional input features on the model quality and the practical usability of the model by comparing different exclusion criteria.

3. Case study

For our case study, we choose an OEM in the German machinery manufacturing industry that builds complex products made up of several thousand individual components and several hundred sub-systems. The products are typically individually designed for customers' needs. The upstream supply chain consists of Tier 1 suppliers for finished components that are used directly in assembly, as well as Tier 1 suppliers for raw materials that are then mechanically processed in the company's production facilities. Approximately 80% of the total number of components are purchased as finished components and 20% are processed within the company. Most of the raw material supply for in-house production is decoupled via storage with suitable safety stocks. Further, in-house production is set-up to be flexible so that potential delays in raw material supply can be partially compensated. In contrast, most of the finished components are ordered individually for each customer project, so delays directly influence the timely start of the assembly. Thus, in our case study, we focus on the late deliveries of finished components.

To answer the research questions, we set up a two-stage experimental plan. In stage one we quantify the impact of the time of the prediction on the quality of ML-based regression models predicting late deliveries and simultaneously analyse if an ML-based regression model is capable to predict delivery delays in calendar days. In stage two we evaluate if there is a need to limit the scope to overcome the curse of dimensionality when predicting late deliveries with ML-based regression models within our exemplary case. To develop the different ML models we applied the established CRISP-DM procedure model (see Fig. 1) [13,14]. Further, to ensure comparability across our experimental plan, the datasets, features, and ML algorithms utilized remain equal in both stages of the experimental plan. Thus, the general process to develop the different ML models remain equal as well. Consequently, the next subsections are following the phases of CRISP-DM business understanding, data understanding, data preparation, modelling, evaluation, and deployment. As we mainly focused on the development of the model, we excluded the last phase Deployment.

3.1. Business understanding

The Business understanding phase typically includes a description of the business problem and a transfer of the business problem into concrete requirements and objectives for further data analysis. Thus, the first phase of the CRISP-DM provides a central basis for all the following steps and decisions in the data mining process.

As the business problem is delays in the assembly process due to delayed deliveries of finished components, the objective from a

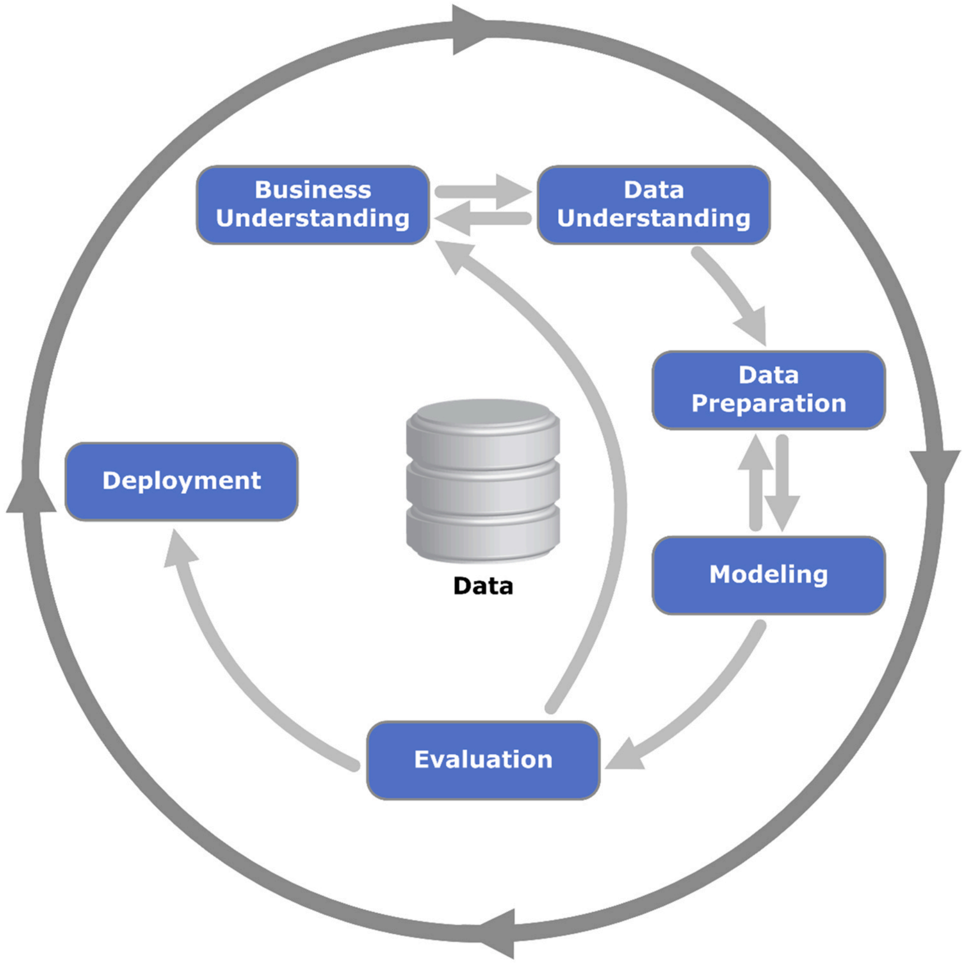

Fig. 1. Procedure of the CRISP-DM framework applied in the case study [13,14,48].

business perspective was to prevent these delays. Predicting potentially delayed deliveries of finished components ordered at the suppliers can support the OEM's purchasing department to take countermeasures such as speeding up the supplier's manufacturing process, choosing a different means of transport for the delivery, or even choosing a different supplier. The OEM's business process itself is typical for a machinery manufacturer. The relevant components for assembly are first defined in a design and material planning process. Then, after a make-or-buy decision, either a production order for in-house production or a purchasing request is created in the company's Enterprise Resource Planning (ERP) system for each required component. The purchasing request then initiates the purchasing process following the established standards. Here, we wanted to support the purchasing process with a prediction of potentially delayed deliveries as early as possible. Together with the domain experts at the OEM, we identified three potential times of the prediction, which is defined as the point in time within a procurement process a prediction model is applied:

(1) *Purchasing request:* A prediction based on all information available after creating the purchasing request in the ERP System.
(2) *Order placement*: A prediction immediately after placing an order at a selected supplier.
(3) *Delivery confirmation*: A prediction immediately after receiving a delivery confirmation of the supplier including the confirmed delivery date.

Although it may seem trivial that a prediction at a later point in time might be of higher quality due to more available information about the purchasing process, an early prediction would be more helpful to take effective countermeasures. Here, one objective of our case study is to identify a good trade-off between the time of the prediction and the prediction quality.

Further, the type of prediction is interesting as well. Binary classification of the deliveries in *late* and *in time* as applied by Brintrup et al. [9] and Baryannis et al. [10] is less valuable than a regression model predicting potential delays in working days. A prediction in working days would give additional information about the severity of a delay. Thus, a regression was our modelling approach of choice. Here, we applied several ML algorithms such as tree-based algorithms, support vector machines, or neural networks utilizing the Scikit-learn library in Python.

Hence, we transformed the business problem – missing components at the start of assembly – into a machine learning problem – predicting delivery delays in the supply of externally purchased finished components utilizing a supervised learning approach. To compare the prediction quality of the different ML models to be set up in the modelling phase first defined metrics for each prediction model. More specifically, we selected the following established metrics for regression models as our metrics of choice: mean absolute error (MAE), root mean squared error (RMSE), and coefficient of determination (R^2). Here MAE and RMSE are defined as

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |y_{i,true} - y_{i,predicted}| \tag{1}$$

$$RSME = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (y_{i,true} - y_{i,predicted})^2} \tag{2}$$

where n is the number of samples considered, $y_{i,true}$ is the actual value, and $y_{i,predicted}$ is the predicted value. R^2 is defined as

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_{i,true} - y_{i,predicted})^2}{\sum_{i=1}^{n} (y_{i,true} - \bar{y}_{true})^2} \quad (3)$$

where $\bar{y}_{true}$ is the empirical mean of all y_{true}. Further details about the equations can be looked up in [49,50].

