Case study on delivery time determination using a machine learning approach in small batch production companies

Alexander Rokoss[1] · Marius Syberg[2] · Laura Tomidei[3] · Christian Hülsing[1] · Jochen Deuse[2,3] · Matthias Schmidt[1]

Abstract

Delivery times represent a key factor influencing the competitive advantage, as manufacturing companies strive for timely and reliable deliveries. As companies face multiple challenges involved with meeting established delivery dates, research on the accurate estimation of delivery dates has been source of interest for decades. In recent years, the use of machine learning techniques in the field of production planning and control has unlocked new opportunities, in both academia and industry practice. In fact, with the increased availability of data across various levels of manufacturing companies, machine learning techniques offer the opportunity to gain valuable and accurate insights about production processes. However, machine learning-based approaches for the prediction of delivery dates have not received sufficient attention. Thus, this study aims to investigate the ability of machine learning to predict delivery dates early in the ordering process, and what type of information is required to obtain accurate predictions. Based on the data provided by two separate manufacturing companies, this paper presents a machine learning-based approach for predicting delivery times as soon as a request for an offer is received considering the desired customer delivery date as a feature.

Keywords Delivery time · Machine learning · Production planning and control · Artificial intelligence · Case study

Introduction

Customer expectations regarding logistic performance have strongly increased in the last decades. Nowadays, customers expect not only high quality and individual products for low prices but also short and especially reliable delivery times (Paprocka & Cyba, 2015). This results in major challenges for manufacturing companies. On one hand, although Make-to-Stock (MTS) approaches allow fast delivery times, they might not be beneficial for all types of manufacturing companies due to a large number of products or product variants. On the other hand, Make-to-Order (MTO) manufacturers are able to address more customized needs but they often tend to use standard delivery times (Rao et al., 2005). The latter approach leads to a strong fluctuation of capacity utilization and requires high effort in continuously adjusting the manufacturing capacity. In addition to that, customers' requirements can vary greatly. While in some cases the delivery should just be as fast as possible, others demand a specific time window for the delivery or even a just-in-time delivery. The resulting pressure on manufacturing companies to meet communicated delivery dates while offering competitive delivery times pushes them to find reliable approaches for accurate predictions. Often, a reliable delivery date can only be determined by converting a customer order into internal manufacturing orders that are scheduled within the manufacturing execution system. However, since converting and planning manufacturing processes is time-consuming, companies face the choice of delaying the confirmation of the customer order or adding significant buffers to ensure on-time delivery for the communicated delivery date.

✉ Alexander Rokoss
  alexander.rokoss@leuphana.de

1   Institute for Production Technology and Systems, Leuphana University of Lüneburg, Lüneburg, Germany

2   Institute of Production Systems, Technical University Dortmund, Dortmund, Germany

3   Centre for Advanced Manufacturing, University of Technology Sydney, Sydney, Australia

For these reasons, the estimation of delivery dates has been addressed by the literature for decades. However, the majority of contributions are based on numerous assumptions and strongly condense the underlying information from various orders. In recent years, research has been increasingly focusing on the application of machine learning approaches for production planning and control tasks (Panzer & Gronau, 2023; Waubert de Puiseau et al., 2022). Nevertheless, the estimation of delivery dates often only appears as a sidenote while the main objective of such contributions is to optimize a system regarding tasks like order acceptance (Zhang et al., 2021), order release planning (Schneckenreither et al., 2021), or sequencing (Liang et al., 2012). Existing research in manufacturing analytics has shown a notable gap in its coverage, particularly regarding the application of Quantitative Logistic Models (QLM) in extracting valuable insights from raw manufacturing data. To address this shortcoming, our study explores the use of QLM to improve feature extraction from manufacturing data and then validates the improvements achieved through practical applications in two different cases. In addition, current research in manufacturing analytics focuses mainly on the prediction of manufacturing lead times, while neglecting the consideration of desired delivery dates, which may be a crucial aspect. Our results prove the importance of this integration, showing that without considering desired delivery dates, the accuracy of delivery date forecasts is no better than that achieved by conventional methods such as moving averages or mean delivery times. The central role of desired delivery dates in improving forecast accuracy is carefully assessed through the application of Shapley values, shedding light on their profound impact in this area. Through the strategic use of established QLM, we strive to extract valuable features based on domain knowledge. This approach enables domain specialists to understand the underlying factors of predictions. As a result, it has the potential to enhance both the trust and clarity of AI, which addresses a significant challenge in its integration into decision-making processes (Adadi & Berrada, 2018; Golpayegani et al., 2023).

Therefore, the object of this study is to evaluate the use of machine learning in predicting delivery dates through a combined case study from two job shop manufacturers. For the case studies, we use two German companies that operate as contract manufacturers producing small batches from 1 to 10,000 pieces. One manufacturer produces precision mechanics utilizing 17 different machines. The other case deals with a manufacturer of rubber sealings utilizing 52 machines covering 15 different manufacturing technologies.

The contribution of this study can be summarized as follows:

– We prove that machine learning is capable of assisting in predicting delivery times at various stages of the order processing workflow, which leads to sustainable competitive advantage for companies. We especially use the desired delivery date as a feature. This has not been taken into account in predicting the delivery date yet.
– We evaluate how sets of input parameters available at different stages of the process contribute to the prediction accuracy.
– We adapt the CRISP-DM methodology in this paper to focus on the intricate task of delivery time forecasting. In doing so, we extend the scope of the conventional CRISP-DM framework, introducing elements such as the utilization of QLM.
– We demonstrate that machine learning is able to estimate delivery dates *before* confirming incoming orders, which reduces manual efforts throughout the process.
– From a practical perspective, we show that QLM combined with machine learning can significantly improve the prediction of the time of order confirmation. Compared to the previous method in one Use Case, it shows that the forecast error, measured by the Normalised Root Mean Square Error (NRMSE), can be reduced by about two-thirds.
– We show that the selection of the best features and their quantitative importance (measured by SHAP Values) is central to forecast quality. It is not only about finding the best algorithm but also about identifying and quantifying those features that give the best results. Based on the use cases we present, we are able to highlight the essential role of considering the desired delivery date for the accurate prediction of delivery dates. This finding is useful for all companies with small batch production, as it can be applied to their forecasting processes.

Our study is structured as follows. After a literature review that places our contribution in the context of current research, we outline the methodology used to utilize domain knowledge in the feature extraction process. Then, the case studies are described and the application of the steps of the methodology are documented in the main case. The results are compared with the findings of the secondary case before the most relevant input factors are discussed within the context of existing domain knowledge. Results indicate that the extraction of features using domain-specific knowledge significantly improves the quality of the prediction and the explainability of the overall results. In the presented main case, the proposed machine learning-based approach estimates delivery dates with an additional error of 2 business days compared to the estimation of the process planning department. The machine learning approach does not use data that can only be used after process planning has been conducted on the customer order. Therefore, the machine learning-based approach can be used significantly earlier in the process, for example within the offer process or before confirmation of the customer order.

Literature review

The scientific community provides different definitions for delivery times depending on the domain. Within the industrial context, Wiendahl (1997) defines the 'delivery time' as the period of time between the placing of the order by the customer and the receipt of the goods, the 'delivery date'. This definition stems from a logistics-related context in multi-link supply chains. Accordingly, Chapman et al. (2017) define delivery times as the sum of transport time, picking, provision, and order processing. Looking at the delivery time from the production point of view, the definition slightly differs. Following the Hanoverian Supply Chain Model (Schmidt & Schäfers, 2017), delivery times are the result of throughput times of the orders through the customer-order specific processes plus possible transport times. This implies, that the delivery time depends on the stocking strategy (Lödding, 2013). Thus, the delivery time in Engineer-To-Order (ETO) includes different time components than in Make-To-Stock (MTS) approaches. In addition, the delivery time is influenced by various variables. Among other factors, are the sequence-dependent setup times, the share of different product groups in the product mix, the batch size, the order lead time, and thus also the delivery time of a product in discrete manufacturing (Schuh et al., 2019). In ETO environments, the time components for the design and engineering are dominant, other components are procurement or supplier lead times, and lead times for production, including manufacturing, assembly, and testing (Alfnes et al., 2021).

The prediction of delivery times in the highly variable and complex production environment is necessary (Bhalla et al., 2023). Accurate prediction of delivery times is a good foundation for negotiations (Rau et al., 2006), as it allows companies to better plan and allocate resources, and make informed decisions based on a high number of influencing variables and gives, therefore, an advantage in the market (Amaro et al., 1999; Bhalla et al., 2023; Cannas et al., 2020; Grabenstetter & Usher, 2014; Hicks & Braiden, 2000). In the literature, delivery time estimation has been more of a byproduct of models to improve logistics performance, including order release planning or sequencing. Next to estimating delivery dates prior to order confirmation, assessing the feasibility of meeting requested delivery dates is highly important (Bhalla et al., 2023). As we move towards data-driven decision-making, various models have been developed for determining delivery time as well as different individual components such as due dates or lead times in different domains (Bezirgiannidis et al., 2013; Choetkiertikul et al., 2017; Jodlbauer & Tripathi, 2023). The literature describes deterministic, stochastic, and combined methods. In the industrial environment, mathematical calculations, heuristics, and meta-heuristics were first described decades ago (Adam et al., 1993; Ragatz et al., 1984). Thürer et al. (2012)

and Moses et al. (2004) propose simulation-based heuristics for estimating delivery dates, known as simulation-based due date setting and incremental forward simulation based on an approach of Roman and Del Vallei (1996). Even today, these approaches are still being used and further research is being carried out on them (Bhalla et al., 2023).

In recent years, the use of machine learning methods to support decision-making in manufacturing has been growing, as they provide the opportunity to make accurate predictions based on large amounts of data. In particular, the use of such methods to solve production planning and control (PPC) tasks is gaining increasing interest. More recently, there has been a sharp increase in the number of publications on the use of machine learning in the estimation of components of the delivery times as well, especially throughput times (Maier et al., 2022). This can be explained by the fact that there are various variables influencing the total delivery time. Especially in this environment with different data and influencing factors, the use of machine learning helps to identify correlations in the data (Shet et al., 2022). This can be particularly valuable in the manufacturing environment to analyze the dependencies of delivery time components in detail.

The majority of existing methodologies do not utilize QLM for feature extraction in delivery time forecasting. Much of the existing literature mainly focuses on modeling techniques and neglects the crucial step of feature extraction. A more in-depth discussion of the primary influencing factors with the aim of generalization is notably observed in the works of Öztürk et al. (2006) and Bender and Ovtcharova (2021). The authors employ distinct (simulation) models to assess the impacts of various manifestations of influencing factors on forecasting accuracy.

