

Article

Machine Learning in Small and Medium-Sized Enterprises, Methodology for the Estimation of the Production Time

Maria Urban ^{1,2,*} , František Koblasa ¹  and Radomír Mendřický ¹ 

¹ Department of Manufacturing Systems and Automation, Faculty of Mechanical Engineering, Technical University of Liberec, 461 17 Liberec, Czech Republic; frantisek.koblasa@tul.cz (F.K.); radomir.mendricky@tul.cz (R.M.)

² Department of Production Engineering, Faculty of Mechanical Engineering, Zittau/Görlitz University of Applied Sciences, 02763 Zittau, Germany

* Correspondence: m.urban@hszg.de

Abstract: Data mining (DM) and machine learning (ML) are widely used in production planning and scheduling. Their application to production time estimation leads to improved planning and scheduling accuracy, resulting in increased overall efficiency. Small and medium-sized enterprises (SMEs) often have a small amount of data, which results in the limited adoption of DM and ML. Instead, production time estimation is still performed using rough approximations, which are inaccurate and non-reproducible. Therefore, this article proposes an ML methodology for production time estimation. It is adapted to the needs of SMEs and is applied with limited data. The methodology is based on the categorization of four job types (from A to D), the partitioning of data according to the limit theorem of data convergence, and the definition of risk based on metrics of probability and statistics. ML was applied by WEKA Workbench (Waikato Environment for Knowledge Analysis). It is also integrated into the Cross Industry Standard Process for DM. The methodology was implemented on data from a medium-sized company, Schoepstal Maschinenbau GmbH, for job types A and B to estimate machine/job cycle time, manufacturing cycle time, and lead time. Different accuracies were obtained for individual estimation models, confirming the strong dependence of the models on data quality. Suitable models were found for the implementation of the estimation of the manufacturing cycle time and the machine/job cycle time. The modeling of lead time estimation was unsuccessful. This was due to the weak dependence between the learning values and the values of the selected model attributes. The implementation of the methodology for job types C and D is the subject of further research.

Keywords: data mining; machine learning; production time estimation methods; standard time; predictive analytics; small and medium-sized enterprises; modeling evaluation



Citation: Urban, M.; Koblasa, F.; Mendřický, R. Machine Learning in Small and Medium-Sized Enterprises, Methodology for the Estimation of the Production Time. *Appl. Sci.* **2024**, *14*, 8608. <https://doi.org/10.3390/app14198608>

Academic Editors: Christos Douligeris, Dimitris Apostolou, George A. Vouros and Themis Panayiotopoulos

Received: 14 August 2024

Revised: 11 September 2024

Accepted: 16 September 2024

Published: 24 September 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Systematic estimation of production time is essential for small and medium-sized enterprises (SMEs) to reach their full potential and maintain their market position. It enables efficient planning, resource optimization, and workforce management in manufacturing processes. The current progress of each production process can be monitored to identify problems or delays quickly. This allows for rapid response and corrective action to be taken. Overall visibility into the production process increases, reducing the risk of delivery delays.

Many methods for processing data and estimating production time were developed during the 20th century [1–4]. Two major trends in developing time estimation methods emerged in the 20th century. On the one hand, Predetermined Motion Time Systems (PMTSs) were grown. PMTSs are based on motion studies of manual work and time measurement units (TMUs). These methods are mainly developed under the Methods-Time Measurement Association, known as the Methods-Time Measurement (MTM). On the

other hand, time studies based on direct measurement of work time in production were developed. The flagship organization for this approach is the REFA (Reichsausschuß für Arbeitszeitermittlung, renamed to Verband für Arbeitsstudien und Betriebsorganisation e.V.) [3,4].

Production time estimation is based on motion data, time data, and specific data. These are current or historical data. Data are processed by PMTS or by mathematical methods, including descriptive analytics (DA), predictive analytics (PA), or simulation. DAs use descriptive statistics to describe the data they are [5]. The data from the technological systems are almost processed by the PAs [6]. PAs are statistical and probabilistic learning methods, fuzzy systems, optimization, topology, etc., applied by algorithms within Machine Learning (ML) [7].

1.1. Motion Data

The motion data are processed by the PMTS. A system of cameras is used to record movements. The records are used for estimating the time required for manual tasks [8,9], mainly for assembly activities [10,11], but also for the operating tools [12] and vehicles [13].

The MTM methods are widely used. In the MTM-1, basic movements are recorded for each hand, where eight basic hand and shoulder movements, two basic eye movements, and nine basic body and leg movements are defined. The smallest time unit is 1 TMU = 0.036 s. It is a very detailed method with a time consumption of 1:200, i.e., 200 min are used to analyze 1 min of movement [1]. MTM-2 does not differentiate between each hand, because a significant portion of the basic movements of the hands and shoulders are in combination. MTM-2 combines up to three basic movements. MTM-1 and MTM-2 are used in mass production. The methods used in serial production, the Universal Analyzer System (UAS), and for single and small series production (MEK) have been designed. UAS and MEK define nine standard actions consisting of a maximum of five basic movements. Contrary to MTM-1 and MTM-2, they are not based on knowledge of the sequence of movements but on the information of the general conditions, on the knowledge of the working system [2].

Maynard's Technique of Sequence of Operations (MOST) also uses TMU. In MOST, activities are described by a sequential model, a fixed string of letters. These are indexed. The index depends on the environmental conditions and expresses the time evaluation of the attribute. The sum of the indices is multiplied by the basic time unit to obtain the estimated time. The MOST methods are subdivided into MiniMOST for high-volume production, MaxiMOST for custom production, and BasicMOST for serial production with medium-job cycles. Approximately 10 h are required to analyze 1 h of motions [13].