3.2. Data understanding

In the second phase, data understanding, following Wirth and Hipp [14], we collected and analysed the data to develop a solid understanding of the dataset. For our prediction model, we included data from two business processes of the company under consideration. First, the material planning process was collected, which included information about the bill of material (BOM) and demand dates of components for processing in assembly based on backward scheduling. Second, the purchasing process containing information about the orders including the respective deliveries was collected. As both processes are executed and documented in the company's ERP system, this was also our data source of choice. The data export with a period under review of three years consisted of two separate CSV-files containing purchasing orders and BOM items. With one ID field, we were able to merge the two separate files. When merging the data, since predicting delivery delays of the purchasing orders was our target, we kept the orders as our object under consideration and expanded them with information from the BOM. The total data set comprised 119,610 purchasing orders with information from 17 different data fields (see Table 1). As the dataset is confidential, we are not allowed to make it available to third parties and cannot publish it.

As target variable for our prediction model and thus, to predict potential delivery delays we calculated a delivery date lateness (DDL) considering the actual delivery date and the demand date. In detail, we utilized the formula

$$DDL = Deliverydate - Demanddate \quad (4)$$

to calculate the DDL. Here, a negative DDL indicates a delivery before the demand date and a positive DDL indicates a delayed delivery. Thus, on the one hand, companies can predict the severity of delivery delays in calendar days for late deliveries. On the other hand, with the prediction of the duration components are delivered before the demand date, companies can allocate space for inventory and calculate days of inventory turnover predictably.

Next, we performed an exploratory data analysis to understand the main characteristics of our dataset using statistical graphics such as boxplots, scatter plots, and histograms. Here, we first analysed the distribution of the DDL as our target variable for our prediction model (see Fig. 2). It is noticeable that with a portion of approx. 86% of most of the orders were delivered before or in time to the demand date, and with a portion of approx. 14% only a few orders were delivered delayed. Thus, only a slight portion of all orders is the main reason for missing material in the assembly. Further, it is noticeable that few orders are delivered several months or even up approx. 0.75 years before the demand date. Potential reasons for these high DDLs are shifts of the assembly after the actual delivery date due to shifts in the customer's order or an assignment of material from a previous purchasing order, that has been placed in stock, to a different assembly order. Nevertheless, we included these purchasing orders in our prediction model, as these were real and not data errors.

Moreover, we analysed the product portfolio based on CAD drawing numbers (see Fig. 3) and supplier structure based on supplier-IDs (see Fig. 4) to get a better understanding of the variety of categorical data fields in our data set. This analysis was also an indicator of the curse of dimensionality [51] that might affect our prediction model. Looking at the product structure, a portion of 61% of all components were ordered twice or less, whereas with a portion of 4% some components were ordered 20 or more times. Additionally, the components ordered twice or less comprised 17% of all orders placed and the components ordered 20 or more times comprised 30% of all orders placed. Therefore, limiting the scope of consideration to overcome the curse of dimensionality to, for example, components that have been ordered at least twice would mean that 61% of all components and thus 17% of all orders would not be considered in the prediction model. Other authors such as the authors in [9] limit their scope even stricter to components that have been ordered at least five times. Nevertheless, excluding 61% of all components, when limiting the scope to components that have been ordered at least twice was a non-negligible limitation in our case study. Consequently, within our modelling phase, we additionally analysed the trade-off between the scope limitation in the variety of components and the achievable model quality. In detail, we quantified the impact of such a limitation on the model quality by comparing prediction models with different exclusion criteria.

Looking at the supplier structure, in addition to the product portfolio, it was noticeable that 22% of all suppliers received a maximum of three orders, which, however, only accounted for 0.2% of all orders. On the other hand, 0.3% of all suppliers received 10,000 each or more orders and accounted for 22% of all orders. Here, limiting the scope to suppliers who received more than three orders seemed reasonable, since only 0.2% of all orders would have been excluded. Thus, a limitation within the supplier structure could help to overcome the curse of dimensionality excluding only a tiny portion of all orders. Nevertheless, further analysis showed, that 60% of the respective orders placed at

Table 1
Overview of data fields.

Data field	Data format	Description
Product description	Text	Short description of the component
Drawing number	Alphanumeric	Unique drawing identification
Order quantity	Integer	Number of components ordered
BOM item created	Date	When the material planning process is completed
Demand date	Date	When the component is required in the assembly
Order created	Date	When purchasing process was initiated
Order date	Date	When the order was placed
Requested delivery date	Date	Requested delivery date of the purchasing department
Confirmed date	Date	Confirmed delivery date of the supplier
Confirmation received	Date	When the supplier confirmed the delivery
Delivery date	Date	When did the supplier deliver
Order-ID	Integer	Purchase order number
Order method	Alphanumeric	Category indicating how the order was placed
Supplier-ID	Integer	Unique supplier identification
Supplier	Text	Legal name of the supplier
Material	Text	Short description of the material (e.g., S235)
Gross weight	Integer	Gross weight of the component

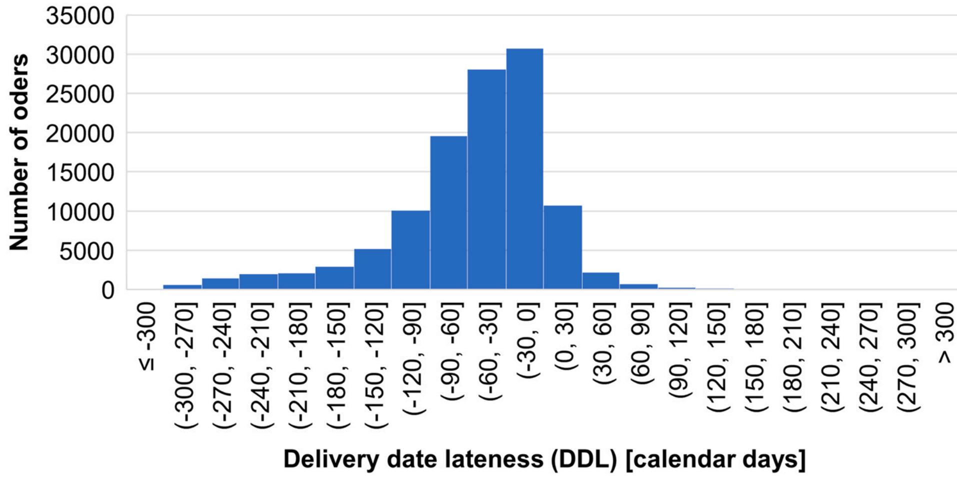

Fig. 2. Distribution of the delivery date lateness (DDL).

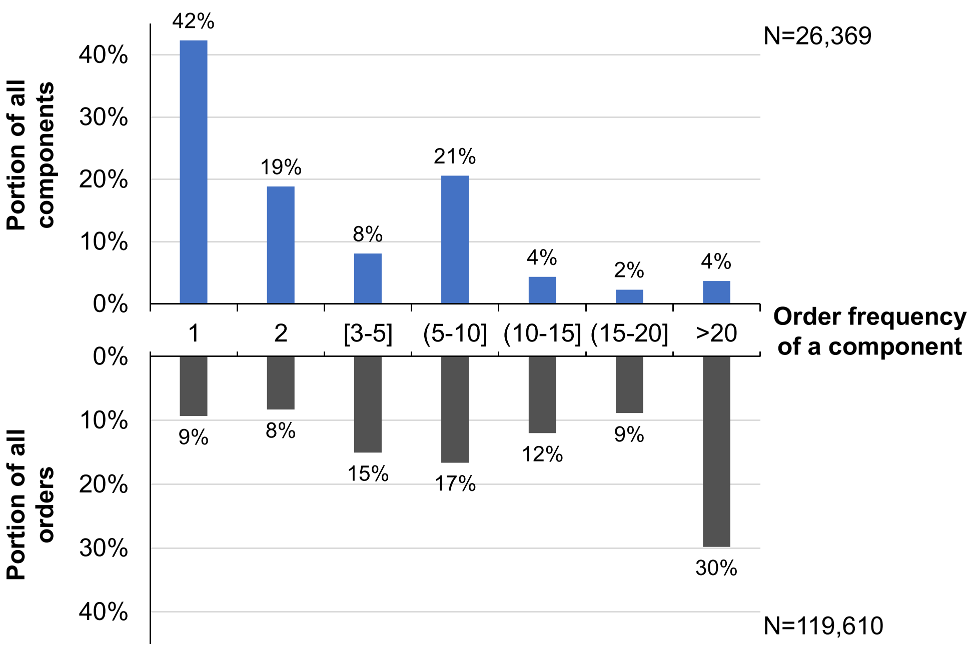

Fig. 3. Distribution of the product portfolio.