A review of the literature in this field reveals that various authors have examined delivery times or some of their key components. In our research, we considered references that deal directly with the assignment of time components, as well as literature in which a prediction of any time components is implied as part of the optimization of a system. Table 1 classifies methods for determining delivery times as well as the type of input data. Based on this, it is possible to draw several conclusions regarding the use of machine learning for predicting delivery times in a practical setting. Firstly, many of the existing publications only utilize simulation data, which may not accurately reflect real-world scenarios (Alenezi et al., 2008; Bender & Ovtcharova, 2021; Murphy et al., 2019; Öztürk et al., 2006). Using simulation-based approaches, Adam et al. (1993) and Thürer et al. (2012) nevertheless demonstrated that it is possible to predict production lead times in complex systems with multi-level product structures, but only a few exceptions specifically focus on delivery time rather than just the order lead time (Alnahhal et al., 2021; Khiari & Olaverri-Monreal, 2020). Second, most contributions continue to use, at least for comparison, classical

Table 1 Overview of data-based determination of delivery time and its components in complex production systems, based on Rokoss et al. (2021) and Bhalla et al. (2023)

| | Data Source | | Used Methods | | | | | | | | | |
| | | | ML-based | | | | | Non-ML-based | | | | |
	Real Data	Simulation Data	Neural Network	Support Vector Machine	Random Forest	Decision Tree	Gradient Boosting	Linear Regression	Other Methods	Identification of features	Use of quantitative logistical models	Predicting Delivery Dates directly
Adam 1993	○	●	○	○	○	○	○	○	●	○	○	◐
Öztürk 2006	○	●	○	○	○	●	○	●	●	●	◐	◐
Alenezi 2008	○	●	●	●	○	○	○	○	●	○	○	◐
Nguyen 2016	●	●	○	○	○	○	●	○	●	○	○	○
Polim 2017	●	○	○	○	●	○	◐	○	●	◐	○	○
Yang 2017	○	●	●	○	○	○	○	○	●	◐	○	○
Hao 2018	●	●	●	○	○	○	○	○	●	◐	○	○
Yang 2018	○	●	●	○	○	○	○	○	●	◐	◐	◐
Murphy 2019	○	●	●	○	●	●	●	○	●	◐	○	○
Khiari 2020	●	○	○	○	●	○	●	●	●	◐	○	●
Alnahhal 2021	●	○	○	○	●	●	●	●	●	◐	○	●
Araujo 2021	●	○	●	○	●	○	○	○	●	◐	○	●
Bender 2021	○	●	●	●	○	○	●	○	○	◐	○	○
Burggräf 2021	●	○	●	●	●	●	●	●	○	◐	○	○
Schneckenreither 2021	○	●	●	○	○	○	○	○	○	○	◐	◐
Lorenzo-Espejo 2022	●	○	○	○	●	○	●	○	○	◐	○	◐
Mohsen 2023	●	○	○	○	●	●	●	○	○	◐	○	○

data-driven methods such as linear regression, heuristics, or other mathematical models (Nguyen, 2016; Polim et al., 2017). Thirdly, publications that use more complex algorithms such as Neural Networks also typically use simulation, which may lack practical validation (Schneckenreither et al., 2021; Yang & Zhang, 2018; Yang et al., 2017). The considered research, applying Neural Networks on real data, deals with the delivery of packages and is therefore assigned to a rather foreign domain (Araujo & Etemad, 2021). Finally, many publications with real-data validation use or compare tree-based models, whose use is often based on easier comprehensibility (Alnahhal et al., 2021; Khiari & Olaverri-Monreal, 2020; Mohsen et al., 2023).

Overall, existing literature maintains a narrow focus on the determination of specific time components such as throughput time, instead of looking at the prediction of the delivery time. A holistic approach to determining delivery times requires understanding the relevance of production management theory-related features, as well as increasing the interpretability of the results with existing domain knowledge. A structured approach to applying the supervised learning models to real production data is needed. This involves several steps, including the need for clear business understanding to include the relevant features on the delivery time, detailed data pre-processing using specific models for feature extraction, as well as suitable modeling to select

the best supervised learning models should be aimed for. This paper intends to consider these aspects from the literature review while shedding light on the development and application of machine learning techniques for delivery time prediction.

Methodology

This paper uses a deductive approach to contribute to the explanatory research on delivery time determination. The methodology used to predict delivery times follows an applied research approach and is based on real world primary data. Using exemplary use cases is an established approach in the field of machine learning, as it allows to gain insights with practical relevance (McCutcheon & Meredith, 1993; Steinberg et al., 2022). Thus, a machine learning-based approach for predicting delivery times has been developed using data and information from two separate manufacturing companies producing in small batches. In contrast to mass producers, small-batch producers have an earlier decoupling point from customer orders. The earlier decoupling point allows for more customization of products to meet the product variability needs of the market, but at the same time missing finished goods buffers lead to longer delivery times (Olhager, 2003). Small batch manufacturers generally have high production

uncertainty due to various complexity and dynamic characteristics like production volumes, product mix, and design changes (Birkie & Trucco, 2016; Braglia et al., 2019; Powell et al., 2014). These characteristics cause variability in the individual components of the delivery time, compared to a mass producer that stores finished goods in stock. It increases the difficulty of the prediction. Analyzing those dynamics is a main challenge to manage their impact on the companies' targets (Richter et al., 2023). External factors like unstable incoming purchases or customer request dates act as a catalyst for internal fluctuations because they lead to rush orders and non-linear scheduling, resulting for example in sub-optimal capacity utilization. As these aspects affect delivery times, the approach presented in this study considers the characteristics of small batch production.

The overall aim is to answer the following research questions:

1. Can delivery times be predicted earlier in the process in the same quality the process planning department provides?
2. Which domain knowledge-specific features do improve the quality of the prediction?
3. What are the different quantitative effects of the different features on the training model?
4. Do the domain knowledge-specific features help a domain expert to follow up on the decisions made?

The methodology used for the development of the approach follows the widespread, industry-independent Cross Industry Standard Process for Data Mining (CRISP-DM) (Chapman et al., 2000). We use domain knowledge by applying QLM described in the Hanovarian Supply Chain Model (Schmidt & Schäfers, 2017) on manufacturing systems to identify relevant input factors regarding the manufacturing process. The underlying funnel model for capacity utilization (Bechte, 1988) and the throughput diagram (Wiendahl & Tönshoff, 1988) represent the main influence on the feature extraction process. Overall, the methodology of this paper aims to specify the very broad CRISP-DM for the task of delivery time forecasting, by taking into consideration desired delivery dates and QLM.

CRISP-DM follows six key phases using a cyclic approach, including business understanding, data understanding, data preparation, modeling, evaluation, and deployment (Chapman et al., 2000). The deployment phase is strongly company-dependent and therefore not part of this paper. Although CRISP-DM itself is well documented, it only provides a very broad structure to a wide variety of data mining projects. To address the requirements of delivery time prediction, the specific methods utilized to adapt the CRISP-DM for forecasting delivery times are discussed below. Figure 1 gives a schematic overview of the specific methods that are used in the respective CRISP-DM phase.

Business and data understanding

The initial phase, business understanding, requires gaining knowledge about requirements and needs from a business perspective. From a value stream perspective, it is possible to identify the key stages of the ordering process, from the first customer inquiry to the product delivery. From an operational perspective, it is possible to understand the manufacturing process and the sequence of operations required to manufacture the goods, both automated and manually. Finally, interviews with stakeholders allow to unveil domain knowledge and in particular what are the key factors influencing delivery times according to the stakeholders' business experience, thus providing guidance for future modelling. To address the specific characteristics of small batch manufacturers, the relevance of external manufacturing processes needs to be examined using value stream mapping and/or interviews.

Key factors of the data understanding phase for delivery time prediction include data on products, customers, and suppliers. Process-wise data on ordering processes, manufacturing processes, and overall market data need to be analyzed. In terms of products, it is possible to identify the structure of the product portfolio, the volume, complexity, manufacturing steps, and whether products follow a Make-to-Stock (MTS) or a Make-to-Order (MTO) approach. Customer data shows the number of orders, order frequency, pricing, and relevant customer groups. Similarly, supplier data highlights the presence of possible stable partnerships, as well as how semi-finished goods and components are delivered (i.e. based on the customer order or in stock), and in what time frame. In general, exploring the available data needs to be split into data that is available prior to the processing by the process planning department and data that is available after processing. To apply QLM during feature engineering it is crucial to extract relevant timestamps from the ERP systems. Timestamps marking the start and finish of every machining step internally are necessary, as well as timestamps on external machining and procurement. The overall focus of the data collection must be to reconstruct the full internal and external processing of a customer order, covering every step from the first offer to the handover of the finished goods to the shipping company.

Data preparation and feature extraction

The data preparation phase involves all activities required to prepare the final dataset before it is fed into any model. In workshop environments, the key manufacturing activities are measured manually by operators, who observe processes
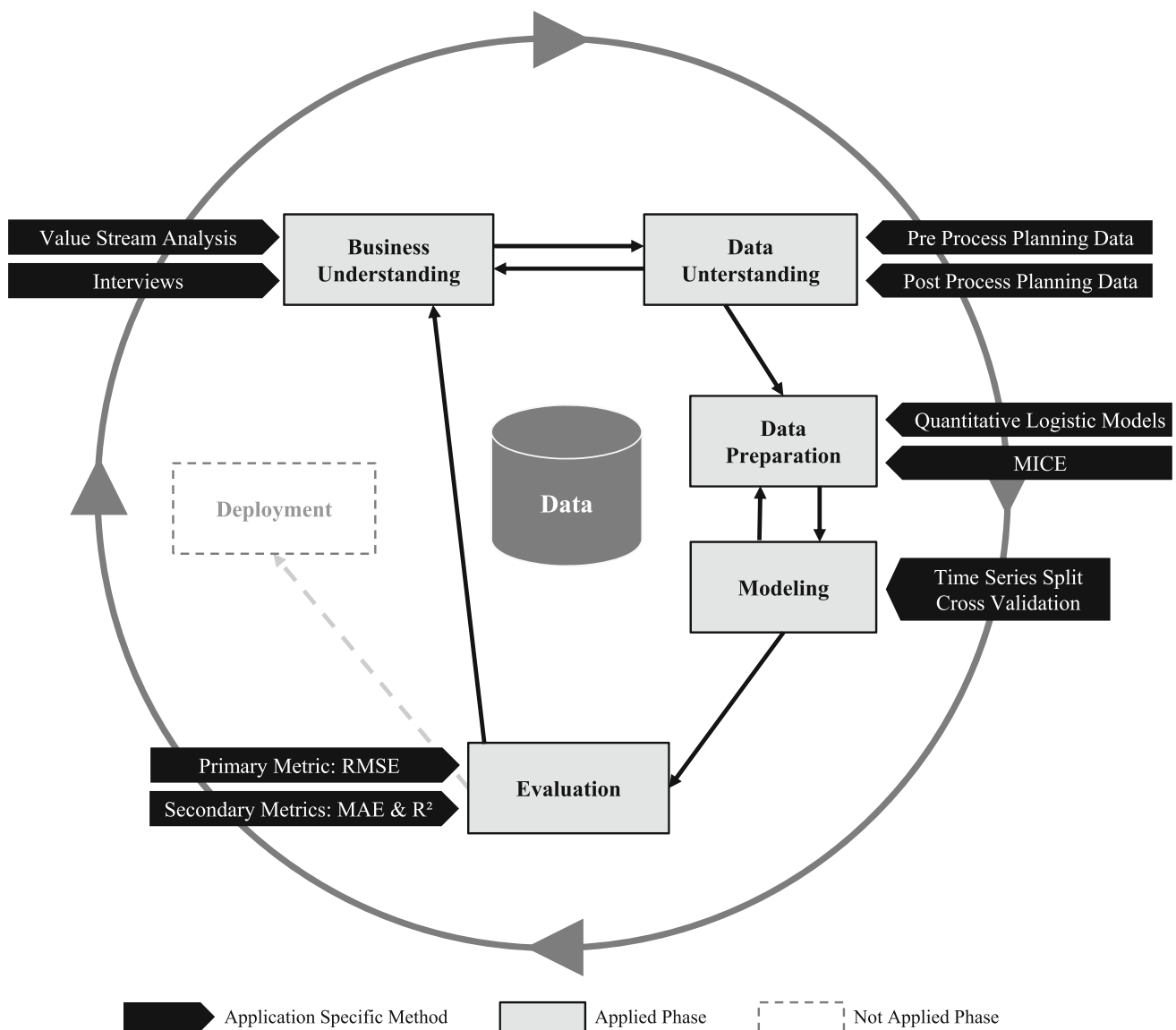
Fig. 1 CRISP-DM scheme (Chapman et al., 2000) enlarged with application-specific methods for delivery time prediction