The Modular Arrangement of Predetermined Time Standards (MODAPTS) is based on the body actions that follow each other [8,14]. Namely, they are the actions of the movements of the different parts of the upper limb, the terminal action, the action of receiving, the action of placing, and the action of auxiliary movements [15]. MODAPS is used to integrate correction factors for implementation into DELMIA (Digital Enterprise Lean Manufacturing Interactive Application) software with the purpose of virtual simulation [16,17].

1.2. Time Data

Time study is a technique used in research and practice to collect time data. Implementation in the automotive industry is common, e.g., [18–22]. The data are processed by mathematical methods. Tools used in time studies are stopwatches, forms and observation boards, video recordings, barcodes, radio frequency data communication (RFCD) and radio frequency identification (RFID), sensors, and cameras [23–28].

Virtual reality tools are also used to collect data. A virtual environment is a computer-generated interface that mimics reality. Workers interact using headsets, haptic gloves, motion trackers, and sensors [29]. Activity is tracked and recorded, and time is measured by a stopwatch or virtually. This application is mainly found for simple manual

assembly tasks [29–31]. Time data are processed by simulation [18], DAs [19,24], and PAs [20–23,25–28].

1.3. Specific Data

Specific data are data that describe a product, process, or production system.

The product-specific data are volume, weight, height, surface, design complexity, dimension of parts, interconnection of parts, shape design, and material [24,32,33], which are collected from CAD (Computer-Aided Design). The process-specific data are given by machining attributes such as tool position, machining speed, acceleration or deceleration [34], feed angle, the ratio of translational to rotational motion [35], and tool trajectory [36]. System-specific data include machine operating status, target completion time, average time to failure, repair time [37,38], system status, job type, actual flow time [39], or number of jobs [40]. Specific data are processed by software applications [24,32,33] and PAs [34–40].

1.4. Predictive Analytics and Data Mining in Production Time Estimation

PAs are powerful analytic techniques for dealing with data. The goal of the application of PAs is to learn from the data, to achieve the prediction of a certain phenomenon [7] by applying algorithms in ML. For production time estimation in research and industry PAs of Supervised Learning (SL), Unsupervised Learning (UL), Ensembled Learning (EL), and Deep Learning (DL) are applied. These are summarized in Figure 1. PAs are used in the sense of building a model for estimation that depends on attributes. Attributes are mostly described by time or specific data.

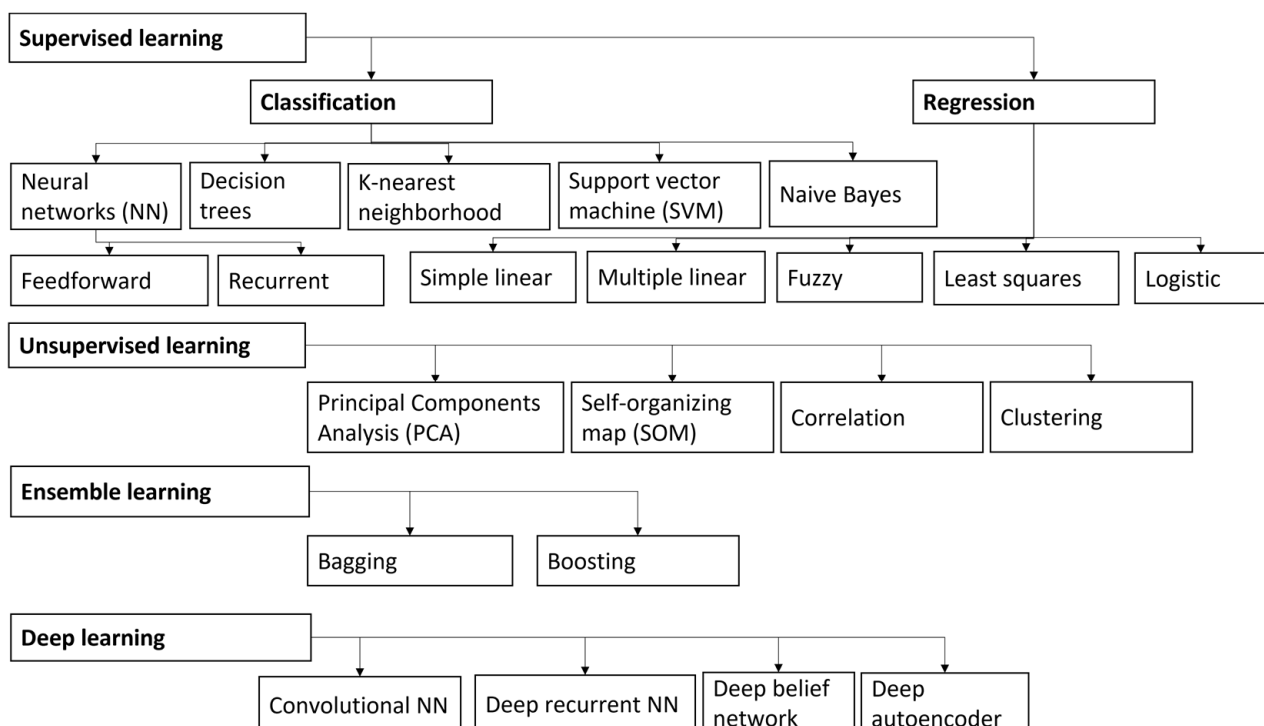


Figure 1. PA learning methods applied for estimating production time.

SL is divided into classification methods and regression. Classification models select the estimated value from training values [26,38,40] by defining constraints on the selection of the assumed nominal value. The predicted nominal values can be in the form of an interval or a single value. The inputs and outputs of classification algorithms are processed as discrete values.