suppliers that received three or fewer orders were delivered too late. Comparing this portion of late deliveries with the overall portion of too late deliveries, which was 14%, revealed an outstanding potential of late deliveries of suppliers who received three or fewer orders. Thus, even this seemingly negligible limitation in the supplier structure would have meant that suppliers with outstanding potential for disruption would have been excluded from the scope, which was another non-negligible limitation in our case study. Consequently, also in the supplier structure, we analysed the trade-off between the scope limitation in the variety of suppliers and the achievable model quality by comparing prediction models with different exclusion criteria.

3.3. Data preparation

With the understanding gained about the dataset, we then started to prepare the final dataset for our prediction models. First, we looked at the features in date-format. Using a date as an input feature for a supervised learning approach means the prediction model is trained on those historical dates as well. In our case, whenever the model will be used in the future, the input dates would be in a different year and not comparable to the dates the model was trained on. Thus, we transformed the features into a date-format. Instead of the actual date as one feature, we used four features: The first three were the year, month, and day of the date in integer-format, and the fourth was the deviation of the respective date to the demand date also in integer-format.

Next, following Kuhn and Johnson [52], we performed a correlation analysis to identify the relevant features for our prediction model. For correlation between categorical features, we calculated Cramér's V, and for correlation between continuous variables, we calculated the Pearson correlation coefficient. The interpretation of the correlation coefficient followed the definitions given by Akoglu [53]. First, we analysed the correlation between the target variable DDL and our input features revealing only weak correlations of the date fields in the first three transformations year, month, and day. Instead, the fourth transformation of all date fields, which was the deviation of the respective date to the demand date, showed strong correlations. Thus, we excluded the transformations year, month, and day and only kept the transformation into the deviation to the demand date. Further, we analysed the correlation within our input features to identify feature dependencies. This analysis showed a strong correlation between the drawing number and the product description, and a strong correlation between the supplier-

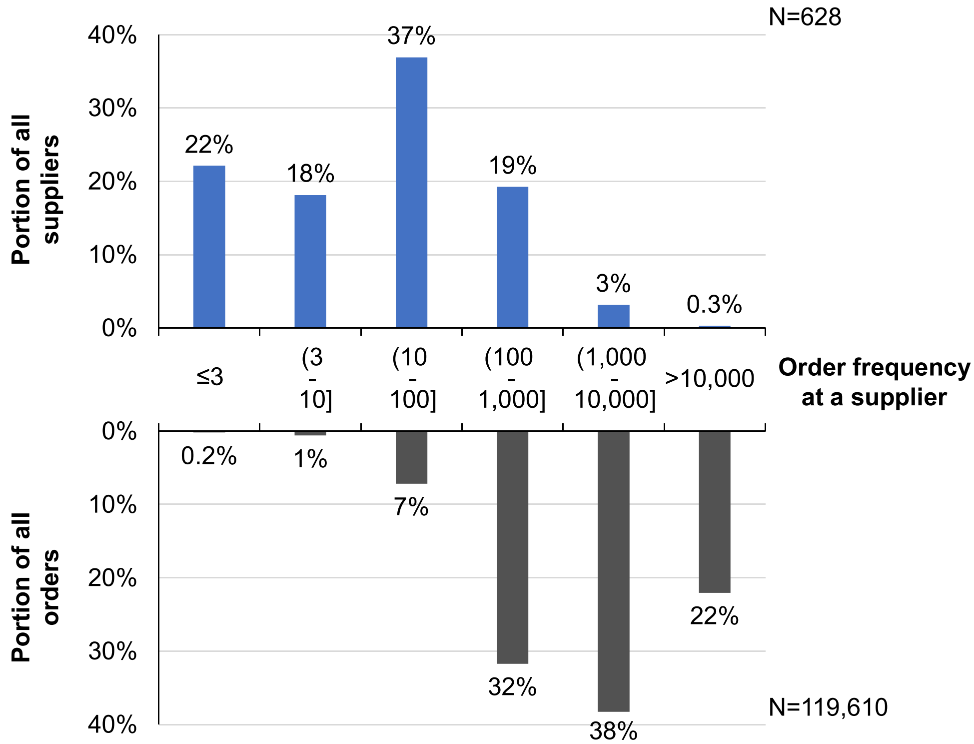

Fig. 4. Distribution of the supplier structure.

ID and the supplier in text format. Consequently, we excluded the product description and the supplier in text format, and only kept the drawing number and supplier-ID to reduce dependencies within our input features. Thus, in summary, we selected the following twelve independent variables to be the input features for our ML models:

- *Drawing Number:* The drawing number is a unique identifier of a component allowing one to determine if a component is ordered multiple times. Thus, based on the drawing number, delays of a component ordered in the past can be used as a base for predicting a potential delivery delay in future. Consequently, this variable is important for predicting delivery delays.
  Order quantity: The lead time of a production lot typically scales with the number of components to be manufactured. Consequently, the number of ordered components influences the delivery times of suppliers and is thus important for the prediction of delivery date delays.
- *Material:* The idea behind using the material as an independent variable is that certain materials from which a component is made could be delivered more quickly than rare materials. Therefore, the use of the material could reveal patterns that have an influence on the prediction of delivery date delays.
- *Gross weight:* The effort required to handle and transport components depends on the weight of a component. Further, the weight of a component could be an indicator of the effort required to manufacture it. Thus, the usage of the gross weight of a component is a relevant variable for predicting delivery delays.
- *BOM-item created:* The creation of the BOM item is the first event at which the need for a component is systemically recorded and thus known. At the company under consideration, the creation of demand is the responsibility of the design department. Without that demand, the purchasing process is not initiated. Thus, the duration between the date of creating the BOM item and the demand date is relevant for predicting delivery delays.
- *Order created:* After the BOM item is created and after the decision has been made to procure the component, a purchasing order is created and thus, the purchasing process is initiated. Consequently, delivery delays depend on the available duration between order creation and the demand date.
- *Order date:* The next step in the purchasing process is ordering the component from a supplier. The duration between the order date and the demand date represents the maximum delivery time for a supplier to deliver on time. Thus, the order date has an influence on potential delivery delays.
- *Request delivery date:* With the order of a component, a requested delivery date is submitted to the supplier. Here, the respective purchaser can make manual adjustments to the demand date – the purchaser can request a delivery date before, contemporary or after the demand date. Consequently, the requested delivery date is essential when predicting delivery delays.
- *Order method:* When placing orders, the company has different levels of atomization and different communication protocols to their supplier – e.g., Electronic Data Interchange (EDI), E-Mail, or phone calls. This might have an influence on the delivery times and thus on predicting delivery delays.
- *Supplier-ID:* The supplier ID is a unique identifier of the supplier allowing us to determine if several components are ordered from the same supplier. Thus, based on the supplier ID, delays of deliveries of the respective supplier in the past can be used as a base for predicting a potential delivery delay in future. Consequently, this variable is important for predicting delivery delays.
- *Confirmed date:* With the confirmation of the order, the supplier provides an estimated delivery date. This information is important for predicting delivery delays.
- *Confirmation received:* The duration between ordering and receiving a confirmation from a supplier might be an indicator of suppliers' available capacities and delivery performance. Thus, this information might also indicate delivery delays.