and record timestamps. This manual process ultimately often leads to gaps in the collected data. The use of Multivariate Imputation by Chain Equations (MICE) approach for handling missing values is an appropriate approach to improve the overall data quality before proceeding to feature engineering. Feature engineering is supported by domain knowledge regarding production planning in extracting features from raw data that best describe the manufacturing, procurement, and shipping processes. The Hanovarian Supply Chain Model offers different approaches to describe the logistic performance of a manufacturing system. Due to the lack of finished goods stock inventory in MTO workshop manufacturing, stock inventory models as well as service level models are not applied to the datasets. Instead, the feature extraction focuses on internal and external manufacturing processes as

well as customer and supplier ratings. Internal manufacturing processes are analyzed by applying the funnel model for capacity utilization (Bechte, 1988) and the throughput diagram (Wiendahl & Tönshoff, 1988) on the existing data. The combination of Work-in-Process (WIP) and the resulting range of the systems, allows to identify internal manufacturing bottlenecks for every order. WIP is defined as the workload that is waiting to be processed by a system. WIP can be measured using a time scale (e.g. manufacturing minutes) or using a quantitative scale (e.g. number of orders). For further analysis, the average output of a system needs to be calculated. The average output of a system is measured by the sum of the output over all orders divided by the time it took the system to finish the orders. The range of the system x is calculated by dividing WIP by the average output over the

past n manufacturing orders. The range measures the time a system needs to finish all waiting orders (Kettner & Bechte, 1981).

$$range_{(n,x)} = \frac{WIP_{(x)}}{\frac{1}{n} \sum_{i=1}^{n} output_{(x)_i}} \tag{1}$$

The range can also be applied to the whole factory to get an idea of the overall waiting time for new orders. The range of the production system is calculated by dividing the WIP of waiting customer orders by the average quantity of dispatched orders. With the range for every system available, every order can be checked for their respective bottleneck system by finding the maximum value of the ranges for the systems k the specific order needs to pass.

$$range_{bottleneck(n,k)} = \max_{x \in k} \{range_{(n,x)}\} \tag{2}$$

External manufacturing processes are covered in the feature set with the average delivery time $\overline{DT}$ and the average desired delivery time deviation $\overline{\Delta DT}$ over the past m orders placed.

$$\overline{DT}_m = \frac{1}{m} \sum_{i=1}^{m} DT \tag{3}$$

$$\overline{\Delta DT}_m = \frac{1}{m} \sum_{i=1}^{m} (DT - desiredDT) \tag{4}$$

The presented features can be combined with an ABC classification or ranking of the suppliers. The same procedure is applicable to the customer side, offering information on key account customers that might result in prioritization of their orders to maintain a high long term delivery time reliability. An approach towards a ranking system is to generate rank by applying a sortation to customers and suppliers by the relative quantity Q of orders that have been placed by customer/supplier within a certain time frame in the past (e.g. three years) compared to all placed orders m.

$$Q_{(z)} = \frac{1}{m} \sum order_{customer/supplier} \tag{5}$$

It is important to note that the presented approaches follow the basic steps of feature engineering, and they should be considered as a baseline for further development. Every production system has its unique mechanisms and resulting data, which needs to be processed accordingly.

Modeling

After the data has been adequately prepared, modeling techniques within the field of machine learning can be applied to predict delivery times and extract relevant input factors. This requires selecting the most performant model out of various proven supervised learning algorithms. In this study, we utilize Support Vector Machine (SVM), Artificial Neural Networks (ANN), Gradient Boosting Tree (XGB), Decision Tree (DT), Random Forests (RF), and Linear Regression (LR). To determine the influence of domain-specific knowledge on feature generation, models are trained with different feature sets. The modeling involves tuning relevant parameters of the models by applying grid search, reducing the number of features by backward feature elimination, and cross-validating the model. Since the second case only consists of 2632 orders, for consistency both cases utilize a 90–10 train-test split to improve the overall prediction quality compared to the more common 80–20 split. This is achieved by increasing the size of the training and validation data set (Xu & Goodacre, 2018) to 90%. The increased risk of overfitting that could be caused by the smaller test data set is reduced by cross validating the prediction results (Hawkins et al., 2003; Vabalas et al., 2019). When validating predicted delivery times, it is important to take into consideration the chronological sequence of the orders. Performing a k-fold cross validation, often used in classification problems, would expose the dataset to the "look-ahead bias". This bias occurs when future data is used during training, thus leading to better results in the training process than later in a production environment. To avoid introducing future knowledge for predictions, a Time Series Split Cross Validation approach is applied. This method allows to avoid the look-ahead bias while using enough data to perform effective cross validation (see Fig. 2). While the training set increases in size, the test set maintains a consistent size and includes observations that happen after the ones in the training set.

Evaluation

The evaluation phase allows to assess the results provided by each model and compare the results. After the data is trained on 90% of the dataset in the modeling phase, the model is tested on the remaining 10% of the data. The process is repeated for various algorithms and feature sets, and the performance is evaluated accordingly. The quality of the predicted delivery times is measured as the deviation between the actual delivery date and the predicted delivery date in business days. To put the measurements into context, the results are compared to the prediction deviation of a) the delivery dates estimated by the process planning department and b) estimated delivery dates if the average delivery time is used. To analyze the performance of the different approaches, various metrics are used. One of the most widely adopted metrics is R^2, which is used in regression problems to evaluate the quality of the model on a scale from $-\infty$ to 1.
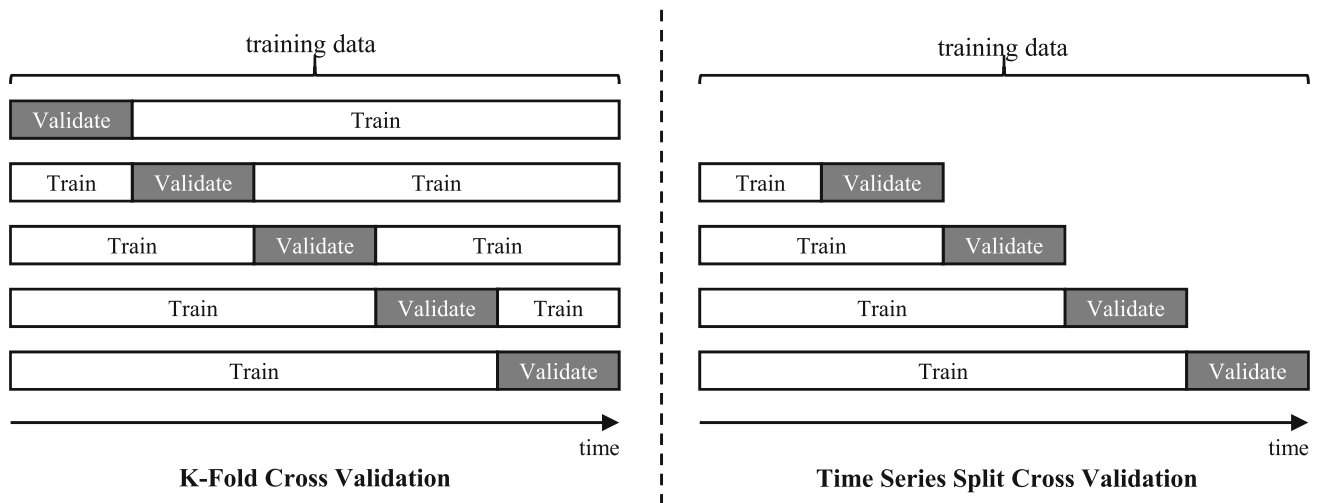
Fig. 2 Time series cross validation vs. K-fold cross validation (adapted from Hyndman and Athanasopoulos (2021))

$$R^2(y, \widehat{y}) = 1 - \frac{\sum (y_i - \widehat{y}_i)^2}{\sum (y_i - \overline{y})^2} \tag{6}$$

Due to its intuitive interpretability, another popular metric is using the mean average error (MAE), which returns the absolute distance of the prediction from the real value. When predicting delivery time, MAE is measured in business days.

$$MAE(y, \widehat{y}) = \sum_{i=1}^{n} \frac{|y_i - \widehat{y}_i|}{n} \tag{7}$$

However, MAE is robust to outliers when there are enough observations close to the real value. For the prediction of delivery times for small batch manufacturers, outliers are extremely important as small deviations can be addressed by adjustments such as rescheduling or increased capacity (e.g. extra shifts), while large differences between the prediction and the real delivery time represent a big issue for the manufacturing company. Therefore, because of its ability to weight outliers more heavily, the Root Mean Squared Error (RMSE) is selected as the preferred metric for the prediction of delivery times.

$$RMSE(y, \widehat{y}) = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \widehat{y}_i)^2}{n}} \tag{8}$$

In addition to the metrics previously mentioned, the normalized RMSE and the normalized MAE are used to allow to comparison of results between different manufacturing systems, as normalization allows to compare datasets with different scales.

$$NRMSE(y, \widehat{y}) = \frac{RMSE(y, \widehat{y})}{\frac{1}{n} \sum |y_i|} \tag{9}$$

$$NMAE(y, \widehat{y}) = \frac{RMSE(y, \widehat{y})}{\frac{1}{n} \sum |y_i|} \tag{10}$$

However, the proposed approach aims to provide an automated method for predicting the delivery date either when the first offer is made or shortly before the incoming customer order gets confirmed. The proposed methodology aims to adapt the broadly applicable CRISP-DM to the task of delivery time forecasting by utilizing desired delivery dates and QLM.