Neural networks (NNs) are often used for classification, almost for modeling of cycle time (CT) and lead time (LT) estimation [38,41–43]. For the flow shop, feedforward NN

with attributes number of operations, product type, and queuing times [44] is applied. In addition, a recurrent NN with the attributes of machine operating status, processing time, target make span, machine parameters, and mean time to failure and repair [37] is presented. The modeling of NN is time-consuming, therefore NN models are used for time estimation of products with high repeatability.

Decision trees are applied to estimate CT [45] and LT [46]. These methods categorize estimated values into disjoint solution sets by constraining attribute values in accordance with the induction strategy [47,48].

The k-nearest neighbor method belongs to the subset of methods known as lazy learning [49]. It does not consider all values but rather recognizes values that have already appeared in the data on the basis of which other values are then classified. K-nearest neighbor is applied for CT estimation of wall panels in a flow shop production system [26].

The Support Vector Machine (SVM) method defines a linear decision boundary. SVM uses only a small subset of values that lie on the boundaries of the classes, and it identifies critical points that lie in different classes, called support vectors [49]. The method then defines their connecting line, where the boundary between classes is a perpendicular bisector passing through the midpoint of this line, known as the maximum margin hyperplane. SVM is presented for time estimation of manual activities [50,51].

Naive Bayes (Bayesian probability) is used when the attributes are independent and equally important. It is based on conditional probability. The algorithm selects the most frequent value and the most frequent values of the attributes associated with it [49]. Naive Bayes within job shop with CNC, turning, and milling for LT estimation is presented [52].

Regression models compute the time estimation value. The regression is a continuous function with coefficients corresponding to the weights of each attribute in the time estimation. The simplest form of regression is the simple linear regression model. It is an estimation based on one attribute. In multiple linear regression, the target value depends on multiple input attributes, and the problem is solved in vector space. CT models for machining [53] and assembly [22,54] are presented. The fuzzy regression model is based on uncertain data. This method finds application in the electronics industry, for the CT estimation of the production of wafers [55].

UL is the application of a set of tools used for data processing [56]. The purpose of UL is to find relationships in the data, discover subsets, or reduce the data [57], to prepare the data for the application of SL. Clustering algorithms are rules for the discovery of similar subsets of data [41,58]. Correlation analysis is a well-known method for identifying relationships. The Self-Organising Map (SOM, Kohonen map) and Principal Component Analysis (PCA) are techniques used for data reduction with minimal information loss. PCA selects appropriate combinations of the original attributes and then proposes these combinations as new attributes [8]. SOM is applied in NNs. The algorithm works by moving neurons in the direction of similar data [45,59].

EL is used to improve the predictive power of the data. Bagging is the process of dividing a dataset into equally sized subsets using sampling with replacement [49]. It is suitable for unstable data, where a small change in the input data significantly affects the model structure. Bagging is combined with NN for real-time estimation modeling [26,60]. Boosting is an iterative process where the current model is influenced by the previous model and by the data points that were misclassified [61]. Boosting is used with NN [62] and with the SVM [50] for LT prediction.

DL is a group of learning methods within neural networks that involve multiple layers [25]. Convolutional NNs are used for image recognition [63] when analyzing manual operations in videos. Deep Recurrent NNs are networks with feedback connections between neurons. They are used to recognize temporal dependencies in manufacturing to estimate LT based on the current state of production [37]. A Deep Belief Network is a stochastic NN based on Boltzmann machines. It is used to estimate LT while considering dependencies between data points [28]. A Deep Autoencoder attempts to reconstruct the input data by

transforming it into hidden layers and using this transformed data as input to the decoder. This technique is used to estimate the remaining time using RFID data [25].

1.5. Status of Production Time Estimation by Machine Learning Applications in SMEs

SMEs often specialize in the custom manufacturing of various products for different customers. The product specification is provided by technical drawings and, for complex products, by a bill of materials. Product types may be repetitive, or they may be one-off jobs. SMEs often have Enterprise Resource Planning (ERP) and shop floor data collection systems. However, these systems only serve as databases for corporate accounting purposes. Job planning is performed in the form of projects, and the scheduling of jobs to resources is often performed operationally. Production time estimation is often based on rough estimates using descriptive analytics or expert knowledge due to the high variability of products.

ML is part of Data Mining (DM) [61], where data are extracted from companies' databases. The Cross Industry Standard Process for DM (CISPDM) is a standard DM process that consists of six stages: business understanding, data understanding, data preparation, modeling, evaluation, and implementation [49,64].

As presented, there is a strong foundation for using ML applications to estimate production time. However, the general implementation of ML in SMEs is insufficient [65]. This is also confirmed by a study of 18 companies [66], which shows that ML methods are increasingly being adopted by companies but that SMEs face challenges in implementing them compared to larger companies. A study of 60 companies in the engineering, construction, automotive, and electrical industries [67] found that the perceived importance of ML, willingness to pay, and readiness to perform data management are key factors influencing the likelihood of ML adoption in SMEs.

However, ML can reduce production times and scrap, improve resource utilization, and derive patterns from data [68]. Therefore, our research proposes a methodology that considers the receptivity factors of SMEs towards DM and ML for production time estimation.

1.6. Research Objective

The research presented in Section 1.4 focuses on concrete products and concrete time estimation. However, there is a lack of information in the mentioned sources on how to proceed in the application of PAs. In addition, the suitability of the model is evaluated based on the Correctly Classified Instances (CCI), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Relative Absolute Error (RAE), or Root Relative Squared Error (RRSE). These metrics are useful in research where the results come from using different PAs with different training setups for model building and where performances of models are compared. Comparing models is possible if they use the same data.

However, the goal of SMEs is to find models for different product groups with limited data. For this reason, our research introduces the risk metric that takes into account both: the accuracy of the modeling (CCI for classification and RAE for regression) and the data used. In the following, the methodology of a DM application is proposed. It includes the selection of a suitable PA for the time estimation model based on experiments and the evaluation of the model based on a risk metric. The proposed methodology is tailored to the conditions of SMEs and it is integrated into the CISPDM in the following steps: see Figure 2 in the color blue:

- Job type categorization to increase data based on similarity criteria;
- Model generation process;
- Model evaluation based on a defined risk metric.