Next, since tree-based classifiers and neural networks from the Scikit-learn library can only be trained on numerical variables in Python [54], we needed to convert our categorical variables into numerical variables. One common and established method for this conversion is One-Hot-Encoding. But, without any limitation in the

Table 2
Overview of the transformed data set for the prediction models.

Feature	Time of the prediction		
	1. Purchasing Request	2. Order Placement	3. Delivery Confirmation
Drawing number (Binary-Encoded to 15 features)	X	X	X
Order quantity	X	X	X
Material (Binary Encoded to 9 features)	X	X	X
Gross weight	X	X	X
BOM-item created (Delta to demand date)	X	X	X
Order created (Delta to demand date)	X	X	X
Order date (Delta to demand date)		X	X
Requested delivery date (Delta to demand date)		X	X
Order method (Binary Encoded to 3 features)		X	X
Supplier-ID (Binary Encoded to 10 features)		X	X
Confirmed date (Delta to demand date)			X
Confirmation received (Delta to demand date)			X

components and suppliers considered One-Hot-Encoding of only these categorical features would have increased the number of dimensions by 26,997 resulting in a high dimensional sparse matrix. To avoid potential memory and computability concerns for our prediction models, and as we wanted to avoid a limitation of our scope, we followed the recommendation of Seger [55] and performed Binary-Encoding instead of One-Hot-Encoding resulting in an increase in the number of dimensions by only 25 without limiting the scope. Nevertheless, we still analysed the impact of a limitation in the scope on the model quality within our modelling phase (see Section 3.4).

Finally, after further data pre-processing operations such as discretization and normalization (see, for details, [56,57]), we finalized the data set for our prediction models. In total, we transformed the initial 17 data fields into 45 features through data pre-processing (see Table 2). As one of our data mining targets was to analyse the trade-off between the prediction quality and the time of the prediction in terms of the three times 'purchasing request', 'order placement' and 'delivery confirmation' we assigned the 45 features to the different times of the prediction. After defining the input features for the different prediction models the Data Preparation was completed.

3.4. Modelling

After understanding the available data and defining the features of our ML models, we set up an experimental plan (see Table 3) to give a quantified response to the objectives of our study – answering the research questions. In this phase of the CRISP-DM procedure, we only define the experimental plan to answer the research question. The execution of the experimental plan – the training and evaluation of the ML models – is considered in the results phase.

The experimental plan consisted of two stages. The first stage focused on the trade-off between the time of the prediction and the model quality utilizing different ML algorithms. Simultaneously, as we set up regression models in this first stage, we can also evaluate if regressions algorithms are capable to predict delivery delays. Thus, in the first stage of the experimental plan, we can answer RQ1 and RQ2. The second stage is about analysing the impact of a limitation in the scope in terms of a limitation in the product portfolio and supplier structure on the model quality and thus, is designed to answer RQ3.

In detail, in the first stage, we plan to compare different regression models for each of the three times of the prediction 'purchasing request', 'order placement', and 'delivery confirmation' in the experimental plan. The models in each of the different times of the prediction differentiate in the ML algorithm used. In detail, we compared the performance of a Linear regressor (LR), a Support Vector regressor (SVR), a Decision Tree (DT) regressor, a Random Forest (RF) regressor, an Adaptive Boosting (AB) regressor, a Gradient Boosting (GB) regressor and a Multilayer Perceptron (MLP). All of them are standard ML algorithms typically used in ML applications, but with a widespread across different learning techniques. RQ1 is primarily concerned with

Table 3
Overview of the experimental plan.

ML algorithm	Stage 1 – Time of the prediction			Stage 2 – Limitation in the scope		
	1. Purchasing request	2. Order placement	3. Delivery confirmation	Five or more orders	Two or more order	No limitation
LR	X	X	X	Selection of the best three performing ML algorithms in stage 1		Already covered with stage 1
SVR	X	X	X			
DT	X	X	X			
RF	X	X	X			
AB	X	X	X			
GB	X	X	X			
MLP	X	X	X			
Total No. of models	21			6		

evaluating the feasibility in general of whether regression algorithms are suitable for predicting delivery delays. RQ2 focuses on the comparison of the performance of prediction models at different times of the prediction. To answer both RQs, it is therefore not necessary to maximize the performance of the ML model by sophisticated ML algorithms; simple and established ML algorithms are sufficient. Instead, it is necessary to ensure comparability across the three times the prediction. Thus, in all three considered times of prediction, the same set of ML algorithms is to be applied. Thus, in the first stage, to analyse the trade-off between the time of the prediction and the model quality, in total 21 prediction models were included in the experimental plan. Depending on the time of the prediction, the set of input features to be used for training and evaluation of the prediction model varied as defined in Table 2.

Further, the focus of the second stage of the experimental plan was to analyse the impact of a limitation in the scope of the prediction model in terms of a limitation in the product portfolio and supplier structure on the model quality. Here, we planned to compare the model quality of prediction models with different levels of limitations. In detail, we planned to compare the following three limitations:

- Limitation to *five or more* orders per component and supplier, following Brintrup et al. [9].
- Limitation to *two or more* orders per component and supplier.
- *No limitation*, which was applied in the 21 prediction models of the first step.

When limiting the product portfolio and supplier structure the number of samples varies, but the data structure itself remains equal. Further, in the second step, we were interested in the impact of the limitation and not in the impact of the ML algorithms nor the time of the prediction. Thus, we only planned to apply the limitations in scope to three selected prediction models of the first step of the experiment – the three best-performing models considering the time of the prediction and the model quality. Thus, in the second step, we included six more prediction models with different levels of limitation in the experimental plan – only six more models instead of nine since the level *No limitation* was already covered in the first step.

Consequently, in total, considering both stages of our experiment 27 prediction models were included in the experimental plan to quantify the impact of the time of the prediction, the impact of the ML algorithm, and the impact of a limitation in the product portfolio and supplier structure on the model quality. All prediction models were implemented in Python 3.7 utilizing the Scikit-learn library.

3.5. Evaluation and results

In the evaluation phase, we trained and compared the performance of all prediction models following the two steps of our experimental plan. In both steps, we evaluated the reached model qualities utilizing MAE, RMSE, and R^2 as defined in the business understanding phase. Further, we split the data sets of all prediction models of the two steps into two separate train and test data sets with a ratio of 80% for training and 20% for testing the models utilizing the train_test_split-function in sklearn with a fixed random_state. The train and test data sets were identical for each model to ensure comparability.

Subsequently, we trained the models of the first step based on the train data set and optimized the hyperparameters. For tuning the hyperparameters we used a grid search algorithm. An overview of the optimized hyperparameters used in each of the prediction models is given in the appendix in Table 6. After the training, we evaluated the achieved model qualities based on the test data set. The results of the first step are documented in Table 4. For the following, we defined a model as a trained ML algorithm at a specific time of the prediction. Additionally, for ease of reading, we used the simplified notation $\text{model}_{\text{ML algorithm, time of the prediction}}$. As an example, the model with the AB regressor as ML algorithm at the time of the prediction 'order placement' is notated as $\text{model}_{\text{AB, 1}}$.