Case studies

The methodology presented above is validated through its application to two case studies from two real world manufacturing companies. Both companies behind the following case studies use workshop production, despite operating in different industries. After a short introduction of both cases, the first case study is described in detail. As the second case study follows the key steps of the main one, only the results are presented. Since the study aims to compare predictions made at different points in time, both case studies follow the experimental design shown in Fig. 3. In the first step (I), manufacturing as well as procurement, article, and customer order data is fetched from different systems within the company. Then, all relevant features are extracted from the existing data (II). The features are split into three groups (A, B, C). Feature group A consists of features that can be directly extracted from the customer order, such as ordered articles or the date of order confirmation. Feature group B consists of features that can be extracted from the underlying data utilizing domain knowledge. The features in group B are extracted
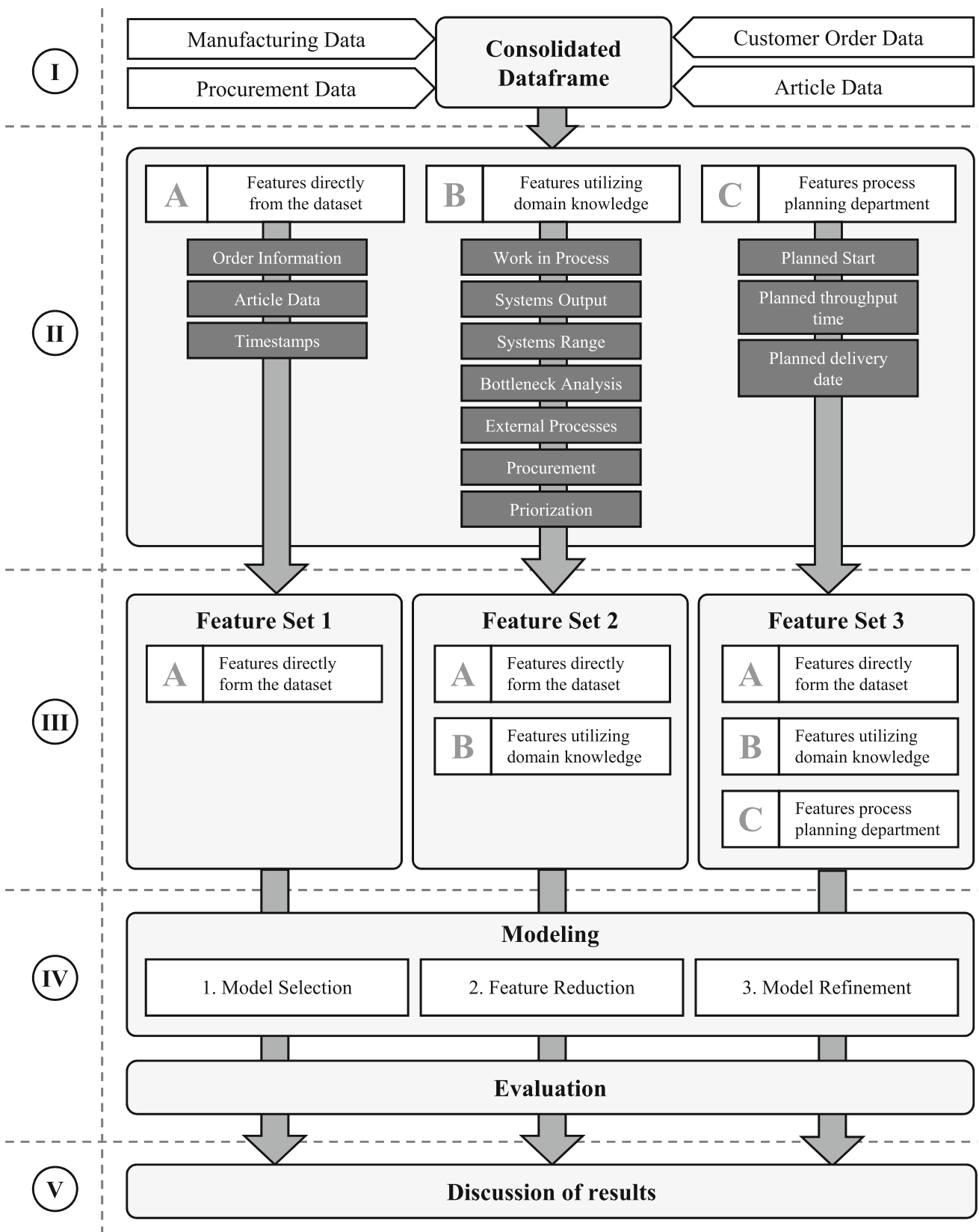
Fig. 3 Experimental design for both examined case studies

from the data by applying QLM as described in the methodology section. Feature group C consists of features that can only be extracted late in the order process when the process planning department has finished processing the order. The features of group C contain the planned start of manufacturing, planned throughput time, and the planned delivery date. In the third step (III), the feature sets for the modeling phase are combined. All three feature sets contain features directly from the dataset (group A). Feature sets 2 and 3 also contain features that derive from domain knowledge (group B). Feature set 3 also contains features that are only available later in the process (group C). The fourth step of the experimental design is the modeling and evaluation phase as described in the methodology section (IV). For both cases, the same model parameters are applied as described. The modeling phase consists of three phases—the model selection, the feature reduction, and the model refinement phase. In the final step, the results of the application of the feature sets are discussed in the context of manufacturing management domain knowledge (V).

Comparison: case study business characteristics

Both examined companies are make-to-order (MTO) manufacturers based in Germany. The manufacturer behind case 1 produces order-specific rubber sealings. The manufacturer behind case 2 produces metallic parts for mechanical use cases. Both companies produce in batches between 1 and 10,000 pieces for a global market. Since the global shipping times vary widely depending on the level of urgency of the shipping and the shipping provider, all delivery dates in both cases are communicated ex works. The two cases were chosen for this study because of the similarity in the underlying business process from the initial offer to the final delivery of the goods. The manufacturing process of the goods usually requires between three to ten machining steps in case 1 and three to seven machining steps in case 2. Although the two examined cases share several similar process patterns, there are some differences. The company behind case 1 does not send unfinished goods out to other manufacturers for external processing such as painting. In case 2, those external processes are part of the manufacturing process and can impact the manufacturing throughput time of the orders. Additionally, the desired delivery date was not specified in the dataset for case 2.

During the offer process, the sales department estimates delivery times based on experience, taking a large buffer time into the calculation. After receiving the customer order, along with CAD (Computer Aided Design) files describing the object details, the order gets confirmed and the details are forwarded to the process planning department. On average, the process planning department is able to estimate a delivery date using manufacturing execution systems and communicate it to the customer within 7 business days for case 1 and four business days for case 2. During this time, the process planning department verifies the CAD files sent by the customer, checks the availability of raw materials, and defines process details related to manufacturing. Decisions on utilized manufacturing methods and machines are finalized and then migrated into software files for production. Finally, the order is placed into the Enterprise Resource Planning (ERP) system and the theoretical delivery date is calculated. Overall, this planning process results in a high risk of the delivery date initially communicated in the order confirmation not being matched by the manufacturing capacities. This Mismatch will first be visible days after the confirmation of the customer order, as shown in Fig. 4. Therefore, the key objective in the given case studies is predicting the delivery date of a customer order based on the information available at the time of the first offer, although a machine learning-based approach can be applied also later in the order processing (i.e. order confirmation and manufacturing order released into the ERP system) (see Fig. 4).

Case study 1

Business and data understanding

Data was fetched from the company's Enterprise Resource Planning (ERP) system including the time frame from 07/23/2019 to 03/31/2022. Table 2 gives a short description of the raw data that has been processed.

The main data set on manufacturing orders contained 16,361 single orders involving 58,344 manufacturing steps. The manufacturer finalizes approximately 300 customer orders on a weekly basis. Orders from 1,630 different customers covered 7,620 different articles to be produced. The full data set contains ten tables counting 73 columns and 2,393,353 rows.

Data preparation and feature extraction

To gain valuable insights, the data collected needs to be cleaned and processed. The main goal of this phase is to identify relevant features and avoid data inconsistencies that might lead to inaccurate results during the modeling phase.

Firstly, the analysis focuses on the entire dataset and the label attribute. Across the entire dataset, duplicate data has been removed, and missing data points were omitted from the dataset using Multiple Imputation by Chained Equations (MICE) and k-nearest-neighbor regression. The label attribute contains the delivery time. For this, the original timestamps have been converted into integers representing the number of business days since the customer order. Removing non-business days allows to eliminate the effect of
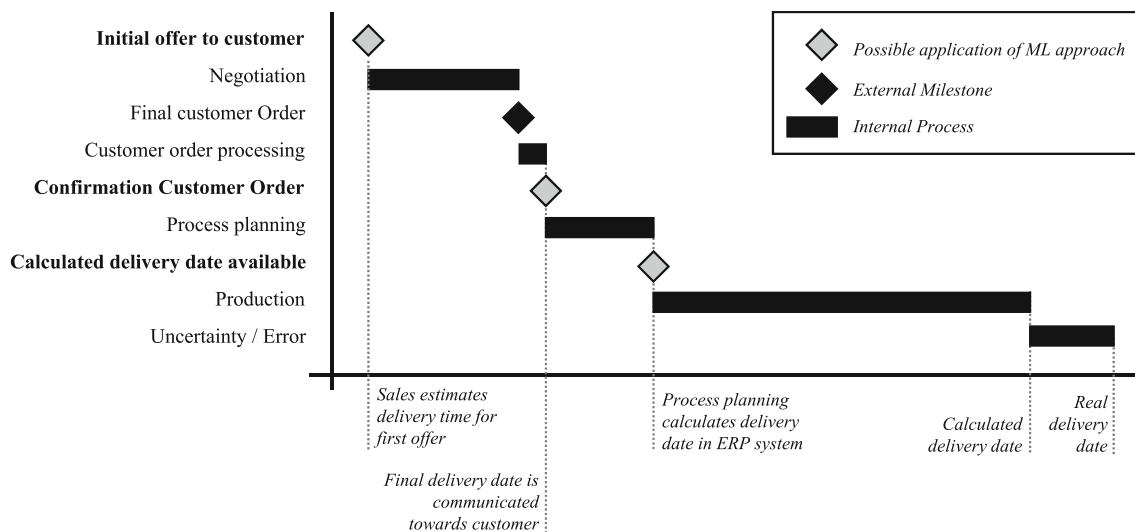
Fig. 4 Customer order processing (both cases)

Table 2 Description of raw data

Category	Data	Included data/description
Customer order	Descriptive data	Quantity ordered, customer ID, timestamps on order, desired delivery date, real delivery date
	Order positions	Article ID, quantity, material, throughput time
Articles	Descriptive data	Name, revision, bill of material ID
	Measurements	Information on shape and measurements (74 different attributes)
	Bill of materials	Quantity of parts needed per article
Procurement	Replenishment lead times	Replenishment lead time per article (overall average, not time specific)
	Historical stock inventory	Stock inventory of every material on a daily basis
Production process	Timestamps operations	operation start, operation end, set up start, set up end
	Descriptive data operations	Sequence of operations, machine number, operator
	Machining areas	Connects machines to machining areas, based on manufacturing technology
	Working time documentation	Registered times of operators being assigned to a machine by opening a manufacturing step on machines operating computer

public holidays (e.g. Christmas) on the manufacturing capacity and delivery times (see Fig. 5). This enables the extraction of clear statistics of the prediction variable. Thus, the average delivery time is around 18 days, and over 99% of deliveries are completed within 50 business days, with only a few outliers over 50 business days (see Fig. 6).