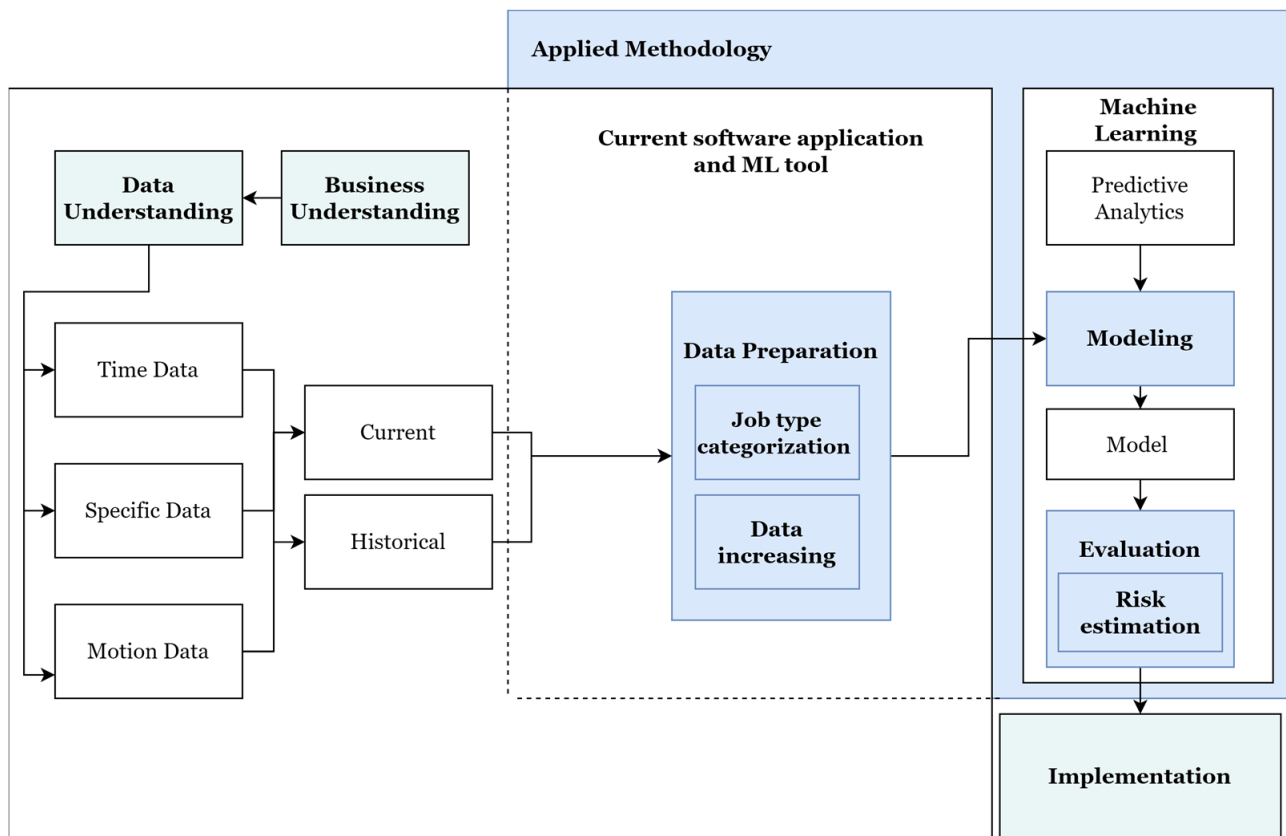


Figure 2. Integration of the applied methodology into the CISPDM.

2. Methods and Materials

2.1. Categorization of Job Types

Jobs in custom manufacturing can be divided into previously executed jobs, i.e., past variants, and new jobs. In relation to the customer, this results in the existence of four types of jobs labeled A, B, C, and D, which are presented in a bivariate Table 1.

Table 1. Categorization of job types.

	Current Customer	New Customer
Previously executed job	A	B
New job	C	D

Depending on the type of job, the proposed methodology considers different data-increasing strategies during model building. For type A jobs, there are specific data that describe this type of job. For type B, C, and D jobs, a similarity criterion must be defined. A similarity criterion is a criterion on the basis of which data are supplemented for modeling. It is an attribute for which the same or very limited values are assumed in the data. These are, for example, the same parts of the job, the same production operations and their queuing, the need to involve an external company in the process, the use of special tools, etc. To define the similarity criterion, it is recommended to go through the specifications of jobs and technical drawings. Finding the similarity criterion is crucial for types B, C, and D; it is essential to increasing the amount of data in modeling.

2.2. Data Increasing

Time is a random variable to which the central limit theorem applies. This says the following:

The central limit theorem: The distribution of a random variable, which is the sum of a large number of independent random variables, approximates a normal distribution.

From the central limit theorem, it follows that, if the amount of data can be increased based on the similarity criterion, it is possible to build a time estimation model even for low amounts of data. In general, $n \geq 30$ is used as a sufficiently large amount of data for the normality assumption. This threshold is used to systematize the data for model building in the following way:

- When the amount of data for the same jobs is $n \geq 30$, the data are divided into training data and independent test data. These are real job execution data. The training data are used for model building and validation by cross-validation. The resulting model is verified using independent test data. These are type A job data.
- If the amount of data for the same jobs is $30 > n \geq 3$, the job data are divided into training data and independent test data. Based on the similarity criterion, the training data are supplemented with data from similar jobs, i.e., data close to the actual execution of the job. These are used for model building and cross-validation. The resulting model is verified on independent test data. These are type B and C jobs.
- If the amount of data for the same jobs is $3 > n$, the jobs are classified as type D jobs. There are no specific data for these jobs. Data from other jobs, which are close to the possible reality of job execution, are used to build models. These models cannot be verified with specific data.