Comparing the metrics revealed that with an R^2 of 92% at the time of the prediction '1. purchasing request', and 98% at the times of the prediction '2. order placement' and '3. delivery confirmation' the AB regressor performed best compared to other ML algorithms, closely followed by the RF regressor and GB regressor. Further, the performance of all ML algorithms increased from the time of the prediction '1. purchasing request' to '2. order placement' and remained almost equal from '2. order placement' to '3. delivery confirmation'. Thus, the information available after placing the order have an impact on the performance of the prediction model, whereas the information on the delivery confirmation has almost no impact. Consequently, the best performing model was $\text{model}_{\text{AB,2}}$ with information available after order placement. However, with an MAE of 8.2 days, an RSME of 19.9 days, and an R^2 of 92% the $\text{model}_{\text{AB,1}}$ with information available after the purchasing request performed only slightly worse. Considering the trade-off between the time of the prediction and model quality, the possibility of taking actions earlier with a slightly lower model quality was preferable over a higher model quality in our case study. Thus, the $\text{model}_{\text{AB,1}}$ was our model of choice. Further, the reach model qualities already showed, that a prediction of delivery delays in calendar days using a regression approach is possible. Thus, with these results we can already answer the research questions RQ1 and RQ2:

- *Answer to RQ1:* Regression algorithms are capable to predict delivery delays of orders placed at suppliers of a low volume high variety manufacturer. Specifically, the prediction has an MAE of 8.2–3.2 days depending on the selected time of the prediction.
- *Answer to RQ2:* With a later time of the prediction and thus the available one of more information about the procurement process the prediction quality increases – which is trivial. Related to the MAE the prediction model improves by approx. 60% (an improvement from 3.2 to 8.2 days) when comparing the creation of a purchasing request and the date of delivery confirmation by the supplier. Nevertheless, the model at the early time of prediction shows good prediction results.

For a practical application of the prediction model in the company under consideration, we suggest applying two models: First, a prediction after the purchasing request with a slightly lower model quality, followed by a second prediction after placing the order with a higher model quality. With this combination, the personnel in the purchasing department are supported with an early but less reliable risk assessment of potentially delayed deliveries which is updated after placing the order. Thus, we ensured both, a possibility to predict early and high model quality.

Further, we analysed the feature importance of our model of choice (see Fig. 6) revealing those input features that significantly influence the model's output. The results showed that the features 'Order created' and 'BOM item created' are most important in the $\text{model}_{\text{AB,1}}$. At the time of the prediction – in this case, at the time of the creation of an internal purchasing request – the time of the demand creation of a component and the time of the transformation of the demand into a purchasing request are therefore decisive for the prediction of the delivery delay. Consequently, it can also be deduced that current delivery delays are largely determined by delays in demand creation and the transformation of the demand into a purchase order. For the practical application of our model at the machinery manufacturer, this indicated that the processes upstream of the purchasing process needed to be accelerated. Thus, based on the feature importance plot, we were able to identify the most relevant features for the prediction of delivery delays and to deduct general optimization potential at the machinery manufacturer in the processes upstream of the purchasing process.

After identifying $\text{model}_{\text{AB, 1}}$ as the model of choice followed by $\text{model}_{\text{RF, 1}}$ and $\text{model}_{\text{GB, 1}}$, we continued with the second step of our

Table 4
Reached model qualities in the first step of the experiment.

ML algorithm	1. Purchasing Request			2. Order Placement			3. Delivery Confirmation		
	MAE	RSME	R^2	MAE	RSME	R^2	MAE	RSME	R^2
LR	29.5	43.7	59%	19.0	27.8	83%	19.3	28.2	83%
SVR	29.7	43.9	58%	19.1	27.8	83%	19.3	28.3	82%
DT	16.2	43.8	81%	7.1	14.3	96%	6.9	14.6	95%
RF	10.2	20.6	91%	4.6	10.7	98%	4.2	11.0	98%
AB	8.2	19.9	92%	3.2	9.9	98%	3.2	10.7	98%
GB	10.9	20.1	91%	4.8	9.9	98%	4.6	10.6	98%
MLP	25.2	35.9	72%	9.4	14.8	95%	9.2	15.0	95%

Table 5
Reached model qualities in the second step of the experiment.

ML algorithm	No limitation			Number of orders per component and supplier ≥ 2			Number of orders per component and supplier ≥ 5		
	MAE	RSME	R^2	MAE	RSME	R^2	MAE	RSME	R^2
RF	10.2	20.6	91%	10.1	21.1	90%	10.1	21.1	90%
AB	8.2	19.9	92%	8.1	19.4	92%	8.3	20.0	92%
GB	10.9	20.1	91%	10.9	20.4	91%	10.2	19.8	91%

experimental plan. Here, we applied the same training and evaluation procedure as in the first step – training and tuning the model based on the train data set and evaluating the model performance based on the test data set. Results are documented in Table 5. It was particularly noticeable that there is no increase nor decrease in the model quality when limiting the product portfolio or suppliers. Thus, in our case study, contrary to Brintrup et al. [9], a limitation was not necessary. Consequently, with these results we can answer RQ3 as follows:

Answer to RQ3: In our exemplary case at the low volume high variety machinery manufacturer the curse of dimensionality was not an issue when setting up a regression model predicting delivery delays of orders placed at suppliers.

Subsequently, we further analysed our model of choice, model$_{AB,1}$, to gain a better understanding of its functionality. Here we conducted a sensitivity analysis of the dataset size (see Fig. 5) quantifying the relationship between dataset size and model performance. In detail, we varied the dataset size from 25% to 100% in 25% steps of the 119,610 orders and evaluated the R^2-value. The results showed that the model quality was only slightly improving with a larger dataset size within the analysed range. Thus, a further increase in the dataset size might improve the model quality. Contrary, a reduction of the dataset size would also be acceptable since the model quality is reduced only minimally with a smaller dataset size. In our case study, the 119,610 orders were placed in three years. Thus, the manufacturer under consideration is placing approx. 40,000 orders per year. Since a dataset size of 25% has provided almost equally good results, our model would also be applicable for a volume of 10,000 orders per year. Consequently, our model might also be applicable to manufacturers that place smaller numbers of orders.

Overall, we were able to fulfil the objectives of our case study. In the first step, we confirmed that a regression model is capable to predict delivery delays and confirmed the influence of time of the prediction on the model performance. Further, we were able to select an AB regressor at the earliest time of the prediction 'purchasing request' as the model of choice, and thus, made a trade-off between the model quality and the time of the prediction. Thus, we could answer RQ1 with our first contribution that an ML-based regression model can predict delivery delays of orders placed at suppliers of machinery manufacturing in calendar days with an R^2 of 92% in our model of choice. In addition, we could answer RQ2 with our second contribution that an early prediction within the purchasing process based on information available after creating the internal order request is possible and only slightly worse than a later prediction. Subsequently, we analysed the impact of a limitation of our data set on the model quality. Here, we could answer RQ3 with our third contribution that in our case study, there is no need to limit the dimensionality of our prediction model.

With our regression model, the company under consideration now can predict delivery delays immediately after creating a purchasing request and getting information about its severity. This helps to select adequate reactive measures in terms of costs for the measures and the

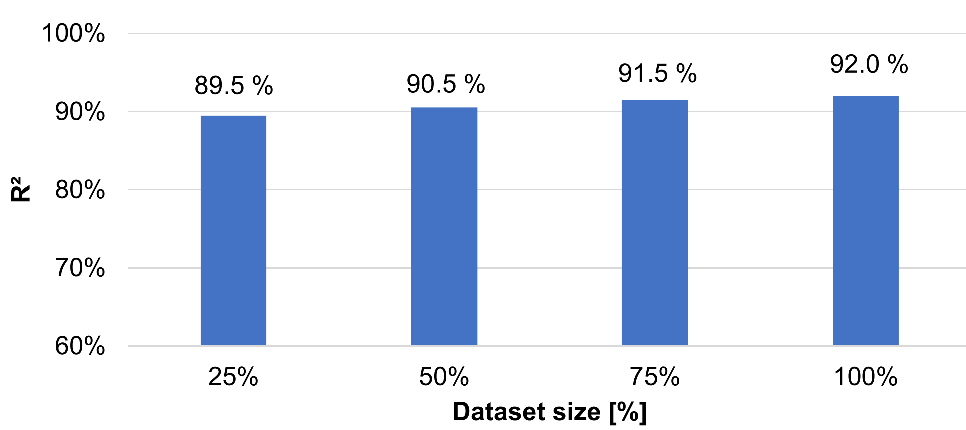

Fig. 5. Impact of the dataset size on the model quality.