Secondly, features have been examined and processed independently to address their individual characteristics. As a result, ad-hoc steps for data preparation have been followed for customer order-related features, features representing

production planning and control information (domain knowledge), and features representing the planned times available after the order has been processed by the process planning department.

Attributes related to customer orders (feature group A) contain various information including order date, desired delivery date, products, and other customer details. These features are converted to numerical or categorical types while ensuring consistency across the dataset. Among these features, there is also the predicted number of machines to be used in the manufacturing process. The company behind case
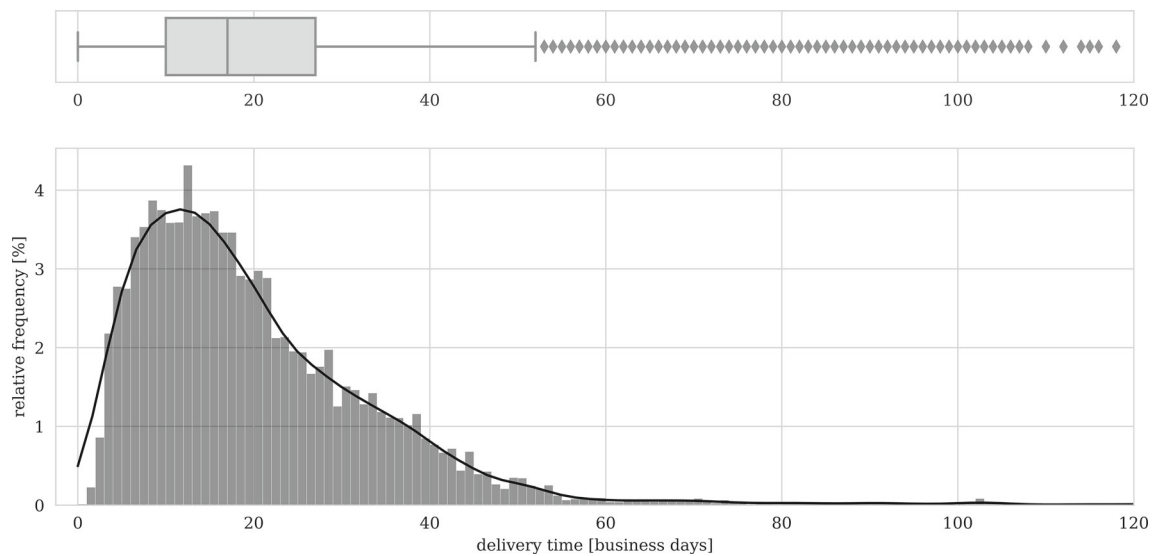
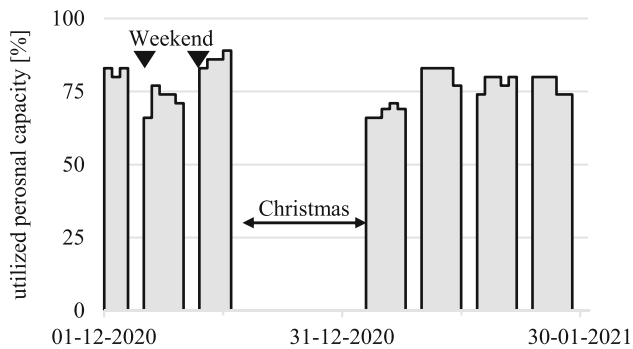**Fig. 6** Distribution of the label attribute (case 1)

Fig. 5 Daily utilization of available work hours (case 1)

1 is organized in job shops, with 13 different departments that group machines of the same type (e.g. milling, drilling, gluing). When the customer first requests an offer, the sales department estimates how often these departments need to be utilized to complete the order.

Additional attributes have been extracted to represent the domain knowledge in production management (feature group B), with the features deriving from QLM. These features aim to provide statistics related to the performance of the manufacturing system, based on historical data. This is due to the influence of the performance of manufacturing systems on delivery times. For example, the presence of bottlenecks, which vary depending on specific product routings, limit the amount of output a manufacturing system can deliver (Goldratt & Cox, 1984).

Finally, after the order has been processed by the process planning department, it is possible to extract the attributes representing the planned times (feature group C), namely the planned throughput time, the planned manufacturing start date (expressed in number of business days since receiving

the customer order), the planned machine set up times and cycle times. Other historical information represents the status of the manufacturing system. This includes the level of work-in-progress (WIP) in the factory and in specific departments, as well as the average number of processed orders in a specific month for both the whole factory and specific departments.

The three feature groups collected represent different inputs to the modeling phase (see Table 3).

Modeling

Modeling and evaluation for the case study are iterated three times with different feature sets. The initial predictive model is developed using two of the three main data groups identified in the previous phase, namely customer order-related data, and domain knowledge (feature set 2). These two types of information are available as soon as an offer is created or a customer order is received, while planned times features become available only later in the process once the process planning department has processed the order. Thus, together these two groups are the input of the baseline scenario. After a model is developed using feature set 2 as input, the model is re-trained changing the input to feature set 1 and feature set 3, thus allowing to compare the contribution of each feature group. The modeling phase consists of the three steps model selection, feature selection, and model refinement.

1. Model selection The dataset is trained using a 90–10 ratio. Various established machine learning algorithms have been evaluated using the metrics previously mentioned. The table below lists the models used in the training phase along with their parameters for the grid search. Training has been conducted using between 18 and 20 different combinations

Table 3 Main data groups and availability in time

Feature group	Features	Availability
Customer order	Desired delivery time in business days, quantity of products ordered, customer class, customer sales rank, product class, product sales rank, novelty of product (has it been ordered/manufactured before), frequency of the product being manufactured in the past, number of materials in the bill of materials (BOM) of the product, maximum replenishment time of the materials in the BOM	Available when the purchase offer is made
Domain knowledge (production planning and control information)	Work in progress of the whole factory, range of the whole factory, average of processed orders per month in the whole factory, work in process of machining areas, average of processed orders machining areas month before, range of machining areas	Historical, always available
Planned times	Planned throughput time (numeric), planned manufacturing start date in business days from customer order, planned machine preparation times, planned machining times	Available after the order has been processed by the process planning department

Table 4 Models and parameter sets for initial grid search

Model	Parameter	Values
Decision tree	max depth	3, 5, 7, 9, 11, 13
	min_samples_leaf	3, 5, 7
Random forrest	n_estimators	50, 100, 250, 500
	max_depth	3, 5, 7, 9
	max_features	log2
XGBoost	objective	reg:squarederror
	colsample_bytree	0.3
	learning_rate	0.1, 0.01
	max_depth	2, 4, 6, 8, 10
	n_estimators	250
Support vector machine	kernel	Rbf
	gamma	10, 1, 0.1, 0.01, 0.001
	C	10, 100, 1000, 10,000
Artificial neural network	activation	relu, tanh
	solver	Adam
	batch_size	Auto
	learning_rate	Adaptive
	max_iterations	200
	hidden_layer_sizes	(1), (20), (100), (150), (200), (50,10), (100,10), (100,50), (200,50), (100,50,20)

Table 5 Results of initial modeling (case 1, feature set 2)

Model	RMSE (days)	MAE (days)	R^2
XGB	7.95	6.09	0.45
ANN	8.19	6.29	0.42
SVM	8.31	6.35	0.44
RF	8.47	6.60	0.46
DT	8.67	6.63	0.38
LR	9.16	6.74	0.37

of parameters (see Table 4) within a separated grid search process for each model. Models have been fitted using Time Series Split Cross Validation.

The results of the initial modeling process are presented in Table 5. As shown by the results in the table, all models perform better than the baseline prediction using the delivery time average (RMSE = 12.1 business days). However, because XGB demonstrates the best performance in RMSE and MAE, further development is conducted using this algorithm.

2. Feature reduction During the data preparation phase, the data is extracted and aggregated from different systems, resulting in 46 features. To reduce the risk of overfitting the model, as well as the amount of data preparation needed in production, only the most relevant features are selected. Thus, features are ordered by importance using Shapley Additive Explanation (SHAP) values. In this specific case, while the number of features for a local optimum would be 41, it is possible to score a close-to-optimal RMSE within 0.1 business days with 13 features (see Fig. 7).

3. Model refinement Once the best model is selected, along with a reduced number of features, the model parameters are tuned again to identify the values leading to the best performance. The final grid search identified the following optimal parameters of the XGB model:

- colsample_bytree: 0.9
- learning_rate: 0.1
- max_depth: 4
- n_estimators: 250

Then, the model is re-trained over the training dataset with 13 features, and prediction values are saved from the application of the model to the test set. In the final step, the metrics are being calculated.

Finally, the whole modeling process is re-tested to compare the relative contribution of the feature groups identified in the data preparation phase.

Evaluation and results

The value of the delivery times prediction can be best discussed by comparing the various applications represented by feature sets. Table 6 shows the performance of the machine learning approach using different feature sets. The baseline application is the simple calculation of the mean delivery time based on historical data. This results in a mean of 12 days RMSE and a negative R^2, indicating a very poor correlation. Additionally, the results are compared to the best-performing approach using moving averages. Moving averages were calculated using different numbers of finished orders (n = 5, 10, 25, 50, 100, 200, 500, 1000, 2000). The prediction using the best-performing moving average (200 orders) does result in comparable results to using the mean delivery time as the
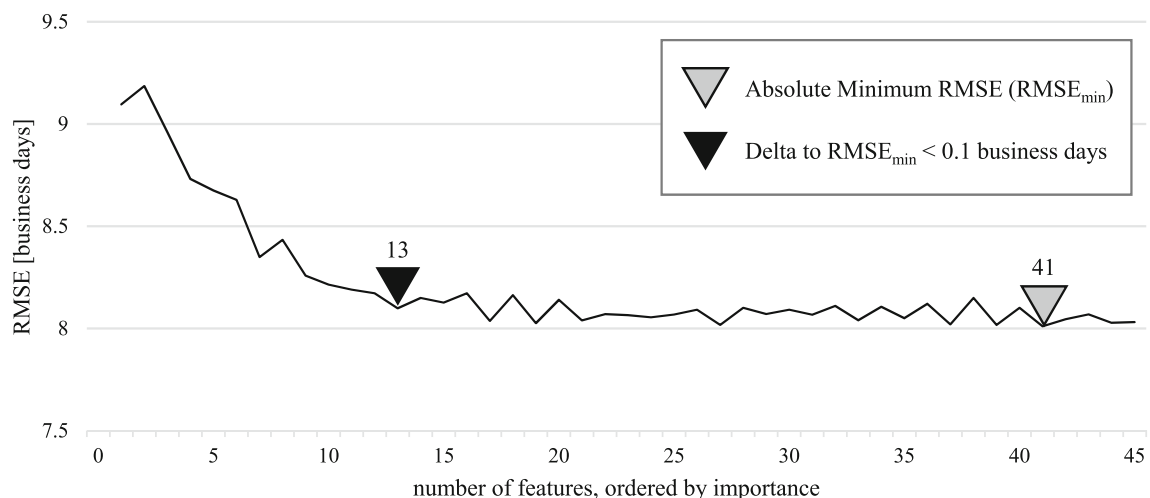
Fig. 7 Feature reduction process by threshold definition