2.3. Modeling

Before modeling, it is necessary to select suitable attributes and PAs. The attributes should be specific to the particular job for which the model is being sought.

Modeling is based on the available original data, which are instances of attributes. For a type A job, the most data are available. Therefore, the selection of suitable PAs for modeling is made based on experimentation with type A data.

During experimentation, models are trained by applying different PAs to data samples. The training data affects the modeling results. Modeling errors can occur during experiments. Underfitting is a model error that occurs when attributes are selected that are too general or when an inappropriate PA algorithm is chosen, while overfitting is a model error that occurs when attributes have been selected for a particular special event [49,61]. Noise in the data can lead to overfitting. Modeling underfitting errors are eliminated by selecting different attributes or by selecting a more appropriate PA. Further training data distinguishability affects the model building. PAs, as decision trees, find more boundaries between classes, therefore they are suitable for data with high separability. For linearly separable data, SVM can be used to find the boundary between two classes. For data with low variability, PAs with lazy algorithms or simple rules that find one class are suitable. Data and model dependencies for different PAs are discussed in [49,69–71].

Based on the experimental results of time prediction for job type A, the PAs whose models achieve the highest CCI or RAE scores are selected. The selected PAs are also suitable for time prediction modeling of job types B, C, and D because the predicted time is assumed to depend on the same attributes. Selection of additional PAs is possible.

The modeling is performed according to the procedure, as shown in Figure 3. In ML applications, data augmentation techniques to increase the data amount by using synthetic data are used [70]. However, this requires a deep knowledge of ML and it is time-consuming because of the high diversity of products in SMEs. For the purpose of increasing the data amount, the similarity criterion is used, as shown in Section 2.2.

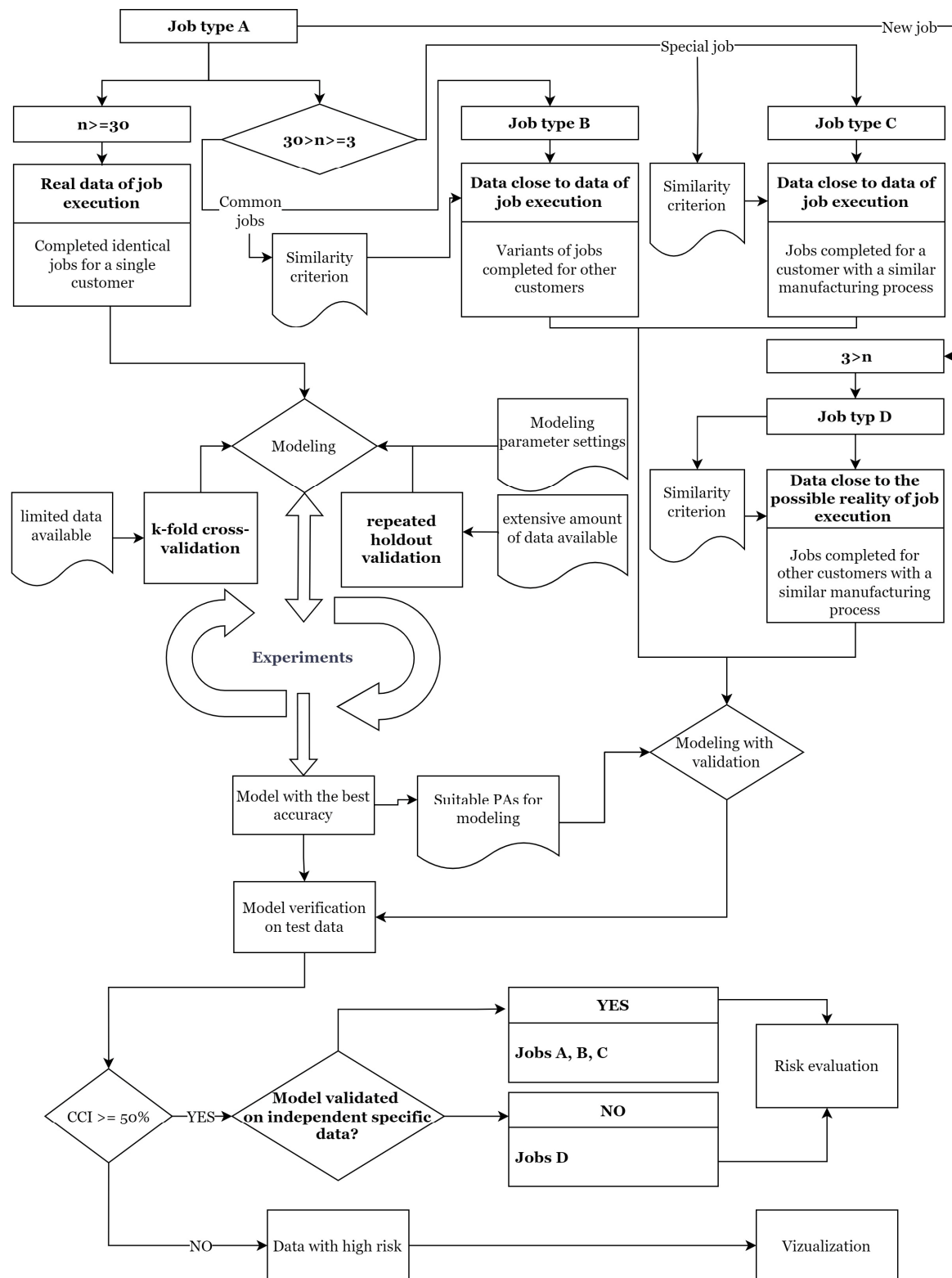


Figure 3. Application scheme of the methodology.