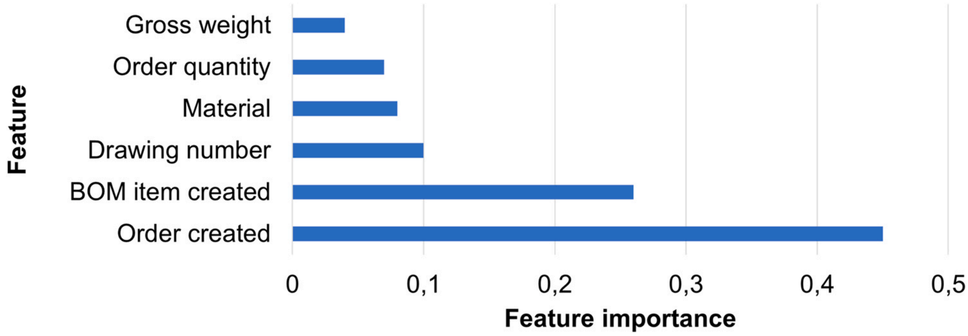

Fig. 6. Feature importance of the model$_{AB,1}$.

predicted severities. In addition, companies can also predict the period a component will be delivered early. Thus, companies can allocate space for inventory and calculate days of inventory turnover predictably. Furthermore, since no restriction in dimensionality was necessary, the model is applicable to all components and suppliers considered.

4. Discussion

In our research, we have considered one machinery manufacturer as an exemplary case study. Although a case-based approach is common in the field of machine learning (see, for example, [58–62]) the findings might remain case-specific and might not be generalizable to other sectors. Nevertheless, our study confirmed the findings of previous case studies of Brintrup et al. [9] and Baryannis et al. [10] – a prediction of potentially delayed deliveries of suppliers is possible. However, it should be noted that all three available case studies including ours focus on similar industries in terms of product complexity and quantity. In other industries, such as automotive, the product complexity and the number of components ordered may be different, potentially affecting the achievable model quality. Thus, further research should validate the achieved findings in other case studies and focus on different industries as well.

Further, we were able to prove that a limitation in the product portfolio and the supplier structure is not necessary. In our case, a limitation did not affect the model quality – neither positively nor negatively. Thus, further researchers can limit their scope if they want to reduce complexity but are not forced to do so by the curse of dimensionality if they use appropriate encoders for high-dimensional categorical features such as a binary encoder.

Moreover, our model quality with an R^2 of 92% at the time of the prediction ‘order request’ and 98% at the time ‘order placement’ is still afflicted with inaccuracy. Here, we identified the ML algorithm to be an important factor for model performance. Accordingly, the use of other regression ML algorithms could be a way to further improve the model quality. Future research should therefore investigate other regression ML algorithms for the prediction of potential delivery delays. Moreover, including additional information about the purchasing process in the ML model could further increase the model quality. For example, we considered so far only a little information about the required components – product number, weight, and material. Further information about the purchased components, such as their complexity or details from the drawing or CAD data, could improve the model performance. The trade-off between the effort in gathering additional data and improvements in the prediction model should also be further investigated by future research. In addition to data about the purchased component, information about the supplier’s manufacturing process, such as the planned work schedule or available capacity, could add value to the prediction model as well. Future research could therefore try to set up an interface to the suppliers and thus include direct information from the supplier’s manufacturing process in the prediction model to further increase its quality.

Further, the dataset of the case study originates from the ERP system of the considered machinery manufacturer. Here, the data fields of a purchasing order are overwritten with every update and there is no history of the recorded data. For example, a supplier can report an update of his expected delivery date, but the dataset in the system always contains only the last one reported. Consequently, it is possible that the dataset considered at the time of the prediction would not have been identical to the dataset that is available retrospectively. Therefore, future research should investigate whether overwriting the data set in ERP systems has an impact on the predictive performance of delivery delay prediction models. For this purpose, it is conceivable to record a data set with historical changes.

Moreover, a critical limitation for the practical applicability of the model is that the model in its current form only uses knowledge from the past to predict the future. However, if previously unknown events occur that have a significant impact on the supply chains (e.g., Covid, Ukraine war), the model will not take these influences into account and will make incorrect predictions. Therefore, with the current setting of the model, manual control is still necessary as a supplement to the model. Here, the integration of additional data sources such as the evaluation of the daily worldwide news would be beneficial.

5. Conclusion

Complex manufacturing industries that rely on externally sourced components need to ensure the on-time start of assembly processes, as delayed deliveries can cause costly assembly disruptions. To streamline operations that depend on external supply, the use of Machine Learning for the prediction of supply delays has been proposed in recent research with promising results. However, extant research considered classification approaches that depict whether a delay will occur but omitted its timing and duration. Furthermore, data-intensive approaches limit their application to high-volume settings whereas low volume high variety industries could also benefit from delay prediction. Finally, we posit that for a delay prediction to be meaningful, it needs to be early enough in the procurement cycle, such that mitigative actions can be taken.

In this work, we address these gaps through a systemic procedure setting up defined research questions and answering these with a two stages experimental plan containing a set of different ML regression models. Here, we show that regression algorithms are capable to predict delivery delays of suppliers of low volume high variety manufacturers. Additionally, we show that the severity of a delay can be predicted, early enough for facilitating action. Our approach also has the advantage of mitigating the curse of dimensionality, thereby making it applicable to low volume high variety settings.

For the development of the ML-based regression models, we followed the established Cross Industry Standard Process for Data Mining (CRISP-DM). Here, first, a detailed understanding of the company's procurement process was built and relevant features for our model were identified by performing a correlation analysis. Subsequently, we set up an extensive experimental plan to identify the best ML model. Our experimental plan consisted of two steps. The first step focused on the performance of different ML algorithms such as AB, RF and GB regressors at different times of the prediction to analyse whether a prediction in the early phases of the procurement process is possible. The second step included comparing different approaches to handle high-dimensional input features within a regression model. Executing the experimental plan revealed that an AB-regressor with an R^2 of 92% trained on information available after the creation of an internal order request performed best, meaning that delay prediction can indeed be performed at the point of an order request.

Nevertheless, our model has scope for improvement. For example, the inclusion of further information such as component identifiers or supplier's manufacturing processes may further improve the prediction quality. The models have been tested on a single use case from the manufacturing industry. Further tests in low-volume-high-variety settings would increase confidence in the validity of our approach. Further, the model in its current form only uses knowledge from the past to predict the future. However, if previously unknown events occur that have a significant impact on the supply chains (e.g., Covid, Ukraine war), the model will not take these influences into account and will make incorrect predictions.

From a managerial standpoint, we recommend the use of regression models in the purchasing process, as it enables an early reaction to delivery delays even before an order is placed. Thus, companies can select adequate reactive measures in terms of cost and predicted severity. This can have a positive effect on the adherence to the start of assembly and ultimately on the adherence to the delivery date to the customer. In addition, companies can also predict the time a component will be delivered early. Thus, companies can allocate space for inventory and calculate days of inventory turnover predictably. Consequently, the regression model also can be used to reduce inventory costs.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The research leading to these results received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Appendix

See Table 6.

Table 6
Hyperparameters of the prediction models.

ML algorithm	Hyperparameter	Time of the prediction		
		1. Purchasing Request	2. Order Placement	3. Delivery Confirmation
SVR	C	1	1	1
	dual	True	True	True
	loss	squared epsilon insensitive	squared epsilon insensitive	squared epsilon insensitive
	Max iter	1000	1200	1200
DT	cc alpha	0.0028	0.0030	0.0032
	max features	auto	auto	auto
	max depth	25	25	25
	min samples split	4	4	4
	min samples leaf	5	5	5
RF	bootstrap	false	false	False
	max features	sqrt	sqrt	sqrt
	n estimators	200	220	230
AB	base estimator	decision tree (depth 3)	decision tree (depth 3)	decision tree (depth 3)
	n estimators	200	240	230
	learning rate	0.20	0.18	0.17
GB	n estimators	200	250	270
	max features	sqrt	sqrt	sqrt
	max depth	14	14	14
	min samples split	14	14	14
	min samples leaf	2	2	2
	learning rate	0.10	0.12	0.13
MLP	alpha	1,00E-06	1,00E-06	1,00E-06
	hidden layer size	35	37	34
	max iter	1000	1250	1150
	activation	relu	relu	relu
	solver	adam	adam	adam
	batch size	auto	Auto	auto
	learning rate	0.10	0.11	0.10

The hyperparameters used in the prediction models were optimized utilizing a grid search and cross-validation algorithms (GridSearchCV) from Scikit learn. Table 6 summarizes the utilized hyperparameters in the different prediction models.

