IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 21 November 2023, accepted 2 January 2024, date of publication 11 January 2024, date of current version 22 January 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3353029

APPLIED RESEARCH

Throughput Time Predictions Along the Order Fulfilment Process

TOBIAS HILLER, TABEA MARIE DEMKE, AND PETER NYHUIS
Institute of Production Systems and Logistics, Leibniz University Hannover, 30823 Garbsen, Germany
Corresponding author: Tobias Hiller (hiller@ifa.uni-hannover.de)

This work was supported in part by the German Federal Ministry of Education and Research; and in part by the Aviation Research and Technology Program of the Lower Saxony Ministry of Economics, Labor, Transport and Digitalization under Grant ZW 1-80157862.

ABSTRACT Planned times for the throughput of production are key components to production planners for determining delivery dates with customers, capacity planning, scheduling and order coordination. While traditional estimation methods often rely on basic statistics and expert knowledge, data mining respectively machine learning offers the potential to compute more precise predictions for order-specific planned throughput times. Factors that lead to deviations from the plan are diverse and thus challenging to consider in the various production planning tasks along the order fulfilment process. Intelligent throughput time predictions promise a remedy. Yet, predictive models are often not designed to be practically applicable due to a lack of consideration of the various characteristics of each stage of the order fulfilment process. To address this gap, this paper takes a closer look at the prediction of throughput times for the various stages of order fulfilment. Based on the Cross Industry Standard Process for Data Mining, the characteristics of the individual steps to build a prediction model are elaborated with a focus and business and data understanding and then examined in a case study. From that, practical implications are derived and guidance for practitioners is given. A key finding is that predictions are less accurate in the early stages of order fulfillment. Prediction quality naturally enhances over time, since more and more order details are known. In conclusion, an iterative prediction process with an evolving database ensures good prediction quality, especially in the late stages of order fulfillment.

INDEX TERMS Production planning, machine learning, throughput time prediction, data analysis.

I. INTRODUCTION

On-time delivery of order-related products is a key success factor for companies. Ensuring a high level of delivery reliability remains a top priority for manufacturers and, along with costs and quality, is among the most important prerequisites for a successful standing in global competition [1], [2], [3], [4].

Along the order fulfilment process, planned times for the throughput are used for different means. While at the beginning of the order fulfilment process, planned throughput times are used to coordinate due dates, compute delivery dates, or plan rough capacities. In short-term planning activities, planned throughput times are used to conduct fine resource and schedule planning [5], [6]. For determining planned throughput times, there are various methods based on statistics, general assumptions, or queuing theory [7], [8]. These are often based on mean value estimate or calculated using only a few key figures, such as work content [8], [9]. The application of Machine Learning (ML) leads to more sophisticated predictions of throughput times [6], [10], [11], [12]. Additionally, technical development can be greatly simplified by several ML frameworks, including AutoKeras, AutoSk-Learn and TPOT, which automate the development process [13], [14], [15].

Along the order fulfilment process, Production Planning and Control (PPC) manages the processes in the internal supply chain. Different PPC tasks use planned values for throughput time at different points in order processing to meet different objectives, have specific constraints, and have

The associate editor coordinating the review of this manuscript and approving it for publication was Zhiwu Li.

access to varying information [5]. However, these aspects are often not considered in predictions [6]. Therefore, this paper examines the application of ML-based throughput time predictions in different phases of order fulfilment to discover differences between the phases and the resulting opportunities. In addition, we provide a practice-oriented discussion of important elements in creating ML predictions.

To achieve this, the paper is structured as follows: First, we present the theoretical foundations, followed by a discussion of the problem of how to embed throughput time predictions in PPC (Section III). Based on the established Cross Industry Standard Process for Data Mining (CRISP-DM), critical elements for developing ML-based throughput time predictions along order fulfilment are elaborated (Section IV). Finally, Section V provides a case study to illustrate the resulting conclusions in terms of throughput time predictions along the order fulfillment process. This paper builds on a previously conducted systematic literature analysis (see [6]).

II. THEORETICAL FOUNDATIONS

This section includes the basics of order fulfilment and throughput time, PPC, and data-driven prediction approaches using ML.

A. ORDER FULFILMENT AND THROUGHPUT TIME

The order fulfilment process describes the execution of various activities to complete an order. In the context of production, this usually includes the acquisition of (customer) orders, purchasing secondary requirements, production, shipping and post-delivery activities [16], [17]. The process varies depending on the type of company and the order fulfilment strategy. In customized production (or engineer-to-order), order fulfilment includes product development and manufacturing engineering phases [18]. In make-to-order (m-t-o) companies, these phases are not part of the order fulfilment process. For the further course of this paper, the order fulfilment process is defined as shown in Figure 1.

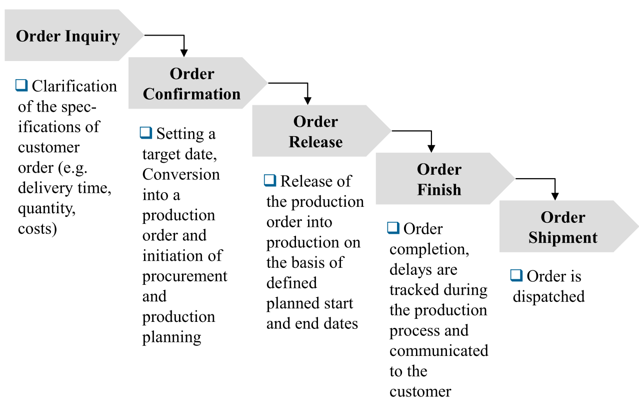

FIGURE 1. Order fulfilment process [17], [19].

The definition of these five phases outlines all key milestones in the processing of an order in an customized production (such as m-t-o). This includes the initiation of a customer order and thus the starting point of a customer-oriented production. It contains the order confirmation, which typically triggers procurement processes as well as production planning tasks. Once the secondary procurement or allocation of requirements have been fulfilled or scheduled, the order can be released for production. As soon as an order has been completed, the product can be delivered [5], [20].

Throughput time (TTP) or lead time is defined as the time that elapses between the release of an order (to production) to its completion [21]. Equation 1 defines the calculation for a production order k:

$$\mathrm{TTP}_{\mathrm{k}}^{\mathrm{Order}} = \mathrm{FO} - \mathrm{RO} \tag{1}$$

where FO (in calendar days) denotes the time of finishing of an order and RO (in calendar day) the time of order release.

A distinction can be made with regard to the level of detail: Operation throughput times ($\mathrm{TTP}^{\mathrm{Operation}}$) describe the time span of a sub-processes of an order (e.g. turning shop, milling center, etc.). Thus, the sum of operation throughput times equals the order throughput time [3]:

$$\mathrm{TTP}_{\mathrm{k}}^{\mathrm{Order}} = \sum\nolimits_{\mathrm{i}=1}^{\mathrm{N}} \mathrm{TTP}_{\mathrm{k,i}}^{\mathrm{Operation}} \tag{2}$$

where i corresponds to one step of N necessary productions process steps to fulfil the order k. Operation throughput times can be decomposed in processing (TOP) as well inter-operation time (TIO) for an operation [3]:

$$\mathrm{TTP}_{\mathrm{k,i}}^{\mathrm{Operation}} = \mathrm{TIO}_{\mathrm{k,i}} + \mathrm{TOP}_{\mathrm{k,i}} \tag{3}$$

In contrast to $\mathrm{TTP}_{\mathrm{k}}^{\mathrm{Order}}$, delivery time (TD) describes the time between order confirmation and shipment and includes administration time (TA), procurement time (TP), shipping time (TS) and time buffers (TB) [21]:

$$\mathrm{DT}_{\mathrm{k}} = \mathrm{TTP}_{\mathrm{k}}^{\mathrm{Order}} + \mathrm{TP}_{\mathrm{k}} + \mathrm{TA}_{\mathrm{k}} + \mathrm{TS}_{\mathrm{k}} + \mathrm{TB}_{\mathrm{k}} \tag{4}$$

B. PRODUCTION PLANNING AND CONTROL

PPC manages the order fulfilment process. PPC finds itself in a field of tension between competing logistical targets [5]. Central targets comprise a low work in progress (WIP), a high schedule compliance, short throughput times, and a high capacity utilization [3]. A central purpose of production planning is to schedule the production program (short and long term) and plan all activities assigned with the manufacturing process such as procurement, capacity planning, supplier coordination etc.) [22]. Production controls primarily responsible for ensuring that all production plans are successfully implemented even in the event of disruptions [21]. In the context of computation and utilization of planned TTP, we focus on four relevant tasks within PPC: Order classification, throughput scheduling, capacity planning and order coordination (cf. [5], [6]). Scheduling and capacity planning are treated jointly in this context, since they usually overlap (in the coming only referred to as throughput scheduling) [9]:

- Order clarification: Determination of due dates based on the delivery dates requested by the customer. This requires calculating (rough) planned TTP at order level, checking available capacities and, if necessary, setting time buffers.
- Throughput scheduling and capacity planning: Determination of start and end times of orders and the individual sub-processes by allocating operations to resources in a given time period [23]. This requires the determination of detailed planned TTP at the order and operation level. The scheduling can be done e.g. by forward or backward scheduling [9].
- Order coordination: Comparison of the current progress of individual production orders with the plan. This task requires determining remaining TTP to notify customers in terms of any delays.