2.4. Risk Definition and Evaluation

When building a model to estimate production time for job types B, C, and D, it is necessary to increase the amount of data by using data from other jobs. These data are not

specific to the job in question. Therefore, there must be an assessment of the estimation risk in the use of the model:

$$R^G = E \frac{\sigma}{\mu}; \quad (1)$$

where

- R Risk;
- G Risk class, where a higher G indicates a lower risk of estimation, $G \in \{1, 2, \dots, k\}$; k is the number of classes, for our purpose four classes were defined;
- E Weight of estimation error (where a larger error in the model indicates greater estimation risk), $E \geq 1$;
- $\frac{\sigma}{\mu}$ The coefficient of variation is a measure of data variability, where σ is the standard deviation and μ is the mean value of data.

In our research, the behavior of models was observed, based on which four classes of modeling risk were defined:

- Models of risk class $G = I$ are not suitable for application because they represent an insufficient proportion of both training and test data. These models are inadequate and it is necessary to change the estimation method, select a different PA, select different attributes, or select a method other than ML.
- Risk class $G = II$ models are models where the CCI is significantly different between the training and test data, but it is greater than 50% in both cases. This means that, if the cause of the difference can be identified, these models are conditionally applicable.
- Risk class $G = III$ models describe the behavior of both the training and testing data and are suitable for use.
- Risk class $G = IV$ models describe the behavior of the data very accurately, and it is possible to consider simplifying them if the possible subsequent decrease in the CCI is acceptable.

For the purpose of the methodology, the following risk limits have been defined in Table 2.

Table 2. Types of risks.

CCI by Model	Model Error	Weight of Estimation Error (E)	Risk (R^G)
$\geq 50\%$	$< 50\%$	≤ 1.5	R^I
$\geq 65\%$	$< 35\%$	≤ 1.35	R^{II}
$\geq 80\%$	$< 20\%$	≤ 1.2	R^{III}
$\geq 95\%$	$< 5\%$	≤ 1.05	R^{IV}

Based on the properties of the normal distribution, the limits of the data variability were defined:

- Data are stable if 99.73% of the data values belong to the interval $\langle \mu - 3\sigma; \mu + 3\sigma \rangle$ or if 95.45% of the data values belong to the interval $\langle \mu - 2\sigma; \mu + 2\sigma \rangle$.
- Data with variability, where 80% of the data values belong to the interval of $\langle \mu - \sigma; \mu + \sigma \rangle$.
- Data with significant variability, where 50% or less of the data values fall within the interval $\langle \mu - 0,67\sigma; \mu + 0,67\sigma \rangle$.

Individual combinations of weight of estimation error and coefficient of variation imply steps. Table 3 is a guideline for evaluating the modeling.

Table 3. Evaluation guideline.

Risk	Model Evaluation	Data Evaluation	Recommendation
$0 < R^I \leq 0.75$	Model not suitable	Data stable	Underfitting, selection of multiple attributes, selection of another modeling method
$0.75 < R^I \leq 1.5$	Model not suitable	Data variation	Choosing a different modeling method and incrementing data
$1.5 < R^I \leq 3$	Model not suitable	Significant data variation	Choosing a different modeling method and incrementing data
$0 < R^{II} \leq 0.675$	Model conditionally suitable	Data stable	Choice of a different modeling method, or conditional implementation
$0.675 < R^{II} \leq 1.35$	Model conditionally suitable	Data variation	Choice of a different modeling method, or conditional implementation
$1.35 < R^{II} \leq 2.7$	Model conditionally suitable	Significant data variation	Selection of a different method of modelling or data incrementing
$0 < R^{III} \leq 0.6$	Model suitable	Data stable	A suitable model for implementation
$0.6 < R^{III} \leq 1.2$	Model suitable	Data variation	A suitable model for implementation
$1.2 < R^{III} \leq 2.4$	Model suitable	Significant data variation	A suitable model for implementation
$0 < R^{IV} \leq 0.525$	Very accurate model	Data stable	High probability of overfitting, simplification of the model possible
$0.525 < R^{IV} \leq 1.05$	Very accurate model	Data variation	High probability of overfitting, simplification of the model possible
$1.05 < R^{IV} \leq 2.1$	Very accurate model	Significant data variation	High probability of overfitting, simplification of the model possible

3. Methodology Application

The methodology was applied to model MJ/CT, manufacturing CT, and LT time estimations of job types A and B using data from a German SME, Schoepstal Maschinenbau GmbH, founded in 1991. The company currently has 130 employees and a cash flow of approximately EUR 15 million per year. The application is carried out for the estimation of machine/job cycle time (M/J CT), manufacturing CT without delays, and LT (manufacturing CT with delays). The company is not specialized in a specific industry and the products

are diverse, with various functions. They are equipment, components, or assembly products used in agriculture, fishing, machinery, mining, chemical, transportation, construction, defense, and other industries.

3.1. Business Understanding

Products range from simple sheet metal components to complex assembly groups. Depending on the operations and manufacturing technologies required to complete the job, the parts can be classified into the following groups: cut parts, machined parts, welded parts, drilling parts, painted parts, and assembled parts.

The company uses the ERP system, which contains basic information about jobs. In addition, the company uses a shop floor data collection system, in which the operating times of machines/jobs for individual jobs are manually entered.

3.2. Data Understanding

The historical time and specific data of jobs from 2021 to 2023 were collected from the systems to apply the methodology.

The systems contain the following information used for modeling: job acceptance and packing slip dates (job contains multiple packing slips), packing slip items, count of items, customer name, payment volume, worker name, and date and time the task on the job was performed by the workplace. A total of 3704 jobs are assigned to 29 workplaces and 370 customers.