References

[1] G. Reinhart, R. Cuiper, and M. Loferer, "Die Bedeutung der Montage als letztes Glied inder Auftragsabwicklung," in *TCW-Report*, vol. 6, *Montage-Management: Lösungen zum Montieren am Standort Deutschland*, G. Reinhart, Ed., München: TCW, 1998, pp. 7–11.

[2] H. Lödding, Handbook of manufacturing control: Fundamentals, description, configuration, Springer, Heidelberg, 2013.

[3] G. Wang, A. Gunasekaran, E.W. Ngai, T. Papadopoulos, Big data analytics in logistics and supply chain management: certain investigations for research and applications, Int. J. Prod. Econ. vol. 176, (2016) 98–110, https://doi.org/10.1016/j.ijpe.2016.03.014

[4] T. Nguyen, L. Zhou, V. Spiegler, P. Ieromonachou, Y. Lin, Big data analytics in supply chain management: a state-of-the-art literature review, Comput. Oper. Res. vol. 98, (2018) 254–264, https://doi.org/10.1016/j.cor.2017.07.004
[5] A. Brintrup, Artificial Intelligence in the Supply Chain, in: T.Y. Choi, J.J. Li, D.S. Rogers, T. Schoenherr, S.M. Wagner (Eds.), The Oxford handbook of supply chain management, Oxford University Press, Oxford, 2020.
[6] Y. Sheffi, The resilient enterprise: Overcoming vulnerability for competitive advantage, MIT Press,, Cambridge, Mass., London, 2007.
[7] K. Katsaliaki, P. Galetsi, S. Kumar, Supply chain disruptions and resilience: a major review and future research agenda, Ann. Oper. Res. (2021) 1–38, https://doi.org/10.1007/s10479-020-03912-1
[8] F.F. Rad, et al., Industry 4.0 and supply chain performance: a systematic literature review of the benefits, challenges, and critical success factors of 11 core technologies, Ind. Mark. Manag. vol. 105, (2022) 268–293, https://doi.org/10.1016/j.indmarman.2022.06.009
[9] A. Brintrup, et al., Supply chain data analytics for predicting supplier disruptions: a case study in complex asset manufacturing, Int. J. Prod. Res. vol. 58, (11) (2020) 3330–3341, https://doi.org/10.1080/00207543.2019.1685705
[10] G. Baryannis, S. Dani, G. Antoniou, Predicting supply chain risks using machine learning: the trade-off between performance and interpretability, Future Gener. Comput. Syst. vol. 101, (2019) 993–1004, https://doi.org/10.1016/j.future.2019.07.059
[11] R.K. Yin, Case study research and applications: Design and methods, Sage publications, 2017.
[12] D.M. McCutcheon, J.R. Meredith, Conducting case study research in operations management, J. Oper. Manag. vol. 11, (3) (1993) 239–256, https://doi.org/10.1016/0272-6963(93)90002-7
[13] C. Shearer, The CRISP-DM model: the new blueprint for data mining, J. data Warehous. vol. 5, (4) (2000) 13–22.
[14] R. Wirth and J. Hipp, “CRISP-DM: Towards a standard process model for data mining,” Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining, no. 1, 2000.
[15] M. Christopher, Logistics & supply chain management. Harlow, England, New York: Pearson Education, 2016.
[16] M. Christopher and H. Peck, Building the resilient supply chain, *0957–4093*, 2004, doi: 10.1108/09574090410700275.
[17] G. Baryannis, S. Validi, S. Dani, G. Antoniou, Supply chain risk management and artificial intelligence: state of the art and future research directions, Int. J. Prod. Res. vol. 57, (7) (2019) 2179–2202, https://doi.org/10.1080/00207543.2018.1530476
[18] M.A. Waller, S.E. Fawcett, Data science, predictive analytics, and big data: a revolution that will transform supply chain design and management, J. Bus. Logist. vol. 34, (2) (2013) 77–84, https://doi.org/10.1111/jbl.12010
[19] J.K. Bae, J. Kim, Product development with data mining techniques: a case on design of digital camera, Expert Syst. Appl. vol. 38, (8) (2011) 9274–9280, https://doi.org/10.1016/j.eswa.2011.01.030
[20] D.M.-H. Chiang, C.-P. Lin, M.-C. Chen, The adaptive approach for storage assignment by mining data of warehouse management system for distribution centres, Enterp. Inf. Syst. vol. 5, (2) (2011) 219–234, https://doi.org/10.1080/17517575.2010.537784
[21] J. Cui, F. Liu, J. Hu, D. Janssens, G. Wets, M. Cools, Identifying mismatch between urban travel demand and transport network services using GPS data: a case study in the fast growing Chinese city of Harbin, Neurocomputing vol. 181, (14) (2016) 4–18, https://doi.org/10.1016/j.neucom.2015.08.100
[22] C. Ghedini Ralha, C.V. Sarmento Silva, A multi-agent data mining system for cartel detection in Brazilian government procurement, Expert Syst. Appl. vol. 39, (14) (2012) 11642–11656, https://doi.org/10.1016/j.eswa.2012.04.037
[23] C.-F. Chien, A.C. Diaz, Y.-B. Lan, A data mining approach for analyzing semiconductor MES and FDC data to enhance overall usage effectiveness (OUE), IJCIS vol. 7, (Supplement 2) (2014) 52, https://doi.org/10.1080/18756891.2014.947114
[24] J. Wang, J. Zhang, Big data analytics for forecasting cycle time in semiconductor wafer fabrication system, Int. J. Prod. Res. vol. 54, (23) (2016) 7231–7244, https://doi.org/10.1080/00207543.2016.1174789
[25] C. Wang, X. Li, X. Zhou, A. Wang, N. Nedjah, Soft computing in big data intelligent transportation systems, Appl. Soft Comput. vol. 38, (7209) (2016) 1099–1108, https://doi.org/10.1016/j.asoc.2015.06.006
[26] P. Helo, Y. Hao, Cloud manufacturing system for sheet metal processing, 6-8, Prod. Plan. Control vol. 28, (2017) 524–537, https://doi.org/10.1080/09537287.2017.1309714
[27] A. Kumar, R. Shankar, A. Choudhary, L.S. Thakur, A big data MapReduce framework for fault diagnosis in cloud-based manufacturing, Int. J. Prod. Res. vol. 54, (23) (2016) 7060–7073, https://doi.org/10.1080/00207543.2016.1153166
[28] F. Bienhaus, A. Haddud, Procurement 4.0: factors influencing the digitisation of procurement and supply chains, BPMJ vol. 24, (4) (2018) 965–984, https://doi.org/10.1108/BPMJ-06-2017-0139
[29] D. Ivanov, C.S. Tang, A. Dolgui, D. Battini, A. Das, Researchers' perspectives on Industry 4.0: multi-disciplinary analysis and opportunities for operations management, Int. J. Prod. Res. vol. 59, (7) (2021) 2055–2078, https://doi.org/10.1080/00207543.2020.1798035
[30] M.M. Queiroz, D. Ivanov, A. Dolgui, S. Fosso Wamba, Impacts of epidemic outbreaks on supply chains: mapping a research agenda amid the COVID-19 pandemic through a structured literature review, Ann. Oper. Res. (2020) 1–38, https://doi.org/10.1007/s10479-020-03685-7
[31] A. McAfee, E. Brynjolfsson, T.H. Davenport, D.J. Patil, D. Barton, Big data: the management revolution, Harv. Bus. Rev. vol. 90, (10) (2012) 60–68.