Within production configuration as described in [4], the tasks defined in the context of computation and utilization of planned TTP can be assigned to the tactical or operational level.

Numerous approaches exist for the calculation of planned TTP. For example, according to Wiendahl [9], common static methods for calculating planned TTP at order level are based on past mean values (see Equation (5) or (6)) or multiples of the processing time (see (7)).

$$\mathrm{TTP}_{k}^{\mathrm{Order,plan}} = \mathrm{TTP}_{\mathrm{mean}}^{\mathrm{Order}} \tag{5}$$

$$\mathrm{TTP}_{k}^{\mathrm{Order,plan}} = \sum\nolimits_{i=1}^{N} \mathrm{TTP}_{\mathrm{i,mean}}^{\mathrm{Operation}} \tag{6}$$

$$\mathrm{TTP}_{k}^{\mathrm{Order,plan}} = \sum\nolimits_{i=1}^{N} b \cdot \mathrm{TOP}_{i} \tag{7}$$

These calculations can be performed by clustering the mean values by product groups or globally [9]. In addition to static methods, there are dynamic approaches to determine planned TTP at order level. An exemplary dynamic approach is based on exponential smoothing, such as shown in [24]:

$$\mathrm{TTP}_{m,t}^{\mathrm{Order,plan}} = \alpha \cdot \mathrm{TTP}_{m,t}^{\mathrm{Order,LO}} + (1-\alpha) \cdot \mathrm{TTP}_{m,t-1}^{\mathrm{Order,plan}} \tag{8}$$

where the calculation of planned TTP is performed for a time period t and a product group m. By subtracting the last observed value of the TTP ($\mathrm{TTP}^{\mathrm{Order,LO}}$) with the TTP of the past time period, dynamic effects shall be taken into account. The coefficient α describes the smoothing parameter (cf. [24]).

C. MACHINE LEARNING FOR DATA-DRIVEN THROUGHPUT TIME PREDICTION

In contrast to traditional TTP calculation methods (static or dynamic), ML enables production planners to consider numerous influencing factors when calculating planned TTP.

ML is a branch of artificial intelligence applications that employs statistical/ computational methods to provide data-based solutions for specific issues, and is increasingly being used in manufacturing [25], [26], [27], [28]. ML uses these methods to learn from past experiences expressed in data and performs different tasks such as classification, regression, ranking or clustering [26]. It can be divided into different scenarios based on the training data available to the user and objective; the most common are: supervised learning, unsupervised learning and reinforcement learning [26]. Supervised learning uses labelled data and performs predictions for unseen data. Mainly, this involves regression and classification tasks [26].

Unsupervised learning is based on unlabeled data, two possible applications are clustering and dimensionality reduction. Since there are no labelled examples, it is difficult to quantify the performance of such models [26]. Applications in the production context comprise e.g. predictive maintenance [29].

Reinforcement learning combines training and testing phases by immediately rewarding every action so that it is possible to interact with and sometimes influence the environment. The goal is to maximize the reward through a series of actions and iterations with the environment [26]. A potential field of application is production control [30]. Predicting TTP is usually a regression task in the context of supervised learning [6].

The process for creating (supervised) ML models can be described as shown in Figure 2:

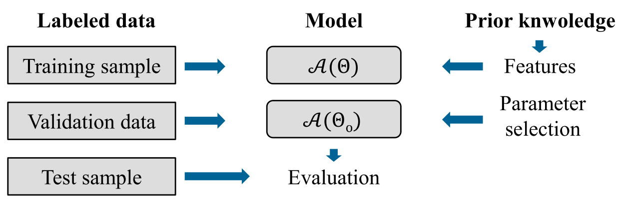

FIGURE 2. Principle depiction of a learning process (supervised) [26].

In the beginning, a labelled data set containing the target of a prediction is needed. In our case of TTP prediction, these are the (nominal) values for TTP of different production orders.

Furthermore, we need features, a set of attributes that represent the independent variables of the model. Feature engineering, the process of creating and enriching raw data into attributes that have a positive impact on prediction performance, is one of the most time-consuming and at the same time most essential tasks in the creation of an ML application [31], [32]. Suitable features can lead the learning algorithm effectively, while uninformative features can be misleading. The choice of features is largely up to the user and reflects their prior knowledge of the issue [26].

The selected features are subsequently used to train a learning algorithm ($\mathcal{A}$), using the training sample by adjusting its free parameters (hyperparameters, Θ). For example, learning algorithms are linear regressions, support-vector regression, or regression trees [26].

Using the validation data, the set of hyperparameters with the best results is selected (Θo). Finally, the performance of the learning algorithm is evaluated using the test sample [26]. Relating to regression tasks, common metrics for the evaluation are the mean absolute error (MAE), the mean square error (MSE) or the root mean squared error (RMSE) [33].

In our context, the error describes the deviation of the prediction (marked with the superscript "pred") from its actual value (marked superscript "actual") which can be denoted as follows [33]:

$$e_i = TTP_i^{actual} - TTP_i^{pred} \tag{9}$$

The MAE and MSE can be calculated as follows [33]:

$$MSE = \frac{1}{N} \sum_{i=1}^{N} e_i^2 \tag{10}$$

$$MAE = \frac{1}{N} \sum_{i=1}^{N} |e_i| \tag{11}$$

Increasingly, automated approaches are used to create ML models. This topic is titled as "AutoML", see e.g. [14], [15], [34]). These automate the selection of a suitable algorithm, (e.g. whether to take a linear regression or a regression tree) and take over the hyperparameter setting (see e.g. [15]). The user is therefore left with selecting a business case, the input data and the evaluation as well as application. We discuss these steps in greater detail in Section IV relating to the CRISP-DM.

In the current state of research, there are many different approaches to the prediction of throughput times. A systematization and literature review can be found in [6] or [12]. As mentioned in the introduction, often, not enough attention is paid to the different phases of order fulfilment and the purpose of prediction, as they are different for each step and the available information changes drastically. This problem is addressed in the following section.

III. ON THE PROBLEM OF EMBEDDING THROUGHPUT TIME PREDICTION IN ORDER FULFILMENT

Planned values for the TTP of production orders are used in several phases along the order fulfilment process. As already elaborated, we consider three PPC tasks where planned TTP are deployed: order clarification, throughput scheduling and order coordination [6]. The use of TTP for these tasks differs in terms of objective, purpose and available information (cf. [5], [17]). In view of this, Figure 3 differentiates the three PPC tasks under consideration, the corresponding order fulfilment phases and the associated prediction targets.

In order clarification, a company needs to define a TTP (hereafter denoted as rough TTP) to communicate a delivery date to the customer based on the customer's requirements. At the same time, this defined rough TTP ($TTP^{Order,rough-plan}$) must be feasible for production. Here, a conflict arises in terms of short delivery times that the customer usually wants as well as the ability to meet short delivery times and deliver on time [5], [35]. The result of this task is a target date.

Once the target date has been set and the order has been confirmed, throughput scheduling needs to schedule the order and its operations in such a way that the target

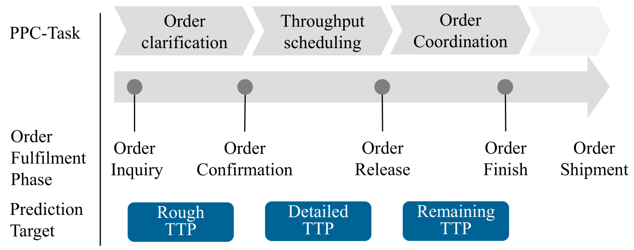

FIGURE 3. Phases of order fulfilment, PPC tasks and prediction targets based on [6].

date is met. To schedule orders and their operations optimally, the planned TTP for the operations are required in addition to the planned TTP for the entire order [5], [35]. The objective is to plan the throughput in the most efficient way, taking into account the logistical objectives. The objective of this task is to ensure that TTP and WIP are low whilst capacity utilization and schedule compliance are maintained high [3]. Estimated planned start and finish dates for the production order ($TTP^{Order,detail-plan}$) and its operations ($TTP^{Operation,detail-plan}$) describe the results of this task. A safety time buffer is usually included to consider schedule deviations [5], [35].

Order coordination takes over the task of coordinating the orders after release. The objective is to compare the progress of the orders with the planned schedule and to inform the customer in the event of delays [5], [35]. For this purpose, delivery time deviations and remaining TTP ($TTP^{Order,remain-plan}$) may have to be calculated.