3.3. Data Preparation

Data preparation involved merging data from both systems. Power Query and Excel functions were used to sort and consolidate the data. The basic consolidated data set contains 3136 terms of items assigned to customer orders. Based on the available data, the item term is an identified similarity criterion. Item terms need to be unified. The goal was to identify a limited number of items for modeling, at least 50 types of item names. The identification was performed through an iterative procedure using the least error value selection rule (OneR). OneR selects as the true value the value of the attribute corresponding to the highest frequency of the searched variable [49].

As part of the identification, 10 iterative steps were performed. A total of 121 item terms were identified, corresponding to the number of 2007 items of jobs. Based on these results, data were categorized in Table 4. Some item terms are repetitive; 95 unique item terms were identified.

Table 4. Categorization of the results of the item identification.

Items Type A $n \geq 30$	Items Type B and C $30 > n \geq 30$	Items Type D $3 > n$
16 item terms 6 customers	74 item terms 18 customers	31 item terms 12 customers

3.4. Modeling Application

The methodology of creating a time estimation model was applied to M/J CT, CT manufacturing, and LT estimation of item types A and B:

- M/J CT for one type A item of one customer, with 10 manufacturing tasks;
- Manufacturing CT for seven type A items of one customer;
- LT for seven type A items of one customer;
- M/J CT for two type B items with five manufacturing tasks and six customers;
- Manufacturing CT for one type B item of four customers;
- LT for one type B item of four customers.

To model the estimated time of B items, the data were increased by data from other customers according to the similarity criterion item term. For type C and D items, it was

not possible to define a similarity criterion from the data obtained from the databases. Therefore, the methodology was not applied to type C and D items.

The proposed methodology implements the selection of suitable PAs for modeling in the form of model validation in the WEKA Experimenter.

3.4.1. J48 Decision Tree Algorithm

Classification methods were chosen for the estimation of M/J CT and LT. Since the estimated time is a continuous variable, it was first necessary to discretize the values into classes so that individual applications of the J48 algorithm would assign values to classes of different ranges. Number of classes (NC) was chosen based on the number of possible training values:

$$NC = \{2, 3, \dots, \sqrt{n}\} \quad (2)$$

where n is the number of possible training values.

The J48 decision tree algorithm aims to create the smallest possible tree. The smallest tree is achieved when a single branch can be assigned as many estimates of the same value as possible, resulting in a simple node. This node provides the most information. The algorithm uses a top-down recursive approach, leveraging the entropy function (information value) from information theory [48,49]. Entropy is calculated as follows:

$$Entropy(p_1, \dots, p_n) = -p_1 \log_2 p_1 - p_2 \log_2 p_2 \dots - p_n \log_2 p_n; \quad (3)$$

where n is the number of possible training values and $p_1, \dots, p_n$ is the probability of occurrence of the training value calculated as the ratio of the number of training value to the total number of possible training values, $p_1, \dots, p_n \leq 1$.

The algorithm computes the entropy for individual attributes and the entropy of the learning value. The starting node is then chosen as the one where the difference between the entropy of the learning value and the entropy of the attribute is the greatest.

The models were built with pruning (for at least two objects in the learning value classes) and without pruning (for at least two, three, and four objects in the learning value classes). By pruning the trees, a simplification of the tree structure by sufficient metrics of the models was expected. M/J CT estimation modeling was possible based only on the attribute count of the items. An example of the experimental inputs for the CT assembly of the machine base item are summarized in the Table 5; the model generated by WEKA is shown in Figure 4.

Table 5. Experiment result for CT assembly machine base item.

Training Data (2021, 2022)	Test Data (2023)	NC in Experiment	Selected Number of Classes	Selected Setting of J48
20	10	4, 3	3	Unpruned J48/M 2

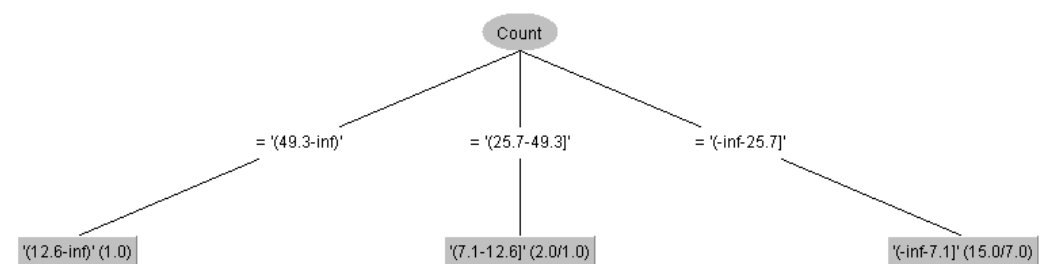


Figure 4. Selected training model for assembly time estimation of item machine base by item count.

For the LT modeling attributes, manufacturing CT, payment volume, number of items in production on the day of receipt, and number of items were tested for dependency on the estimated LT by applying correlation and the relief ranking filter. Attributes redundancy was tested using the Wrapper Subset Evaluator in combination with the Swarm or Hill-climber algorithms. Only the attributes manufacturing CT and payment volume showed some dependence on LT. These two were selected for modeling. Algorithms with different induction strategies (J48, PrismRule, ID3) were used to model decision trees, and EL PA rotation forest was applied. For an explanation of PAs algorithms, please see [49,57,72].

3.4.2. Regression Methods

Regression methods were chosen for estimating the manufacturing CT since this time is conditioned by the sum of the individual M/J CTs. In linear regression, the expected value y to be found depends on several input attributes, X , and the problem is solved in a vector space:

$$y = \beta X + \varepsilon; \quad (4)$$

where each attribute has its own β -coefficient and ε is the error of prediction; it is the difference between the training value and the predicted value.