[32] M.M. Queiroz, S. Fosso Wamba, M.C. Machado, R. Telles, Smart production systems drivers for business process management improvement, BPMJ vol. 26, (5) (2020) 1075–1092, https://doi.org/10.1108/BPMJ-03-2019-0134
[33] X. Zhu, A. Ninh, H. Zhao, Z. Liu, Demand forecasting with supply-chain information and machine learning: evidence in the pharmaceutical industry, Prod. Oper. Manag vol. 30, (9) (2021) 3231–3252, https://doi.org/10.1111/poms.13426
[34] P. Liu, A. Hendalianpour, J. Razmi, M.S. Sangari, A solution algorithm for integrated production-inventory-routing of perishable goods with transshipment and uncertain demand, Complex Intell. Syst. vol. 7, (3) (2021) 1349–1365, https://doi.org/10.1007/s40747-020-00264-y
[35] S.K. Sardar, B. Sarkar, B. Kim, Integrating machine learning, radio frequency identification, and consignment policy for reducing unreliability in smart supply chain management, Processes vol. 9, (2) (2021) 247, https://doi.org/10.3390/pr9020247
[36] A. Hendalianpour, Optimal lot-size and price of perishable goods: a novel game-theoretic model using double interval grey numbers, Comput. Ind. Eng. vol. 149, (1) (2020) 106780, https://doi.org/10.1016/j.cie.2020.106780
[37] P. Liu, A. Hendalianpour, M. Hamzehlou, Pricing model of two-echelon supply chain for substitutable products based on double-interval grey-numbers, IFS vol. 40, (5) (2021) 8939–8961, https://doi.org/10.3233/JIFS-201206
[38] A. Hendalianpour, M. Hamzehlou, M.R. Feylizadeh, N. Xie, M.H. Shakerizadeh, Coordination and competition in two-echelon supply chain using grey revenue-sharing contracts, GS vol. 11, (4) (2021) 681–706, https://doi.org/10.1108/GS-04-2020-0056
[39] K. Nayal, R.D. Raut, M.M. Queiroz, V.S. Yadav, and B.E. Narkhede, Are artificial intelligence and machine learning suitable to tackle the COVID-19 impacts? An agriculture supply chain perspective, *0957–4093*, ahead-of-print, ahead-of-print, p. 438, 2021, doi: 10.1108/IJLM-01–2021-0002.
[40] I.N. Pujawan, A.U. Bah, Supply chains under COVID-19 disruptions: literature review and research agenda, Supply Chain Forum.: Int. J. vol. 23, (1) (2022) 81–95, https://doi.org/10.1080/16258312.2021.1932568
[41] M.O. Alabi, O. Ngwenyama, Food security and disruptions of the global food supply chains during COVID-19: building smarter food supply chains for post COVID-19 era, BFJ vol. 125, (1) (2023) 167–185, https://doi.org/10.1108/BFJ-03-2021-0333
[42] M.M. Bassiouni, R.K. Chakrabortty, O.K. Hussain, H.F. Rahman, Advanced deep learning approaches to predict supply chain risks under COVID-19 restrictions, Expert Syst. Appl. vol. 211, (2023) 118604, https://doi.org/10.1016/j.eswa.2022.118604
[43] P. Ghadimi, C. Wang, M.K. Lim, C. Heavey, Intelligent sustainable supplier selection using multi-agent technology: theory and application for Industry 4.0 supply chains, Comput. Ind. Eng. vol. 127, (5) (2019) 588–600, https://doi.org/10.1016/j.cie.2018.10.050
[44] M.T. Ahmad, M. Firouz, S. Mondal, Robust supplier-selection and order-allocation in two-echelon supply networks: a parametric tolerance design approach, Comput. Ind. Eng. vol. 171, (1) (2022) 108394, https://doi.org/10.1016/j.cie.2022.108394
[45] M.M. Khan, I. Bashar, G.M. Minhaj, A.I. Wasi, N.U.I. Hossain, Resilient and sustainable supplier selection: an integration of SCOR 4.0 and machine learning approach, Sustain. Resilient Infrastruct. (2023) 1–17, https://doi.org/10.1080/23789689.2023.2165782
[46] Z. Du Ni, Xiao, M.K. Lim, A systematic review of the research trends of machine learning in supply chain management, Int. J. Mach. Learn. Cyber vol. 11, (7) (2020) 1463–1482, https://doi.org/10.1007/s13042-019-01050-0
[47] J. Wagner, P. Burggräf, M. Dannapfel, C. Fölling, Assembly disruptions. Empirical evidence in the manufacturing industry of Germany, Austria and Switzerland, Int. Referee J. Eng. Sci. vol. 6, (3) (2017) 15–25.
[48] K. Jensen, Cross-industry standard process for data mining. [Online]. Available: https:// en.wikipedia.org / w/ index.php ? title = Cross-industry_standard_process_for_data_mining& oldid = 1117200778 (accessed: Jan. 30 2023).
[49] R.J. Hyndman, A.B. Koehler, Another look at measures of forecast accuracy, Int. J. Forecast. vol. 22, (4) (2006) 679–688, https://doi.org/10.1016/j.ijforecast.2006.03.001
[50] G.G. Judge, R.C. Hill, W.E. Griffiths, H. Lütkepohl, T.-C. Lee, Introduction to the theory and practice of econometrics, 2nd ed., Wiley, New York, 1988.
[51] R. Bellman, *Dynamic programming.* Princeton, NJ: Princeton Univ. Pr, 1984.
[52] M. Kuhn, K. Johnson, Applied predictive modeling, 5th ed., Springer, New York, 2016.
[53] H. Akoglu, User's guide to correlation coefficients, Turk. J. Emerg. Med. vol. 18, (3) (2018) 91–93, https://doi.org/10.1016/j.tjem.2018.08.001
[54] V. Verdhan, Supervised Learning with Python: Concepts and Practical Implementation Using Python, 1st ed. Berkeley CA: Apress; Imprint: Apress, 2020.
[55] C. Seger, “An investigation of categorical variable encoding techniques in machine learning: binary versus one-hot and feature hashing,” School of electrical engineering and computer science, Royal Institute of Technology, Stockholm, Schweden, 2018.
[56] J. Han, M. Kamber, and J. Pei, Data mining: Concepts and techniques, 3rd ed. Amsterdam: Elsevier/Morgan Kaufmann, 2012.
[57] G. Dong and H. Liu, Feature Engineering for Machine Learning and Data Analytics. Milton: Chapman and Hall/CRC, 2018.
[58] I. Tirkel, Forecasting flow time in semiconductor manufacturing using knowledge discovery in databases, Int. J. Prod. Res. vol. 51, (18) (2013) 5536–5548, https://doi.org/10.1080/00207543.2013.787168
[59] A.D. Karaoglan, O. Karademir, Flow time and product cost estimation by using an artificial neural network (ANN): a case study for transformer orders, Eng. Econ. vol. 62, (3) (2017) 272–292, https://doi.org/10.1080/0013791X.2016.1185808
[60] L. Lingitz, et al., Lead time prediction using machine learning algorithms: a case study by a semiconductor manufacturer, Procedia CIRP vol. 72, (2018) 1051–1056, https://doi.org/10.1016/j.procir.2018.03.148
[61] S. Singh and U. Soni, “Predicting Order Lead Time for Just in Time production system using various Machine Learning Algorithms: A Case Study,” in *2019 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, Noida, India, Jan. 2019, pp. 422–425.
[62] P. Burggräf, J. Wagner, B. Heinbach, F. Steinberg, Machine learning-based prediction of missing components for assembly – a case study at an engineer-to-order manufacturer, IEEE Access vol. 9, (2021) 105926–105938, https://doi.org/10.1109/ACCESS.2021.3075620