In order to make the best possible prediction, it is vital to pay close attention to the PPC task in question. Since the objective differs in each task, the requirements for the calculated planned TTP as described above will differ. Furthermore, during the course of order fulfilment process, the information that can be used at each phase is very different and increases as time goes on (cf. [6]). This information growth provides the potential for better predictions. A good example for this provides the WIP.

WIP, as a measure of the amount of jobs currently in production, usually measured in (working) hours or number of jobs, has a significant impact on the TTP according to Little's Law (see e.g. [3], [36]). Usually, information about the current system load, e.g. in the form of WIP, is not available at the beginning of the order clarification (at most in the form of estimates or planned values). If this indicator is to be meaningfully included in a prediction, the prediction has to be as close to the release (close before or after), so that the WIP is actually known. As pointed out in [6] most approaches investigated use the WIP as a key figure for the prediction.

In a nutshell, TTP predictions can be used in different means along the order fulfilment process in different PPC tasks. At each point in time, different information is available to make predictions. Existing approaches tend to focus on a single point in time and do not systematically compare the different stages of the order fulfilment process taking into account the information available. Therefore, the following section presents a procedure for predicting TTP based on the CRISP-DM process, which considers these issues and other critical aspects.

IV. PREDICTING THROUGHPUT TIMES ALONG ORDER FULFILMENT

The CRISP-DM (see Figure 4) is model for data mining applications. The CRISP-DM comprises six steps: Business Understanding, Data Understanding, Data Preparation, Modelling, Evaluation and Deployment [37]. There are alternative approaches as described in [38] or [39] that adapt the CRISP-DM relating to the use of ML, that we also consider in the following. The delimitation of ML and DM as subdomains of artificial intelligence is done in [25] and is not the subject of this paper. For example, the process model discussed in [38] provides an extra phase for monitoring and maintenance in addition to the deployment phase (which we do not consider) and combines phases of Business & Data Understanding. This adjustment is suitable for our purposes, so we will also follow it.

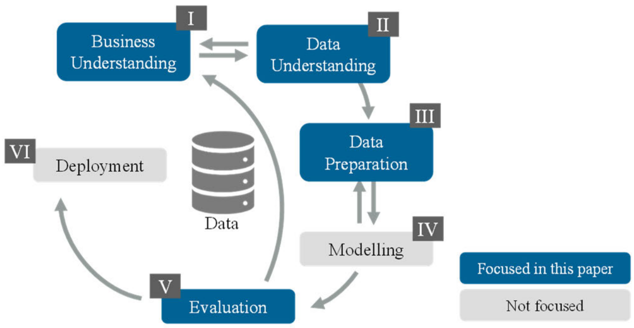

FIGURE 4. CRISP-DM [37].

In the following, we describe the relevant contents and tasks of the individual steps with regard to our field of application, the prediction of TTP, considering the CRISP-DM [37] as well as the extension by Studer et al. [38].

A. BUSINESS & DATA UNDERSTANDING (STEP I AND II)

This phase includes the definition of the scope as well as setting *success criteria* (A.1), *data collection* (A.2), *data exploration* and *assessment of initial situation* (A.3) and *data quality verification*(A.4) [37], [38].

1) SCOPE AND SUCCESS CRITERIA

The first step in defining the scope is to select the phase within the order fulfilment process to which the TTP prediction will be applied through the corresponding PPC task and what business units will be involved. Understanding the requirements and needs of the business unit is crucial for the ML application to be accepted [38]. While order clarification is often handled by key account management or sales, throughput scheduling is the responsibility of production planners.

Scope definition, often the key step in creating a prediction (cf. [40]), involves defining the specific area of production to be covered by the TTP prediction. Therefore, it is necessary to look at the associated material and process flow. It is also necessary to define the objectives (such as improvement of plan stability) to be achieved and the criteria by which the use of ML can be considered a success. In general, a distinction can be made between business and ML success criteria [38].In this context, the metrics MAE or MSE described in Section II-C, can be applied. For example, a specific value for the prediction's MAE can be defined as a target. It is also possible to measure against currently used planning values. Conducting feasibility checks before starting an ML project is also crucial to minimize the risk of premature failure due to unrealistic objectives etc. [38]. The general availability of datasets should be also assessed. In addition, the applicability of the ML technology should be examined, e.g. through proof of concept or literature reviews [38]. It may also be necessary to consider legal restrictions [37] (e.g. when dealing with personal data).

2) DATA COLLECTION

Data collection is an essential part of carrying out an ML project and should not be confused with feature selection (cf. [32]), where the data used for modelling are selected. At this stage, the aim is to gather a sufficient amount of generally relevant data from various sources [38].

As ML tries to find patterns between the relationships between our input features and our target variable (TTP), we need to collect data that we assume has an influence/relationship in or on TTP. Irrelevant information can affect the predictive performance of many models negatively. Domain-specific knowledge helps to separate potentially meaningful information from irrelevant information [41]. A first guide to collect initial data is provided by the TTP driver tree by [42] and [43]. The TTP driver tree contains the main factors that generally affect TTP and their deviation from plan (cf. Figure 5).

These include physical factors such as the production structure and planning factors that influence the throughput itself and its deviation from the plan.

In addition, it is necessary to take into account whether the data collected is available at the time of the prediction. Table 1 shows which information would be available at which phases of order fulfilment and the associated PPS tasks. Please note that this is difficult to generalize and is domain specific. Parentheses indicate that availability cannot be given in general terms as it may depend, for example, on the type of order fulfillment. For example, in the case of engineer to order, the number of process steps within order clarification may not be known.

However, certain information is only physically available at a certain point in time (e.g. an unpunctual start can only be known after the order has been released). Nevertheless, other information, such as lot size, could potentially be determined

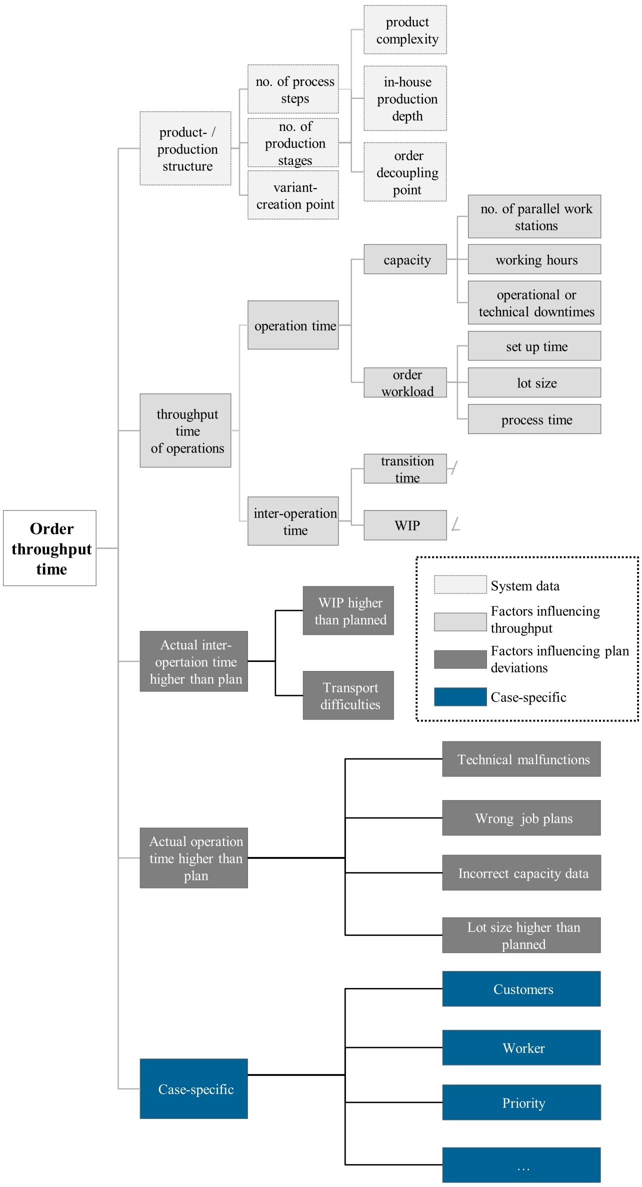

FIGURE 5. Drivers influencing order throughput time (based on [42], [43]).

during order clarification. Typically, however, this information is only available at the scheduling stage, i.e. after the order has been clarified.

It can be stated that the amount of available data is increasing along the order fulfilment process and orders are becoming more predictable. In other words, there is less data available at the beginning of the order fulfilment process than at later phases.

3) DATA EXPLORATION AND ASSESSMENT OF INITIAL SITUATION

This task takes a deeper look at the issue at hand by analyzing the data collected. Univariate analyses (i.e. of the target variable TTP) and bivariate analyses of the relationship between the TTP and the other attributes should be carried out [37]. With regard to univariate analysis, one could look at the distribution of TTP and the current target achievement by comparing it to the current planned values. For bivariate analysis, correlations and relationships between TTP and other features could be calculated, for example, using scatter matrix plots and non-graphical correlation analyses. For this task in general, it may be appropriate to follow the procedures of explorative data analysis (cf. [44]).