Table 6 Final results for XGB model (case 1)

Approach	RMSE (days)	MAE (days)	R^2	NRMSE	NMAE
Mean delivery time	12.15	10.42	− 0.01	65%	55%
Moving average delivery time (200 orders)	11.81	9.04	− 0.05	63%	53%
Process Planning Department	5.87	3.83	0.76	31%	20%
ML with planned times (C)	5.56	3.98	0.79	30%	21%
ML without planned times (B)	7.73	6.03	0.59	41%	32%
ML without planned times, without domain knowledge features (A)	9.47	7.48	0.39	50%	40%
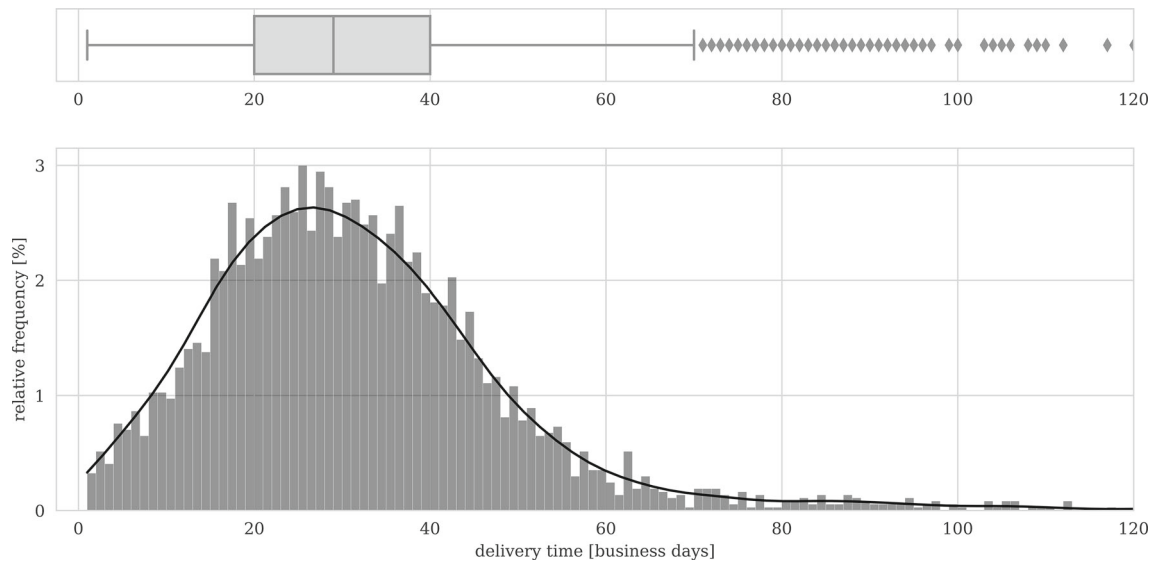
Fig. 8 Distribution of the label attribute (case 2)

prediction. Training a dataset without domain knowledge features improves the prediction by almost 3 business days in RMSE (9.5). This shows that the use of machine learning can help with predicting delivery times, even without utilizing domain knowledge. In this scenario, the prediction error is reduced by almost 50% compared to the estimates produced by the process planning department. However, by training a dataset that includes production management domain knowledge features, the RMSE can be further reduced by 1.8 business days (7.7), reducing the remaining error almost by a further 50%. In other words, including information about the current state of the performance of the manufacturing system allows considerable knowledge gain. Finally, when the planned dates defined during process planning are also included in the dataset, the results are more accurate but only by a small margin. This shows that the planned dates originate mostly from the information derived from the status of the manufacturing system. It is important to note that this last step can be performed only late into the process when the process planning department calculated the planned delivery date.

Case study 2

Business and data understanding

To validate the method presented in the previous section, a second case study was examined. Since the majority of the steps applied from the methodology, this section only addresses differences in the processing of Case 2 compared to Case 1, described in the section above. Figure 8 visualizes the delivery times of the second case, showing overall longer delivery times compared to case 1.

In case 2, the mean delivery time is 27 business days, resulting in an RMSE of 10.18 business days if used as prediction. After applying the methodology, the initial results proved to be significantly less performant compared to case 1 (see Table 7). The results show that machine learning does not allow significant improvements in the prediction, compared to using the mean delivery time.

Table 7 Initial results for XGB model (case 2) without constructed desired delivery time

Approach	RMSE (days)	MAE (days)	R^2	NRMSE	NMAE
Mean delivery time	10.18	8.34	0.00	37%	31%
With domain knowledge features (B)	9.90	7.99	0.05	36%	29%
Without domain knowledge features (A)	10.11	8.05	0.01	37%	30%

Table 8 Results of initial modeling (case 2, feature set 2)

Model	RMSE (days)	MAE (days)	R^2
XGB	7.38	5.52	0.56
ANN	7.73	5.82	0.50
SVM	7.79	5.75	0.50
RF	7.81	6.27	0.54
DT	7.83	6.07	0.51
LR	8.08	6.23	0.49

Data preparation and feature extraction

Although the dataset structure is similar to the one presented in case study 1, the desired delivery date was not specified in the dataset for case 2. This results in higher uncertainty, as it is unclear whether the order should be completed by a desired date in the future or as soon as possible. To address this issue, a new feature was engineered and added to the dataset. The algorithm checks whether the orders in the dataset have been completed within the longest range of the used machines and calculates the throughput time of the whole order. Combined with a normalized buffer (+ 3/-3 business days), the algorithm decides if the order was produced as fast as possible depending on the capacity at the time, or if there was an additional delay. If there was an additional delay, the algorithm establishes that there was a desired delivery date causing the delay.

Modeling, evaluation and results

To ensure the comparability of both case studies, Case Study 2 went through the same modeling phase as described above for Case Study 1. With the new feature included in the dataset (see Table 8), the results improve by 2 business days without including the domain knowledge as input. Again XGB showed the best results.

Table 9 shows that compared to case study 1, adding the manufacturing domain knowledge does not improve the RMSE and MAE significantly (0.4 business days), while the R^2 score is enhanced by 0.13. As opposed to the first case, adding planned times information (feature group C) improves the RMSE results by 4 business days, with predictions being 2 business days more accurate than the dates planned by the process planning department. This is due to the fact that the planned dates defined by the process planning department tend to be 2 days late on average (see Fig. 9).

Discussion of results over both cases

Table 10 summarizes the key results of the two case studies presented in the previous sections. The R^2 score shows that with the mean delivery time being used as the delivery time prediction, there is no relationship between the real values and the prediction.

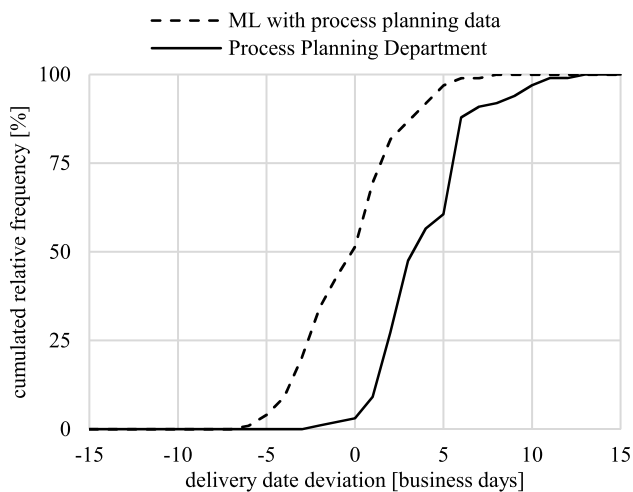The same effect can also be observed in the comparison between the RMSE of the machine learning-based

Table 9 Final results for XGB model (case 2)

Approach	RMSE (days)	MAE (days)	R^2	NRMSE	NMAE
Mean delivery time	10.18	8.34	0.00	37%	31%
Moving average delivery time (1000 orders)	10.41	8.53	− 0.04	38%	31%
Process planning department	5.80	5.00	0.67	21%	18%
ML with planned times (C)	3.81	2.65	0.86	14%	10%
ML without planned times (B)	7.90	6.37	0.48	29%	23%
ML without planned times, without domain knowledge features (A)	8.46	6.71	0.40	31%	25%

Fig. 9 ML model compensates mean delivery time deviation of process planning department

Fig. 10 Added error of approaches compared to prediction results of process planning department

approaches to the RMSE of the prediction from the process planning department (see Fig. 10). The diagram also shows an increasing added RMSE correlating with a decreasing amount of information given to the ML approach.

As mentioned earlier, using the baseline prediction of the mean delivery time calculated on historical data results in significant error, leading to a R^2 score of 0, which means that there is no correlation between prediction and reality. The same applies to the usage of the best approach utilizing the moving average of the delivery time. In both cases, the NRSME highlights the gain in prediction accuracy when different groups of information are included in the datasets. In both cases, the NRSME decreases when all available information is given as input to the machine learning model (feature set 3). However, the decrease in case 2 derives from a global delay of all orders. After the correction of this delay, the additional decrease in prediction error for both cases is below 10%. This indicates that after scheduling all processing steps, unforeseeable changes result in an error that can neither be corrected by the ML model, nor by the process planning department. If the prediction is made earlier in the

process—before the customer order is confirmed or the offer sent out—using only customer order-related data (feature set 1), the prediction error increases compared to the delivery date determined by the process planning department. Without planned delivery times from the process planning department and domain knowledge-specific features (feature set 1), the ML model scores a RMSE that adds around 50% relative error. If domain knowledge-specific features are added (feature set 2), the additional error is reduced to around 35%. However, as the domain knowledge is based on historical data, and therefore it is known prior to the order, the delivery date prediction can be done before the order gets processed by the process planning department. Although this results in less risk of late delivery when confirming the customer order, the added uncertainty results in a higher prediction error. This demonstrates the importance of the knowledge about the performance of the production system. Knowing details about production system dynamics allows the early prediction of the delivery date. From a manufacturing management perspective, the early prediction of the delivery time before the customer order is confirmed has several positive impacts on

Table 10 Comparison of both cases (NRMSE, NMAE, R^2)

Approach	NRMSE		NMAE		R^2	
	Case 1	Case 2	Case 1	Case 2	Case 1	Case 2
Mean delivery time	65%	37%	55%	31%	− 0.01	0.00
Moving average delivery time (best)	63%	38%	53%	31%	− 0.05	− 0.04
Process planning department	31%	21%	20%	18%	0.76	0.67
ML with planned times	30%	14%	21%	10%	0.79	0.86
ML without planned times	41%	29%	32%	23%	0.59	0.48
ML without planned times, without domain knowledge features	50%	31%	40%	25%	0.39	0.40

Table 11 Most important feature groups

Feature group	Directly from dataset	Generated using domain knowledge
Desired delivery time	x	
Quantity ordered (customer order)	x	
Quantity of processes/elements in BOM	x	
WIP calculations		x
Range calculations		x
Bottleneck system identification		x
Summary of external processing		x
Customer rank by quantity of orders		x

the overall production performance. Early knowledge about realistic delivery dates helps avoiding rush orders. Fewer rush orders improve the reliability of the planned schedule which improves the stability of the overall production process. Assuming a more stable production process leads to less unpredictable delays. Fewer delays will eventually result in an even more precise delivery time prediction.