Especially, it is worthwhile to take a closer look at the key influences described in Figure 5. In addition, production areas or systems that substantially determine TTP (such as bottlenecks) should be investigated.

The initial data exploration also allows us to assess the baseline situation in terms of plan stability and accuracy of the planned TTP. Here it might be interesting to circle back and reassess and evaluate the success criteria.

TABLE 1. Principle availability of features in different PPC tasks.

Feature	Order clarification	Throughput scheduling	Order co-ordination
Physical factors			
No. of process steps	(x)	x	x
No. of production stages	(x)	x	x
Product complexity	x	x	x
Factors influencing throughput			
Capacity	(x)	x	x
Setup time	(x)	x	x
Lot size	(x)	x	x
Processing time	(x)	x	x
Actual WIP		(x)	x
Transition time		(x)	x
Factors influencing plan deviations			
Lot size too high		(x)	x
Capacity lower than plan		(x)	x
Sequence deviation through unsuitable priority			x
(Unexpected) backlog/ unpunctual start			x

4) DATA QUALITY VERIFICATION

The purpose of this task is to determine whether the objectives and success criteria can be achieved with the given set of data. It is based on the previous exploratory data analysis and should include a general description of the data (information on format, units, etc.). A lack of data could trigger step A.2. [38]. Other issues to be considered are whether the data is complete, certain errors are common or there are large numbers of missing values [37]. This step also contains, according to [38], the definition of data requirements as for example expected feature values, format or maximum of missing values. To reduce the risk of bias, process experts should be consulted. In a final verification step, it must be decided whether the data meet the requirements and whether the ML application is feasible [38].

B. DATA PREPARATION (STEP III)

This phase includes *feature engineering* (B.1), *feature selection*(B.2) and *cleaning* (B.3) [37], [38]. Since we use AutoML for modelling, there is partially no need for (technical) data pre-processing steps including formatting, normalization or data transformation. The extent of needed pre-processing depends on the chosen framework. This also applies to feature selection.

1) FEATURE ENGINEERING

Feature engineering refers to the task of transforming and pre-processing raw data into features that are suitable for ML and enable a good output [32]. Feature engineering is difficult to generalize because features can often only be defined in the context of the model and the data, and data and models are very diverse across different cases [32].

In the following, we distinguish between two forms of feature engineering: the context-specific engineering of features, which hopefully has a positive impact on the output, and the rather technical, generalized construction of features, which is often done automatically by different frameworks or AutoML (cf. [14], [15]).

Context-specific (and domain-specific) engineering describes the evaluation of parameters such as WIP (see Figure 5), specific production characteristics (bottlenecks) or company-specific phenomena that need to be taken into account and are not available in the raw data in a suitable form. This may have already happened or been uncovered during the previous exploratory analysis.

The generalized construction of features can be done by standardized frameworks (like scikit learn [45]), or directly by AutoML application. This comprises formatting steps such as transforming categorical features but also feature construction such as the generation of polynomial values. Figure 6 shows a typical AutoML process using the example of Tree-based Pipeline Optimization Tool (TPOT) [15].

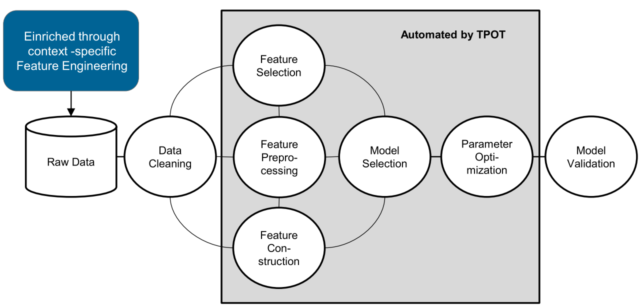

FIGURE 6. AutoML pipeline (based on TPOT-Framework [15]).

As shown in Figure 6, the process of feature construction and pre-processing is automated by TPOT. Users are required to pass a clean (and at best enriched with context or domain-specific features) dataset to the framework.

2) FEATURE SELECTION

Feature selection describes the tasks of the selection of a subset of features that lead to the best prediction result [31]. There are various methods for this, such as filters, wrappers and embedded methods. Filters are based on a relevance index using correlation coefficients, while wrappers use a learning algorithm to select the best subset of features. Both methods use search strategies to find the best solution. Embedded methods are a class characterized by their capability to include the generation of optimal feature subsets in the learning algorithm itself [31]. As can be seen in Figure 6, this is a task automated by AutoML.

3) DATA CLEANING

This step of data preparation aims to improve the quality of the dataset. This includes matters such as the selection of clean subsets of the data, handling noise, or dealing with special and/ missing values [37], [38].

Especially the handling of outliers should also be considered. Outliers are data points that are different from the norm, also known as abnormalities or deviants. Outliers can be detected using (simple) statistics such as the Tukey Fence [46], outlier algorithms (e.g. density-based) or using expert respectively domain knowledge [47]. Only a combination of these techniques and interaction with end users will produce good results [47].

In order to apply these findings to TTP predictions, it is necessary to identify which orders in principle show abnormal behaviour and should be excluded from the data set. The classification in Figure 7 shows an example of how this can be implemented. The aim is to remove orders that result from data errors or unclear demand (Type D) and are not representative in terms of the throughput behavior of an area or work system.

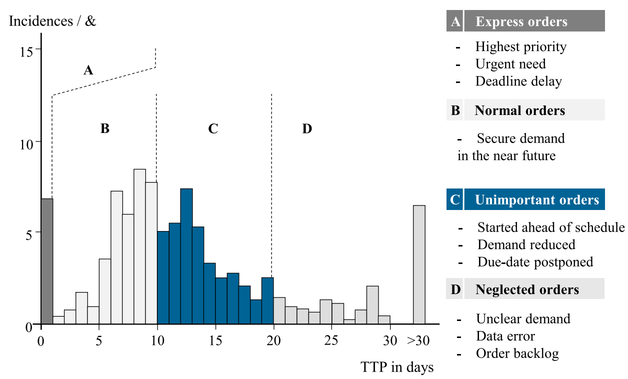

FIGURE 7. TTP distribution of a work system (based on [48]).

As the example in Figure 8 shows, it would be advisable to remove orders with a TTP of more than 30 days, which represent about 5 % of all orders. However, the removal of outliers remains very context-specific.

C. MODELLING (STEP IV)

Usually, these steps include the task of selecting a model, generating a test design as well as building and assessing the model (cf. [37]). As we suggest the use of AutoML frameworks for the prediction of TTP, the extent of these steps is noticeably reduced. Instead of choosing a ML technique or selecting a specific model [37], [38], we choose a AutoML framework that automates the selection process (cf. Figure 6). The following steps are similar to the modelling step without AutoML and include setting up a test design (splitting test, training and/or validation data) and monitoring the training results.

D. EVALUATION (STEP V)

The evaluation includes an overall *assessment of the results* (D.1), an *analysis of the explainability* (D.2) of the results and a *comparison with previously defined success criteria* (D.3) [37], [38].

1) ASSESSMENT OF THE RESULTS

At first, the evaluation of the results includes the evaluation of the ML model's performance. To investigate the performance of the ML model, different metrics such as MAE or MSE (see Section II-C) can be used. Further, the robustness of the model needs to be determined. For this purpose, it is possible to add perturbing inputs to the model to test the generalization ability. There are also robustness validation methods that use cross-validation [38]. Regarding the prediction of TTP, the MAE has the same dimension as the measured variable, i.e. typically days. Thus, it indicates the absolute deviation averaged over the calculated test sample. Therefore, a MAE of zero indicates an (quite unlikely) always accurate prediction. Higher values indicate a greater deviation between predicted and actual values.

2) ANALYSIS OF THE EXPLAINABILITY

Explainability of ML-models helps to understand results and gain trust [49]. It also provides further insights to improve model performance. According to [38], it is best practice to observe how the individual features of a model influence the outcome and to check whether it is plausible or not.

There are a variety of different approaches that support the understanding of the results of ML applications, referred to as Explainable AI (XAI) (cf. [50], [51]). XAI can be divided into agnostic and model-specific approaches [52]. While model-specific frameworks are only suitable for specific model types, agnostic models can be applied to different model types. Generally, XAI helps to explain local phenomena or determine the influence of certain features on the result [52]. A common element of these approaches is that the analyses are visual and easy to understand [52], [53]. A well-known approach is SHAP (SHapley Additive exPlanations) [53], [54]. It is based on the interpretation of Shapley values, a technique from cooperative game theory. To calculate SHAP values, subsets are formed from the total quantity of all features. Then, for each subset, the model output is calculated once with the feature in question and once without [54], [55]. Applying this to the TTP prediction, the influence of each feature on the final outcome of the prediction can be viewed. This allows to check whether the previously assumed influences can be generalized by the model, whether there are systematic errors or whether the model has uncovered previously unknown correlations.

3) COMPARISON OF SUCCESS CRITERIA AND NEXT STEPS

Finally, model and domain experts assess the usability of the model. In order to determine success, a comparison with the previously defined success criteria in terms of the ML model and the underlying business process is required. If the criteria are not met, it is necessary to go back to earlier phases of the CRISP-DM [38].

E. INTERIM CONCLUSION

In this section, along the different steps of the CRISP-DM, we have elaborated upon general matters that need to be considered when making TTP predictions.

Given that the prediction of TTP is not a new issue, it is important to note that little attention has yet been paid to the order fulfilment phase and related PPC tasks. In this context, particular attention should be paid to the definition of the scope and the available data. We also discuss how to select data based on driver trees, the benefits of AutoML, and provide guidance on how to remove outliers. Our procedures allow both theorists and practitioners to follow guidelines for constructing adequate TTP predictions.

V. CASE STUDY

To illustrate the use of TTP predictions along the order fulfilment process and highlight differences between the necessary PPC tasks, a case study is presented below. The investigated company, with approximately 300 employees, manufactures printed circuit boards of varying complexity. Products are made-to-order and according to customer specifications. The company's philosophy is to compete through high logistical performance in terms of high schedule compliance and short TTP. The case study's presentation follows the course of the CRISP-DM.

A. BUSINESS & DATA UNDERSTANDING

1) SCOPE AND SUCCESS CRITERIA

We intend to predict the planning values for TTP for each relevant task within PPC, i.e.: $TTP^{Order,rough-plan}$ in the context of order clarification, $TTP^{Order,detail-plan}$ for throughput scheduling and $TTP^{Order,remain-plan}$ for order coordination. The predicted share of time is the same each case, i.e. the time from release to completion (cf. formula 1). However, the time perspective and the target of the prediction change. While in order clarification, the TTP must be determined in order to calculate a delivery date, in throughput scheduling, the TTP defines the detailed start and finish of an order and its associated operations. In the context of order coordination, the remaining TTP of an order is used to communicate delays to the customer. We define the point in time for the prediction of the remaining TTP directly after order release. Thereby, it easier to compare results as if predicting the remaining TTP at a particular process step.

We do not consider a specific production subarea but the entire manufacturing process. The overall objective is to calculate more accurate planned TTP specified for the relevant PPC tasks.

2) DATA COLLECTION

The job shop production process contains 33 workstations, including order release, packing & shipping. During coating, a bottleneck can be observed. Available data comprises master data such as work schedules (e.g., operation processing and setup times) and material specifications. In addition, planning data is provided that includes a target end date of an order as well as the currently used planned TTP. Further, each order is assigned to a priority tag, that characterizes an express or less important order. Also, production feedback data is available, which includes actual start and end times of orders as well as individual operations. The time specifications are given in shop calendar days [SCD], that comprise only business working days. As an initial assumption, the following features were considered relevant:

- Actual Order TTP (the target variable) [SCD]
- Product group [-]
- No. of process steps [-]
- Target end date [SCD]
- Planned Order TTP [SCD]
- Lot size [-]
- Processing times [hours]
- Priority [-]
- Release date [SCD]
- Input schedule deviation (Deviation from planned release) (Input schedule deviation) [SCD]

The data collected covers a period of 220 SCD and includes 3004 production orders. More features are constructed in Section B.

3) DATA EXPLORATION AND ASSESSMENT OF INITIAL SITUATION

Essentially, three aspects will be analyzed in this section: the current situation regarding the achievement of the set planned TTP, a univariate analysis of the target variable (Order TTP) and a bivariate analysis to examine the relationship of the overall data with our target variable.

The average planned TTP is 15,2 SCD, with a minimum of 1 and a maximum of 26 SCD. The standard deviation is 6,2 SCD. The average value for the actual Order TTP is 18,1 SCD and thus approx. 3 SCD higher than the planned TTP. The standard deviation is 14,4 SCD (also quite higher than the planned TTP). The MAE of the used planned TTP is 7,4 SCD while the MSE is 184,0 SCD.

A correlation analysis is performed to examine the deviation of actual and planned TTP in more detail, also referred to as the relative schedule deviation (cf. [2]). We use the Spearman correlation coefficient (r_s) and search for attributes with a strong influence. The limit that differentiate an influence to be considered as strong varies depending on the discipline, but in the following, it is considered to be 0,5 or higher (cf. [56]). Only one feature that has an impact of 0,5 or higher could be identified: the deviation from the planned release date, also referred to as input schedule deviation ($r_s = -0{,}77$). Thus, a strong relation between a delayed release and the deviation of actual and planned TTP can be suspected.

Figure 8 visualizes this relationship. Obviously, orders with a high input delay are accelerated and completed faster than planned, and orders with a too-early input are completed later. This indicates a deadline-oriented fulfilment of the orders (cf. [57]).

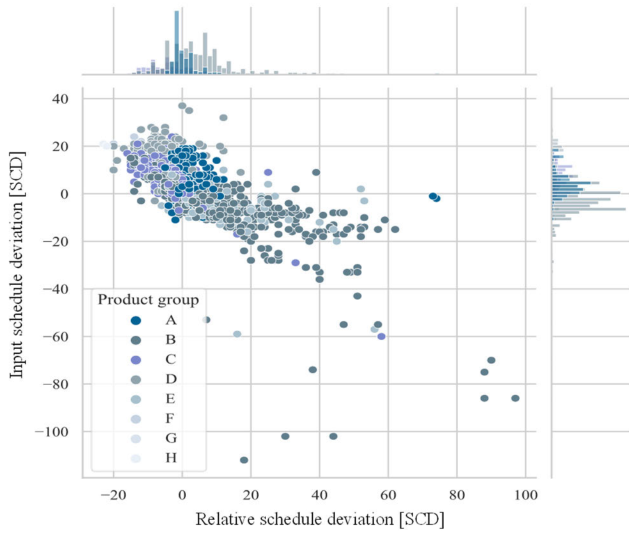

FIGURE 8. Correlation of input and schedule deviation clustered by product groups.

Further on, a univariate analysis of our target variable, the actual Order TTP is advisable. The distribution can be seen in the figure 9. The minimum TTP is 1 day while the maximum observed TTP in the data set is 202 SCD. It can be speculated that this could be a data error.

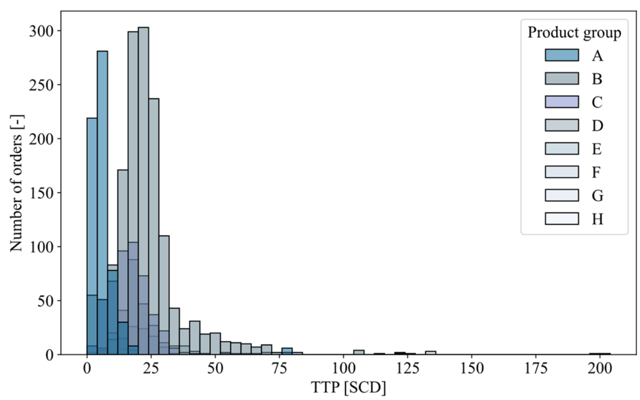

FIGURE 9. Distribution of Order TTP clustered with product groups.

To understand the relationship of the target variable to the other features, we use a scatter matrix plot. To keep it

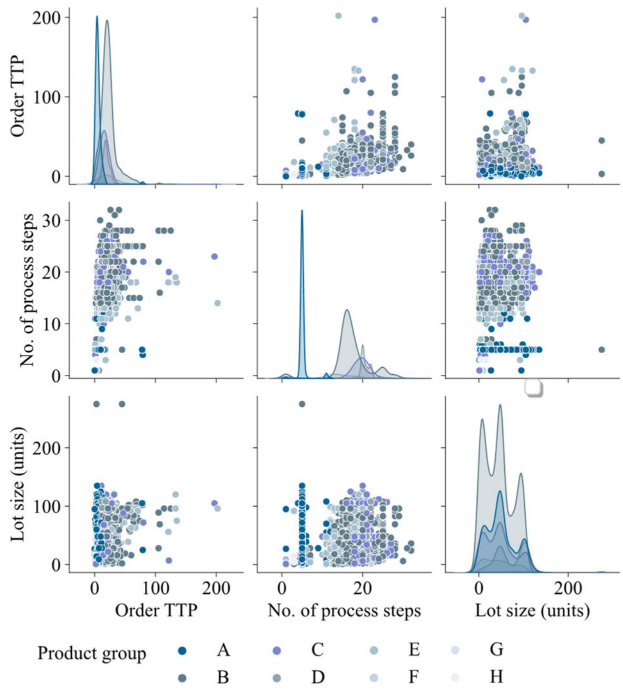

FIGURE 10. Scatter matrix plot.

compact, we have used only two other features that generally could have a high influence on actual Order TTP (cf. [42]), namely lot size and number of process steps).