The contribution of different feature groups can be further explained by analyzing the respective top 10 SHAP values. The tables below list the most important features of the two case studies except for the planned times defined by the process planning department (feature set 2). The features are split into two groups: features extracted directly from the dataset (group A) and features that utilize QLM (group B) to be calculated (Table 11).

In both cases, with a SHAP value above 4, the most valuable feature is the desired delivery time. From a manufacturing management perspective, this information highlights the difference between manufacturing lead time and delivery time prediction. Orders may be received with a desired delivery date in the far future. The resulting desired delivery time is the only feature in the examined cases, that provides information to the model, whether the manufacturing processes need to start as soon as possible or at some point in the future, generating a buffer that can help meeting the desired delivery time. The desired delivery time is directly fetched from the input dataset. So are the quantity of items ordered by the customer and the resulting quantity of processes and

elements of the BOM. From the manufacturing management perspective, both features enable the model to estimate the quantity of operations needed to fulfill the order. In the group of features generated using domain knowledge WIP, range, and bottleneck calculations are important features describing the internal manufacturing process. WIP calculations generate information for the model on how many orders are currently in the system. The range calculations add information on the existing capacity of machines and employees to the WIP and condense this information to an estimation of the time needed for a specific system to finish all existing orders. Range calculations allow the identification of bottleneck systems for an individual order, therefore giving the model insight into the earliest possible start for the manufacturing processes. The last two domain knowledge-related feature groups are the processing of external orders and the customer rank based on the quantity of orders placed. From a manufacturing management perspective, the analysis of external manufacturing processes is as important as the analysis of internal manufacturing processes, although most of the underlying information on the suppliers' manufacturing process is expectedly not available. However, information on the suppliers' on-time delivery reliability in the past and the quantity and value of the external order condense into useful information for the model. The high impact of the customer rank based on the quantity of placed orders can be explained by domain knowledge. A high number of orders indicates a higher chance of recurring processes and therefore shorter and less spreading processing times. All of the above-mentioned feature groups improved the prediction of the actual delivery time and are therefore considered mandatory for implementation in practice. Another observation that matches domain knowledge is that the estimated machining times of the single manufacturing steps do not show in the top 10 most important features in both cases. Research shows that for manufacturing lead time determination the transition times (not available in the observed cases) are more important than the actual machining times (Schuh et al., 2019). For both cases, the replenishment times and overall procurement processes did not have a significant impact on the quality of the predicted delivery time. Since procurement processes can significantly delay a production process, if raw materials are not in stock, further research needs to be undertaken to investigate the impact of procurement delays on the delivery time in real world scenarios.

The top 10 most important features for both cases are listed in Table 12 and Table 13.

Table 12 Top ten most important features (case 1)

Feature	SHAP value	(%)
Desired delivery time	4.12	36
WIP bottleneck system Cutter	1.60	14
Quantity of elements in BOM	0.75	7
Quantity ordered (customer order)	0.73	6
Range bottleneck system 'cutter'	0.67	6
WIP bottleneck system 'manual work'	0.63	6
Quantity of utilizations of system 'cutter'	0.51	5
Customer rank by quantity of orders	0.50	4
Max. bottleneck system range	0.49	4
WIP factory (min)	0.40	4

Table 13 Top ten most important features (case 2)

Feature	SHAP value	(%)
Desired delivery time	4.82	34
Price of external machining processes	0.94	7
Quantity of external machining processes	0.93	7
Customer rank by quantity of orders	0.84	6
Range bottleneck system	0.62	4
WIP factory [orders]	0.54	4
Quantity of utilizations of system 'A'	0.53	4
Value of ordered goods	0.47	3
Number of processes needed	0.47	3
Quantity ordered (customer order)	0.35	2

Conclusion

The proposed study aims to provide a machine learning-based approach for delivery date prediction. For the development and testing of such approach, a case study methodology has been used. Case studies are suitable for investigating the evidence of an approach—in this case using machine learning—in a real-world application and evaluating the transferability of the research idea (Yin, 2018). Based on the data collected from the two case studies, three groups of information were identified, namely customer order-related data, production planning and control information or domain knowledge, and information about planned times. Only the first two of these three groups are available for a first offer or as the customer order is received, and as this study aims at the early prediction of delivery dates, these two categories have been used to build the machine learning model. Referring to the four research questions formulated in the methodology, the case studies revealed the following results:

1. Can delivery times be predicted earlier in the process in the same quality the process planning department provides?

Applying ML approaches in the initial offering process or just before confirming the customer order results in less risk of late delivery. However, the added uncertainty results in a higher prediction error of + 35%. Using ML approaches in process planning just before releasing the manufacturing order can help to further reduce the prediction error of the delivery time.

2. Which domain knowledge-specific features do improve the quality of the prediction?

For the machine learning prediction, domain knowledge is extremely valuable, as it helps reducing the model error significantly. In other words, historical knowledge about the performance of the manufacturing system increases the prediction accuracy. The SHAP values showed a high relevance for all features extracted according to the Hanovarian Supply Chain Model. The identification of the current range of the bottleneck system as well as external manufacturing processes and customer rank by quantity of orders showed a high impact on the prediction in both examined cases. However, results across both case studies also show that the specification of the desired delivery date is the most valuable feature for the prediction of delivery times, as it allows the model to distinguish whether to aim for the fastest possible delivery or the delivery in a certain time frame.

3. What are the different quantitative effects of the different features on the training model?

All generated features utilizing QLM showed a high impact on the predicted output in the single-digit percentage range. However, the results showed a significant importance of the desired delivery date. In both cases, one third of the average quantitative change of the prediction value of a customer order relied on the desired delivery date.

4. Do the domain knowledge-specific features help a domain expert to follow up on the decisions made?

The extracted SHAP values show that similar features are used by the ML models in both cases. The trained models utilize input factors as features that enable comprehensible decision-making for a domain expert. The key reason for this lies in the fact that the underlying theory on QLM, that were used for feature generation, has been subject of research for decades and is popular within the domain of manufacturing management.

Overall, this research provides contribution to both academia and industry practice by addressing existing research gaps and providing a practical solution for delivery time prediction at the same time. The ability of machine learning to make predictions earlier in the process allows to define a delivery date at the time of the first offer. This can result in increased efficiency, lower costs for the process planning department, and ultimately it can represent a competitive advantage for manufacturing companies using this approach.

Future research on machine learning methods for delivery time prediction should aim in two directions. First, existing cases should be extended by including external input factors such as stock market indexes. Second, the application of new ML approaches to the field under investigation should be considered. A detailed examination of the impact of the desired delivery date on the prediction results seems reasonable. As this research is based on two cases, more examples are required to strengthen the observed results.

Author contributions All authors contributed to the study conception and design. Data collection, preparation and analysis were performed by AR and CH. The first draft of the manuscript was written by AR, MS and LT. All authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Data availability The datasets analyzed during the current study are not publicly available due confidential company data. The anonymized datasets are available from the corresponding author on reasonable request.

Declarations

Conflict of interest All authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript.

References

Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: A survey on explainable artificial intelligence (XAI). *IEEE Access, 6*, 52138–52160. https://doi.org/10.1109/ACCESS.2018.2870052

Adam, N. R., Bertrand, J. M., Morehead, D. C., & Surkis, J. (1993). Due date assignment procedures with dynamically updated coefficients for multi-level assembly job shops. *European Journal of Operational Research, 68*, 212–227. https://doi.org/10.1016/0377-2217(93)90304-6

Alenezi, A., Moses, S. A., & Trafalis, T. B. (2008). Real-time prediction of order flowtimes using support vector regression. *Computers & Operations Research, 35*, 3489–3503. https://doi.org/10.1016/j.cor.2007.01.026

Alfnes, E., Gosling, J., Naim, M., & Dreyer, H. C. (2021). Exploring systemic factors creating uncertainty in complex engineer-to-order supply chains: Case studies from Norwegian shipbuilding first tier suppliers. *International Journal of Production Economics, 240*, 108211. https://doi.org/10.1016/j.ijpe.2021.108211

Alnahhal, M., Ahrens, D., & Salah, B. (2021). Dynamic lead-time forecasting using machine learning in a make-to-order supply chain. *MDPI Applied Sciences, 11*, 10105. https://doi.org/10.3390/app112110105

Amaro, G., Hendry, L., & Kingsman, B. (1999). Competitive advantage, customisation and a new taxonomy for non make-to-stock companies. *International Journal of Operations & Production Management, 19*, 349–371. https://doi.org/10.1108/01443579910254213

Bechte, W. (1988). Load-orientated manufacturing control: case study: how to make the funnel-model work. APICS 31st Annual Conference, Las Vegas, pp.17–21.

Bender, J., & Ovtcharova, J. (2021). Prototyping machine-learning-supported lead time prediction using AutoML. *Procedia Computer Science*. https://doi.org/10.1016/j.procs.2021.01.287

Bezirgiannidis, N., Burleigh, S., & Tsaoussidis, V. (2013). Delivery time estimation for space bundles. *IEEE Transactions on Aerospace and Electronic Systems, 49*, 1897–1910. https://doi.org/10.1109/TAES.2013.6558026

Bhalla, S., Alfnes, E., & Hvolby, H.-H. (2023). Tools and practices for tactical delivery date setting in engineer-to-order environments: A systematic literature review. *International Journal of Production Research, 61*, 2339–2371. https://doi.org/10.1080/00207543.2022.2057256

Birkie, S. E., & Trucco, P. (2016). Understanding dynamism and complexity factors in engineer-to-order and their influence on lean implementation strategy. *Production Planning & Control, 27*, 345–359. https://doi.org/10.1080/09537287.2015.1127446

Braglia, M., Frosolini, M., Gallo, M., & Marrazzini, L. (2019). Lean manufacturing tool in engineer-to-order environment: Project cost deployment. *International Journal of Production Research, 57*, 1825–1839. https://doi.org/10.1080/00207543.2018.1508905

Cannas, V. G., Gosling, J., Pero, M., & Rossi, T. (2020). Determinants for order-fulfilment strategies in engineer-to-order companies: Insights from the machinery industry. *International Journal of Production Economics, 228*, 107743. https://doi.org/10.1016/j.ijpe.2020.107743

Chapman, P., Clinton, J., Kerber, R., Khabaz, T., Reinartz, T., & Shearer, C. (2000). CRISP-DM 1.0 Step-by-step data mining guide.