While there is a strong correlation between Order TTP and the number of process steps ($r_s = 0{,}56$), there is no clear correlation between lot size and Order TTP ($r_s = 0{,}24$). Further features with a strong influence are the processing time ($r_s = 0{,}57$) and the input schedule deviation ($r_s = -0{,}60$). The latter confirms the relationship suspected earlier.

4) DATA QUALITY VERIFICATION

We consider the data quality to be sufficient to fulfil the task.

B. DATA PREPARATION AND MODELLING

We use AutoML as our prediction tool. As Framework TPOT is applied (cf. [15]), which automates the data preparation (pre-processing, feature construction & selection) and modelling as illustrated in Figure 6. As described before, we distinguish between two forms of feature engineering: the context-specific engineering of features, and the more technical, generalized construction of features, done automatically by TPOT.

After TPOT has selected suitable models for the prediction, we apply a similar model type as a benchmark to compare the results.

1) FEATURE ENGINEERING

In addition to Step A.2, we compute two additional features before cleaning that are not included in the raw data. These are two parameters for WIP. The WIP is an essential parameter influencing the throughput in production. The calculation follows the formula below [3]:

$$\mathrm{WIP}_t = \int_{T=t_0}^{T=t} \mathrm{IN}\,(T)\,dT - \int_{T=t_0}^{T=t} \mathrm{OUT}\,(T)\,dT \qquad (12)$$

where t denotes the current time, t_0 the start time and IN or OUT the input/output of the system, which can be expressed in working hours or the number of jobs. We calculate values for the WIP at a time t, which is the time of prediction and thus reflects the current system load, for the whole production area (denoted as WIP_{PA}) as well as for the bottleneck (denoted as WIP_{BN}). In conclusion, the total number of available features depending on the corresponding PPC task can be seen in Table 2.

TABLE 2. Available features depending PPC tasks.

Feature	Order clarification	Throughput scheduling	Order co-ordination
Physical factors			
No. of process steps	x	x	x
Product group	x	x	x
Factors influencing throughput			
Processing time		x	x
Lot size		x	x
WIP_{PA}			x
WIP_{BN}			x
Factors influencing plan deviations			
Priority		x	x
Target end		x	x
Input schedule deviation			x
Release date			x

The last step of feature engineering contains the manual transformation of categorical features using one-hot encoding.

2) FEATURE SELECTION

TPOT automates the selection of the best subset of available features.

3) DATA CLEANING

According to the classification of orders discussed earlier (see Figure 7), we decide to exclude orders with an actual Order TTP higher than 100 SCD. This means that we eliminate just 14 rows. Also, 1 row is eliminated due to a missing value in the priority column. In total, the cleaned data set contains 2989 rows.

C. MODELLING

We create a training and a test dataset by splitting the data in a ratio of 80/20. The training dataset is used to train and optimize the applied models with TPOT, therefore, further splitting into a validation dataset is not essential for our application.

We create in total three models for predicting $TTP^{Order,rough-plan}$ in the context of order clarification, $TTP^{Order,detail-plan}$ for throughput scheduling and $TTP^{Order,remain-plan}$ for order coordination. We have specified the scoring parameter to be the MAE. TPOT provides a wide range of information about the process of creating the model. We limit ourselves in stating which model was ultimately selected as the most suitable. For predicting $TTP^{Order,rough-plan}$, Gradient Boosting Regressor was selected. For $TTP^{Order,detail-plan}$ as well as $TTP^{Order,remain-plan}$ Extra Trees Regressor was selected.

As all models are tree-based, therefore we use LightGBM (see [58]), a tree-based gradient boosting framework as the benchmark for AutoML.

D. EVALUATION

1) ASSESSMENT OF THE RESULTS

The following table summarizes the results, showing the prediction performance in line with the specified metrics for the three models depending on the PPC tasks computed using TPOT and LightGBM.

As shown in Table 3, each model is able to predict Order TTP with a MAE of less than 6 SCD. It is noteworthy how MAE and MSE improve as order fulfilment progresses. This demonstrates the impact of the additional information in the different PPC tasks. TPOT only slightly outperforms LightGBM. However, it is important to state that we could only select this model type based on the TPOT results, since we saw that obviously tree-based algorithms might be suitable. The influence of TPOT's built-in feature construction and selection tools appears small. However, it should be noted that our raw data set is not very extensive. For larger and more complex data sets, it is likely possible that the results will be different.

TABLE 3. Metric results.

	MAE TPOT	MAE Light GMB	MSE TPOT	MSE Light GBM
Order clarification	5,93 [SCD]	6,09 [SCD]	82,43 [SCD]	82,71 [SCD]
Throughput scheduling	4,63 [SCD]	4,82 [SCD]	53,58 [SCD]	56,25 [SCD]
Order coordinatin	3,58 [SCD]	3,61 [SCD]	24,82 [SCD]	28,06 [SCD]

2) ANALYSIS OF THE EXPLAINABILITY

To analyze the explainability models applied, we use SHAP. Because of the ease of use and the similarity of the results, we do this based on the prediction results of the LightGBM framework. Using the SHAP framework allows us to identify the features that most influence the outcome. To do this, we will first name the top 3 features for each model and then take a closer look at the prediction within order coordination (since this has access to the most information).

The mean absolute SHAP value shows us how much a single feature influenced the prediction. For order clarification, the feature number of process steps is by far the most influential. Further, certain product groups also influence the prediction. An interesting point to note here is that some product groups had no impact.

In the case of throughput scheduling, the priority tag N was the most important feature. Priority tag N indicates "not important" orders. Thus, we assume the model was able to process this type of input correctly. The number of process steps is also a significant feature as well as the process time. Both are to be expected according to the general influencing factors (cf. Figure 5).

TABLE 4. Features sorted by (mean absolute) SHAP value.

Importance	Order clarification	Throughput scheduling	Order co-ordination
#1	No. of process steps: 4,1	Priority_N: 3,1	Input schedule deviation: 4,5
#2	Product group_B: 1,2	No. of process steps: 3,1	No. of process steps: 4,0
#3	Product group_C: 1,0	Process time: 1,4	Release: 1,8

For order coordination, the most important feature processed is the input schedule deviation. Due to the fact that late orders are accelerated (cf. Figure 8), we assume that this feature is also adequately handled by the model. Based on the information described above, we now know which features have a strong influence on the model result, but we do not yet know how this is expressed in detail (positive or negative). To illustrate how certain features impact the prediction outcome in detail, we can use local explanations. Figure 11 shows a waterfall diagram depicting the influence of each feature of a single observation in the test set.

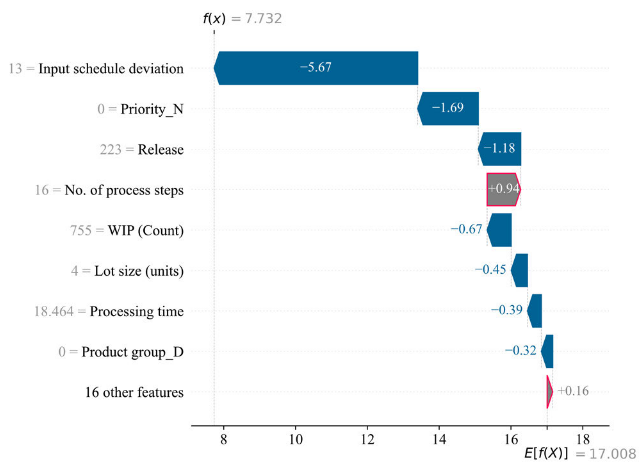

FIGURE 11. Waterfall plot for visualizing the influence of certain features on prediction outcome.

E[f(x)] corresponds to the expected value of our model function, while f(x) denotes the prediction value for that observation. Thus, the sum of the SHAP values corresponds to the difference between the expected value and the actual, predicted value. With an expected value of the model function of approx. 17 SCD, the information that the order's input schedule deviation equals 13 SCD causes the prediction results to be reduced by 5.67 SCD. The priority tag N = 0 indicates that this order is not an unimportant order (1 would indicate it as unimportant). This reduces the model result by 1.69 SCD. This confirms the above-mentioned assumption that the priority tag N is correctly processed by the model. The number of process steps in this example amounts to 16 (hence the order passes 16 workstations), which increases the prediction result by 0.94 SCD.

In terms of analyzing the model's the model's explainability, it becomes clear that only a few features have a strong influence, while many other features have only a small influence on the result. As described in this section, from a domain and process perspective, these results are plausible.

3) COMPARISON OF SUCCESS CRITERIA AND NEXT STEPS

Finally, the degree to which the prediction results are able to produce a more accurate Order TTP in comparison with the planned Order TTP is evaluated. As mentioned above, we use the internally planned TTP for evaluation. Using formula 5, we also computed two comparative models values for planned TTP. They differ according to whether they are calculated globally for all orders (1) or whether they are calculated specifically for each product group (2).

TABLE 5. Metrics of planning data (compared to Order TTP) and Comparison to ML results using TPOT.

Comparative Model	MAE / MSE	Improvement in comparison with TPOT model		
		Order clarification	Throughput scheduling	Order coordination
Company's Planned TTP	7,36 / 184,0	19,4% / 55,2%	37,1% / 70,9%	51,4% / 86,5%
(1)	9,30 / 208,7	36,2% / 60,5%	50,2% / 74,3%	61,5% / 88,1%
(2)	7,16 / 164,3	17,2% / 49,8%	35,3% / 67,4%	50,0% / 84,9%

It is evident that the planned values determined by ML are significantly more accurate than the planned values used by the company and those calculated for further comparison. The additional information generated during the course of order fulfilment considerably influences the prediction quality. Yet even at the beginning of order fulfilment, ML can already help to improve the estimation of planned Order TTP.

However, it is crucial to emphasize that the comparison with "historical" planned values is lagging. In a real-world environment, the value of actual order TTP (that we use to calculate the above metrics) is the result of planning, implementation, and execution. If we retrospectively just apply other plan values, we neglect the fact that one is related to the other. Consequently, the appropriateness of ML-based planned values for the order TTP would need to be empirically tested and then compared to prior planning accuracy. But for a first assessment it is still helpful.

VI. CONCLUSION

In the course of this paper, the problem, or better said, the potential, of looking more closely at the order fulfillment phase when predicting TTP is discussed. Based on the CRISP-DM, it is shown which aspects have to be considered in the different PPC tasks. To support the development of ML applications, different approaches, e.g. for data cleaning or selection, are presented. A key aspect is the availability of diverse information at different times along the order fulfilment process. In a subsequent case study, we showed how to predict TTP along order fulfillment using AutoML and different ML models. Using XAI, we reviewed and validated the models' results. Our results support that ML can be used to improve estimating accurate planned TTP. More accurate predictions are possible by taking into account the different specifics of the various phases of order fulfillment and the peculiarities of each PPC task. Prediction quality improves along the order fulfilment process.

The deployment of predicted TTP in PPC planning tasks has been little explored until now (cf. [24]). Thus, further research should focus more on the integration of ML-based predictions into PPC, taking into account the different aspects of different tasks, and addressing important logistics issues such as lead-time syndrome. Moreover, it is crucial evaluate the overall logistical improvements that can be achieved through ML-based predictions, especially in terms of on-time delivery. This clarification is important in order to determine the concrete benefits for the companies.

REFERENCES

[1] M. Schmidt, J. T. Maier, and L. Härtel, "Data based root cause analysis for improving logistic key performance indicators of a company's internal supply chain," *Proc. CIRP*, vol. 86, pp. 276–281, 2019, doi: 10.1016/j.procir.2020.01.023.

[2] R. Handfield, F. Straube, H.-C. Pfohl, and C. Wieland, *Trends and Strategies in Logistics and Supply Chain Management: Embracing Global Logistics Complexity to Drive Market Advantage*. Hamburg, Germany: DVV Media Group, 2013.

[3] P. Nyhuis and H.-P. Wiendahl, *Fundamentals of Production Logistics*. Berlin, Germany: Springer, 2009.

[4] A. Mütze, T. Lucht, and P. Nyhuis, "Logistics-oriented production configuration using the example of MRO service providers," *IEEE Access*, vol. 10, pp. 20328–20344, 2022, doi: 10.1109/ACCESS.2022.3146420.

[5] M. Schmidt and P. Schäfers, "The Hanoverian supply chain model: Modelling the impact of production planning and control on a supply chain's logistic objectives," *Prod. Eng.*, vol. 11, nos. 4–5, pp. 487–493, Oct. 2017, doi: 10.1007/s11740-017-0740-9.

[6] T. Hiller, L. Deipenwisch, and P. Nyhuis, "Systemising data-driven methods for predicting throughput time within production planning & control," in *Proc. IEEE Int. Conf. Ind. Eng. Eng. Manage. (IEEM)*, Kuala Lumpur, Malaysia, Dec. 2022, pp. 0716–0721.

[7] D. Mourtzis, M. Doukas, K. Fragou, K. Efthymiou, and V. Matzorou, "Knowledge-based estimation of manufacturing lead time for complex engineered-to-order products," *Proc. CIRP*, vol. 17, pp. 499–504, 2014, doi: 10.1016/j.procir.2014.01.087.

[8] A. Baykasoğlu, M. Göçken, and Z. D. Unutmaz, "New approaches to due date assignment in job shops," *Eur. J. Oper. Res.*, vol. 187, no. 1, pp. 31–45, May 2008, doi: 10.1016/j.ejor.2007.02.020.

[9] H.-P. Wiendahl and H.-H. Wiendahl, *Betriebsorganisation für Ingenieure*, 9th ed. München, Germany: Hanser, 2019.

[10] G. Schuh, A. Gützlaff, F. Sauermann, O. Kaul, and N. Klein, "Databased prediction and planning of order-specific transition times," *Proc. CIRP*, vol. 93, pp. 885–890, 2020, doi: 10.1016/j.procir.2020.04.026.

[11] K. J. Kramer, C. Wagner, and M. Schmidt, "Machine learning-supported planning of lead times in job shop manufacturing," in *Advances in Production Management Systems. The Path to Digital Transformation and Innovation of Production Management Systems* (IFIP Advances in Information and Communication Technology), B. Lalic, V. Majstorovic, U. Marjanovic, G. v. Cieminski, and D. Romero, Eds. Cham, Switzerland: Springer, 2020, pp. 363–370.

[12] J. T. Maier, A. Rokoss, T. Green, N. Brkovic, and M. Schmidt, "A systematic literature review of machine learning approaches for the prediction of delivery dates," in *Proc. Conf. Prod. Syst. Logistics*. Hannover, Germany: Publish-Ing, 2022, pp. 121–130.

[13] M. Feurer, K. Eggensperger, S. Falkner, M. Lindauer, and F. Hutter, "Auto-Sklearn 2.0: Hands-free AutoML via meta-learning," 2020, *arXiv:2007.04074*.

[14] H. Jin, F. Chollet, Q. Song, and X. Hu, "AutoKeras: An AutoML library for deep learning," *J. Mach. Learn. Res.*, vol. 24, no. 6, pp. 1–6, 2023. [Online]. Available: httpjmlr.orgpapersv2420-1355.html

[15] R. S. Olson, N. Bartley, R. J. Urbanowicz, and J. H. Moore, "Evaluation of a tree-based pipeline optimization tool for automating data science," in *Proc. Genetic Evol. Comput. Conf.*, Denver, CO, USA, Jul. 2016, pp. 485–492.

[16] K. L. Croxton, "The order fulfillment process," *Int. J. Logistics Manage.*, vol. 14, no. 1, pp. 19–32, Jan. 2003.

[17] M. Schmidt and P. Schäfers. *Hannoveraner Lieferkettenmodell*. Accessed: Oct. 17, 2023. [Online]. Available: https://halimo.education/

[18] B. MacCarthy, P. G. Brabazon, and J. Bramham, "Fundamental modes of operation for mass customization," *Int. J. Prod. Econ.*, vol. 85, no. 3, pp. 289–304, Sep. 2003, doi: 10.1016/S0925-5273(03)00117-8.

[19] H. R'bigui and C. Cho, "Customer Oder fulfillment process analysis with process mining," in *Proc. Int. Conf. Comput. Sci. Artif. Intell.*, Jakarta, Indonesia, Dec. 2017, pp. 247–252.

[20] M. Schmidt and P. Nyhuis, *Produktionsplanung Und-Steuerung im Hannoveraner Lieferkettenmodell*. Berlin, Germany: Springer, 2021.

[21] H. Lödding, *Handbook of Manufacturing Control*. Berlin, Germany: Springer, 2013.

[22] F. R. Jacobs, W. L. Berry, D. C. Whybark, and T. E. Vollmann, *Manufacturing Planning and Control for Supply Chain Management: The CPIM Reference*, 2nd ed. New York, NY, USA: McGraw-Hill, 2018.

[23] M. L. Pinedo, *Scheduling: Theory, Algorithms, and Systems*, 5th ed. Cham, Switzerland: Springer, 2016.

[24] M. Schneckenreither, S. Haeussler, and C. Gerhold, "Order release planning with predictive lead times: A machine learning approach," *Int. J. Prod. Res.*, vol. 59, no. 11, pp. 3285–3303, Jun. 2021, doi: 10.1080/00207543.2020.1859634.