Chapman, S. N., Arnold, J. R. T., Gatewood, A. K., & Clive, L. M. (2017). *Introduction to materials management*. Pearson.

Choetkiertikul, M., Dam, H. K., Tran, T., & Ghose, A. (2017). Predicting the delay of issues with due dates in software projects. *Empirical Software Engineering, 22*, 1223–1263. https://doi.org/10.1007/s10664-016-9496-7

de Araujo, A. C., & Etemad, A. (2021). End-to-end prediction of parcel delivery time with deep learning for smart-city applications. *IEEE Internet of Things Journal*. https://doi.org/10.1109/JIOT.2021.3077007

Goldratt, E. M., & Cox, J. (1984). *The goal: Excellence in manufacturing*. North River Press.

Golpayegani, D., Pandit, H. J., & Lewis, D. (2023). Comparison and analysis of 3 key AI documents: EU's proposed AI Act, assessment list for trustworthy AI (ALTAI), and ISO/IEC 42001 AI management system. In L. Longo & R. O'Reilly (Eds.), Artificial Intelligence and Cognitive Science (Vol. 1662, pp. 189–200, Communications in Computer and Information Science). Springer.

Grabenstetter, D. H., & Usher, J. M. (2014). Developing due dates in an engineer-to-order engineering environment. *International Journal of Production Research, 52*, 6349–6361. https://doi.org/10.1080/00207543.2014.940072

Hawkins, D. M., Basak, S. C., & Mills, D. (2003). Assessing model fit by cross-validation. *Journal of chemical information and computer sciences, 43*, 579–586. https://doi.org/10.1021/ci025626i

Hicks, C., & Braiden, P. M. (2000). Computer-aided production management issues in the engineer-to-order production of complex capital goods explored using a simulation approach. *International Journal of Production Research, 38*, 4783–4810. https://doi.org/10.1080/00207540010001019

Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: Principles and practice*. Otexts Online Open-Access Textbooks.

Jodlbauer, H., & Tripathi, S. (2023). Due date quoting and rescheduling in a fixed production sequence. *International Journal of Production Research*. https://doi.org/10.1080/00207543.2023.2179342

Kettner, H., & Bechte, W. (1981). New methods of production control by load-orientated order release. VDI.

Khiari, J., & Olaverri-Monreal, C. (2020). *Boosting algorithms for delivery time prediction in transportation logistics* (pp. 251–258). IEEE.

Liang, Y.-C., Lee, Z.-H., & Chen, Y.-S. (2012). A novel ant colony optimization approach for on-line scheduling and due date determination. *Journal of Heuristics, 18*, 571–591. https://doi.org/10.1007/s10732-012-9199-1

Lödding, H. (2013). *Handbook of manufacturing control: Fundamentals, description, configuration*. Springer.

Maier, J. T., Rokoss, A., Green, T., Brkovic, N., & Schmidt, M. (2022). *A systematic literature review of machine learning approaches for the prediction of delivery dates*. Publish-Ing.

McCutcheon, D. M., & Meredith, J. R. (1993). Conducting case study research in operations management. *Journal of Operations Management, 11*, 239–256. https://doi.org/10.1016/0272-6963(93)90002-7

Mohsen, O., Petre, C., & Mohamed, Y. (2023). Machine-learning approach to predict total fabrication duration of industrial pipe spools. *Journal of Construction Engineering and Management*. https://doi.org/10.1061/JCEMD4.COENG-11973

Moses, S., Grant, H., Gruenwald, L., & Pulat, S. (2004). Real-time due-date promising by build-to-order environments. *International Journal of Production Research, 42*, 4353–4375. https://doi.org/10.1080/00207540410001716462

Murphy, R., Newell, A., Hargaden, V., & Papakostas, N. (2019). Machine learning technologies for order flowtime estimation in manufacturing systems. *Procedia CIRP*. https://doi.org/10.1016/j.procir.2019.03.179

Nguyen, S. (2016). A learning and optimizing system for order acceptance and scheduling. *The International Journal of Advanced Manufacturing Technology*. https://doi.org/10.1007/s00170-015-8321-6

Olhager, J. (2003). Strategic positioning of the order penetration point. *International Journal of Production Economics, 85*, 319–329. https://doi.org/10.1016/S0925-5273(03)00119-1

Öztürk, A., Kayalıgil, S., & Özdemirel, N. E. (2006). Manufacturing lead time estimation using data mining. *European Journal of Operational Research, 173*, 683–700. https://doi.org/10.1016/j.ejor.2005.03.015

Panzer, M., & Gronau, N. (2023). Designing an adaptive and deep learning based control framework for modular production systems. *Journal of Intelligent Manufacturing*. https://doi.org/10.1007/s10845-023-02249-3

Paprocka, I., & Cyba, S. (2015). Assessment of production capacity and ability of rapid response to changing customer expectations. *Applied Mechanics and Materials, 809–810*, 1378–1383. https://doi.org/10.4028/www.scientific.net/AMM.809-810.1378

Polim, R., Kumara, S., & Gomes, B. M. (2017). Real-time shipment duration prediction. 67th Annual Conference and Expo of the Institute of Industrial Engineers 2017, p. 67.

Powell, D., Strandhagen, J. O., Tommelein, I., Ballard, G., & Rossi, M. (2014). A new set of principles for pursuing the lean ideal in engineer-to-order manufacturers. *Procedia CIRP, 17*, 571–576. https://doi.org/10.1016/j.procir.2014.01.137

Ragatz, G. L., Mabert, V. L., & Deuse, J. (1984). A framework for the study of due date management in job shops. *International Journal of Production Research, 22*, 685–695. https://doi.org/10.1080/00207548408942488

Rao, U. S., Swaminathan, J. M., & Zhang, J. (2005). Demand and production management with uniform guaranteed lead time. *Production and Operations Management, 14*, 400–412. https://doi.org/10.1111/j.1937-5956.2005.tb00229.x

Rau, H., Tsai, M.-H., Chen, C.-W., & Shiang, W.-J. (2006). Learning-based automated negotiation between shipper and forwarder. *Computers & Industrial Engineering, 51*, 464–481. https://doi.org/10.1016/j.cie.2006.08.008

Richter, R., Syberg, M., Deuse, J., Willats, P., & Lenze, D. (2023). Creating lean value streams through proactive variability management. *International Journal of Production Research, 61*, 5692–5703. https://doi.org/10.1080/00207543.2022.2111614

Rokoss, A., Kramer, K., & Schmidt, M. (2021). How a learning factory approach can help to increase the un- derstanding of the application of machine learning on production planning and control tasks. In W. Sihn & S. Schlund (Eds.), Competence development and learning assistance systems for the data-driven future (pp. 125–142, Schriftenreihe der Wissenschaftlichen Gesellschaft für Arbeits- und Betriebsorganisation (WGAB) e.V). Gito Verlag

Roman, D. B., & Del Vallei, A. G. (1996). Dynamic assignation of due-dates in an assembly shop based in simulation. *International Journal of Production Research, 34*, 1539–1554. https://doi.org/10.1080/00207549608904983

Schmidt, M., & Schäfers, P. (2017). The hanoverian supply chain model: Modelling the impact of production planning and control on a supply chain's logistic objectives. *Production Engineering, 11*, 487–493. https://doi.org/10.1007/s11740-017-0740-9

Schneckenreither, M., Haeussler, S., & Gerhold, C. (2021). Order release planning with predictive lead times: A machine learning approach. *International Journal of Production Research*. https://doi.org/10.1080/00207543.2020.1859634

Schuh, G., Prote, J.-P., Sauermann, F., & Franzkoch, B. (2019). Databased prediction of order-specific transition times. *CIRP Annals*. https://doi.org/10.1016/j.cirp.2019.03.008

Shet, A., Hanumanth Naik, D., & Shetty, S. (2022). Stock price prediction using machine learning. *International Journal of Engineering*

Applied Sciences and Technology, 7, 225–228. https://doi.org/10.33564/ijeast.2022.v07i02.034

Steinberg, F., Burggaef, P., Wagner, J., & Heinbach, B. (2022). Impact of material data in assembly delay prediction—a machine learning-based case study in machinery industry. *The International Journal of Advanced Manufacturing Technology, 120*, 1333–1346. https://doi.org/10.1007/s00170-022-08767-3

Thürer, M., Huang, G., Stevenson, M., Silva, C., & Filho, M. G. (2012). The performance of due date setting rules in assembly and multi-stage job shops: An assessment by simulation. *International Journal of Production Research, 50*, 5949–5965. https://doi.org/10.1080/00207543.2011.638942

Vabalas, A., Gowen, E., Poliakoff, E., & Casson, A. J. (2019). Machine learning algorithm validation with a limited sample size. *PloS one, 14*, e0224365. https://doi.org/10.1371/journal.pone.0224365

Waubert de Puiseau, C., Meyes, R., & Meisen, T. (2022). On reliability of reinforcement learning based production scheduling systems: A comparative survey. *Journal of Intelligent Manufacturing, 33*, 911–927. https://doi.org/10.1007/s10845-022-01915-2

Wiendahl, H.-P. (1997). *Manufacturing control: Logistical control of production processes based on the funnel model*. Hanser.

Wiendahl, H.-P., & Tönshoff, K. (1988). The throughput diagram—an universal model for the illustration, control and supervision of logistic processes. *CIRP Annals, 37*, 465–468. https://doi.org/10.1016/S0007-8506(07)61678-3

Xu, Y., & Goodacre, R. (2018). On splitting training and validation set: A comparative study of cross-validation, bootstrap and systematic sampling for estimating the generalization performance of supervised learning. *Journal of Analysis and Testing, 2*, 249–262. https://doi.org/10.1007/s41664-018-0068-2

Yang, D. H., Hu, L., & Qian, Y. (2017). *Due date assignment in a dynamic job shop with the orthogonal kernel least squares algorithm*. IOP Publishing. https://doi.org/10.1088/1757-899X/212/1/012022

Yang, D., & Zhang, X. (2018). *A hybrid approach for due date assignment in a dynamic job shop*. IEEE. https://doi.org/10.1109/ICMIC.2017.8321562

Yin, R. K. (2018). *Case study research and applications: design and methods*. SAGE.

Zhang, H., Leng, J., Zhang, H., Ruan, G., Zhou, M., & Zhang, Y. (2021). A deep reinforcement learning algorithm for order acceptance decision of individualized product assembling. In 2021 IEEE 1st international conference on digital twins and parallel intelligence (DTPI), Beijing, China, 15.07.2021–15.08.2021 (pp. 21–24). IEEE. https://doi.org/10.1109/DTPI52967.2021.9540190.