[25] G. Schuh, G. Reinhart, J.-P. Prote, F. Sauermann, J. Horsthofer, F. Oppolzer, and D. Knoll, "Data mining definitions and applications for the management of production complexity," *Proc. CIRP*, vol. 81, pp. 874–879, 2019, doi: 10.1016/j.procir.2019.03.217.

[26] M. Mohri, A. Rostamizadeh, and A. Talwalkar, *Foundations of Machine Learning*. Cambridge, MA, USA: MIT Press, 2018.

[27] J. P. U. Cadavid, S. Lamouri, B. Grabot, R. Pellerin, and A. Fortin, "Machine learning applied in production planning and control: A state-of-the-art in the era of Industry 4.0," *J. Intell. Manuf.*, vol. 31, no. 6, pp. 1531–1558, Aug. 2020, doi: 10.1007/s10845-019-01531-7.

[28] K. J. Kramer, A. Rokoss, and M. Schmidt, "Do we really know the benefit of machine learning in production planning and control? A systematic review of industry case studies," in *Proc. Conf. Prod. Syst. Logistics*. Hanover, Germany: Publish-Ing, 2021, pp. 223–233.

[29] N. Amruthnath and T. Gupta, "A research study on unsupervised machine learning algorithms for early fault detection in predictive maintenance," in *Proc. 5th Int. Conf. Ind. Eng. Appl. (ICIEA)*, Singapore, Apr. 2018, pp. 355–361.

[30] T. Altenmüller, T. Stüker, B. Waschneck, A. Kuhnle, and G. Lanza, "Reinforcement learning for an intelligent and autonomous production control of complex job-shops under time constraints," *Prod. Eng.*, vol. 14, no. 3, pp. 319–328, Jun. 2020, doi: 10.1007/s11740-020-00967-8.

[31] I. Guyon, M. Nikravesh, S. Gunn, and L. A. Zadeh, *Feature Extraction*. Berlin, Germany: Springer, 2006.

[32] A. Zheng and A. Casari, *Feature Engineering for Machine Learning: Principles and Techniques for Data Scientists*. Sebastopol, CA, USA: O'Reilly, 2018.

[33] A. Saxena, J. Celaya, E. Balaban, K. Goebel, B. Saha, S. Saha, and M. Schwabacher, "Metrics for evaluating performance of prognostic techniques," in *Proc. Int. Conf. Prognostics Health Manage.*, Denver, CO, USA, Oct. 2008, pp. 1–17.

[34] R. Barbudo, S. Ventura, and J. R. Romero, "Eight years of AutoML: Categorisation, review and trends," *Knowl. Inf. Syst.*, vol. 65, no. 12, pp. 5097–5149, Dec. 2023, doi: 10.1007/s10115-023-01935-1.

[35] M. Schmidt and P. Schäfers. *Hannoveraner Lieferkettenmodell*. Accessed: Jun. 27, 2023. [Online]. Available: https://halimo.education/

[36] R. W. Conway, L. W. Miller, and W. L. Maxwell, *Theory of Scheduling*. Reading, MA, USA: Addison-Wesley, 1967.

[37] P. Chapman, *Crisp-Dm 1.0: Step-By-Step Data Mining Guide*, vol. 78. Chicago, IL, USA: SPSS, 2000.

[38] S. Studer, T. B. Bui, C. Drescher, A. Hanuschkin, L. Winkler, S. Peters, and K.-R. Mueller, "Towards CRISP-ML(Q): A machine learning process model with quality assurance methodology," 2020, *arXiv:2003.05155*.

[39] S. Amershi, A. Begel, C. Bird, R. DeLine, H. Gall, E. Kamar, N. Nagappan, B. Nushi, and T. Zimmermann, "Software engineering for machine learning: A case study," in *Proc. IEEE/ACM 41st Int. Conf. Softw. Eng., Softw. Eng. Pract. (ICSE-SEIP)*, Montreal, QC, Canada, May 2019, pp. 291–300.

[40] E. Pelikán, "Principles of forecasting—A short overview," in *Theory and Practice of Informatics* (Lecture Notes in Computer Science), G. Goos, J. Hartmanis, J. van Leeuwen, J. Pavelka, G. Tel, and M. Bartošek, Eds. Berlin, Germany: Springer, 1999, pp. 311–327.

[41] M. Kuhn and K. Johnson, *Applied Predictive Modeling*. New York, NY, USA: Springer, 2013.

[42] L. Härtel and P. Nyhuis, "Data-based identification of throughput time potentials in production departments," in *Proc. Conf. Prod. Syst. Logistics*. Hanover, Germany: Publish-Ing, 2020, pp. 239–248.

[43] L. Haertel, "Modellbasierte datenanalyse zur identifikation logistischer schwachstellen in produktionsbereichen," Ph.D. dissertation, Leibniz Universität Hannover, Institu für Fabrikanlagen und Logistik, Hanover, Germany, 2021.

[44] S. Morgenthaler, "Exploratory data analysis," *WIREs Comput. Statist.*, vol. 1, no. 1, pp. 33–44, Jul. 2009, doi: 10.1002/wics.2.

[45] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and É. Duchesnay, "Scikit-Learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, Jan. 2011.

[46] D. C. Hoaglin, "John W. Tukey and data analysis," *Stat. Sci.*, vol. 18, no. 3, pp. 311–318, Aug. 2003, doi: 10.1214/ss/1076102418.

[47] J. H. Sejr and A. Schneider-Kamp, "Explainable outlier detection: What, for whom and why?" *Mach. Learn. Appl.*, vol. 6, Dec. 2021, Art. no. 100172, doi: 10.1016/j.mlwa.2021.100172.

[48] H.-P. Wiendahl, *Load-Oriented Manufacturing Control*. Berlin, Germany: Springer, 1995.

[49] D. Shin, "The effects of explainability and causability on perception, trust, and acceptance: Implications for explainable AI," *Int. J. Hum. Comput. Stud.*, vol. 146, Feb. 2021, Art. no. 102551, doi: 10.1016/j.ijhcs.2020.102551.

[50] I. Ahmed, G. Jeon, and F. Piccialli, "From artificial intelligence to explainable artificial intelligence in Industry 4.0: A survey on what, how, and where," *IEEE Trans. Ind. Informat.*, vol. 18, no. 8, pp. 5031–5042, Aug. 2022, doi: 10.1109/TII.2022.3146552.

[51] D. Gunning, M. Stefik, J. Choi, T. Miller, S. Stumpf, and G.-Z. Yang, "XAI—Explainable artificial intelligence," *Sci. Robot.*, vol. 4, no. 37, Dec. 2019, doi: 10.1126/scirobotics.aay7120.

[52] G. Sofianidis, J. M. Rožanec, D. Mladenić, and D. Kyriazis, "A review of explainable artificial intelligence in manufacturing," 2021, *arXiv:2107.02295*.

[53] J. Gerlings, A. Shollo, and I. Constantiou, "Reviewing the need for explainable artificial intelligence (xAI)," in *Proc. 56th Annu. Hawaii Int. Conf. Syst. Sci.*, Jan. 2023, pp. 1284–1293.

[54] S. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," 2017, *arXiv:1705.07874*.
[55] S. M. Lundberg, G. Erion, H. Chen, A. DeGrave, J. M. Prutkin, B. Nair, R. Katz, J. Himmelfarb, N. Bansal, and S.-I. Lee, "From local explanations to global understanding with explainable AI for trees," *Nature Mach. Intell.*, vol. 2, no. 1, pp. 56–67, Jan. 2020, doi: 10.1038/s42256-019-0138-9.
[56] H. Akoglu, "User's guide to correlation coefficients," *Turkish J. Emergency Med.*, vol. 18, no. 3, pp. 91–93, Sep. 2018, doi: 10.1016/j.tjem.2018.08.001.
[57] A. Mütze and P. Nyhuis, "Deriving of sequencing strategies for multi-stage productions supported by logistic models and software tools," in *Proc. Conf. Prod. Syst. Logistics*. Hanover, Germany: Publish-Ing, 2020, pp. 198–208.
[58] *Light GMB: Microsoft*. Accessed: Nov. 21, 2023. [Online]. Available: https://github.com/Microsoft/LightGBM

TOBIAS HILLER was born in 1990. He received the M.Sc. degree in management with a focus on production and logistics from the Georg August University of Göttingen. Since 2019, he has been a Research Associate in production management with the Institute of Production Systems and Logistics (IFA), Leibniz University Hannover.

TABEA MARIE DEMKE was born in 1993. She received the M.Sc. degree in production and logistics from Leibniz University Hannover. Since 2021, she has been a Research Associate in production management with the Institute of Production Systems and Logistics (IFA), Leibniz University Hannover.

PETER NYHUIS was born in 1957. He received the Dr.-Ing. degree in mechanical engineering and the Habilitation degree from Leibniz University Hannover. After working as the Manager of Supply Chain Management in the Electronics and Mechanical Engineering Industry, he has been heading the Institute of Production Systems and Logistics, Leibniz University Hannover, since 2003.