Least squares regression divides the data into subsets for which it builds linear regression models. It selects the model with the smallest RMSE as the output model. The M5 algorithm combines a decision tree and a linear regression function. J48 is used to select the best linear regression model [49].

A total of three experiments were run for items of type A, for which 39 method settings were run on 38 datasets with a final number of 24,884 models. A summary of the experiments is shown in Table 6. A total of 24 models with the best metrics in CCI, or RAE, were selected for tests on independent test data. For the modeling of M/J CT, unpruned decision trees with different numbers of classes for MJ/CT estimation of item machine base were selected, as shown in Table 7. Table 8 summarizes the results of the tests for models of manufacturing CT. The results of the modeling experiment and the LT model tests resulted in disappointing CCI values, as shown in Table 9.

Table 6. Summary of experiments.

Modeling	Number of Models in Experiment	Number of Evaluated Models	Selected PAs
M/J CT	384	10	Unpruned J48
Manufacturing CT	12,600	7	Linear regression, and least square regression
LT	11,900	7	Visualization

Table 7. Results of testing the M/J CT models of A type item “machine base”.

Machine/Job	Selected NC	CCI (%)
Laser Programming	4	50
Laser	3	70
Drilling	4	65
Milling	3	55
Welding	5	72
Cleaning works	3	90
Additional works	3	78
Assembly	3	56
Bending	4	65
Cutting	5	50

Table 8. Selected models for A-type items based on experiments for manufacturing CT.

Item	Method	RAE (%)
Holder	LinearRegr./Correl/N	1.56
Baseplate	LinearRegr./Correl/0.3	0.96
Underframe	LeastMedSq/Correl/N	0.67
Machine base	LinearRegr./Correl/N	1.32
Tub	LeastMedSq/Correl/N	3.93
Support part	LinearRegr./Correl/0.3	8.21
Sheet metal part	LinearRegr./Correl/N	26.87

Table 9. Results of testing the LT models of A-type items.

Item	CCI (%)
Holder	40
Baseplate	6
Underframe	17
Machine base	7
Tub	27
Support part	14
Sheet metal part	18

For type B items, no experiments were performed, but PAs with the best performance from modeling time estimation of type A items were selected, specifically, decision trees for M/J CT estimation, as shown in Table 10, and linear regression for manufacturing CT, as shown in Table 11. In the LT modeling experiment, no suitable models were found for the type A items. Therefore, LT modeling was not performed, and data behavior visualization was used for type B items, as shown in Figure 5.

Table 10. Results of testing the M/J CT models of type B items.

Item	Machine/Job	Customer	CCI (%)
Steel plate	Cutting	D10183	72
Steel plate	Cutting	D10605	72
Steel plate	Shearing	D10206	50
Steel plate	Shearing	D10605	60
Lasered and bent part	Laser programming	D10216	50
Lasered and bent part	Laser programming	D10607	75
Lasered and bent part	Laser programming	D10605	75
Lasered and bent part	Laser programming	D10600	75
Lasered and bent part	Laser	D10216	50
Lasered and bent part	Laser	D10607	87
Lasered and bent part	Laser	D10605	80
Lasered and bent part	Laser	D10600	87
Lasered and bent part	Bending	D10216	75
Lasered and bent part	Bending	D10607	75
Lasered and bent part	Bending	D10605	75
Lasered and bent part	Bending	D10600	67

Table 11. Results of testing the manufacturing CT models of type B item “Steel plate”.

Customer	RAE (%)
D10605	5.65
D10216	25.0
D10183	121.38
D10176	37.03



Within the methodology, a risk of estimation was proposed, taking into account the error of the model in combination with the variation of the data. For the model error, the minimum value of CCI and RAE of models on training and test data were considered. For the presented items, the evaluation results are summarized in Tables 12 and 13.

Risk	Use Another Method	Conditionally Suitable Model	Suitable Model for Implementation	Suitable Model, Simplifying Possible
M/J CT models of type A item "Machine base"				
$R^I = 1.04$	x			
$R^{II} = 1.43$		x		
$R^{II} = 1.27$		x		
$R^I = 0.89$	x			
$R^{II} = 1.10$		x		
$R^{III} = 1.10$			x	
$R^{II} = 1.11$		x		
$R^I = 0.91$	x			
$R^{II} = 0.72$		x		
$R^I = 1.22$	x			
Manufacturing CT models of type A items				
$R^{IV} = 1.80$				x
$R^{IV} = 1.70$			x	
$R^{IV} = 0.74$				x
$R^{IV} = 0.98$				x
$R^{IV} = 0.70$				x
$R^{III} = 1.07$			x	
$R^{II} = 1.43$	x			
LT models of type A items				

els is greater than 50%, which corresponds to excessive risk. Therefore, all models are unsuitable.

Table 13. Evaluation of models of B type items.

Machine/Job/Item	Customer	$\frac{\sigma}{\mu}$	Risk	Use Another Method	Conditionally Suitable Model	Suitable Model for Implementation	Suitable Model, Simplifying Possible
M/J CT models of type B item “Steel plate”							
Cutting	D10183	0.51	$R^{II} = 0.66$		x		
Cutting	D10605	0.51	$R^{II} = 0.66$		x		
Shearing	D10206	0.54	$R^I = 0.81$	x			
Shearing	D10605	0.54	$R^I = 0.75$	x			
M/J CT models of type B item “Lasered and bent part”							
Laser programming	D10216	1.14	$R^I = 1.72$	x			
Laser programming	D10607	1.14	$R^{II} = 1.43$	x			
Laser programming	D10605	1.14	$R^{II} = 1.43$	x			
Laser programming	D10600	1.14	$R^{II} = 1.43$	x			
Laser	D10216	1.44	$R^I = 2.17$	x			
Laser	D10607	1.44	$R^{III} = 1.63$			x	
Laser	D10605	1.44	$R^{III} = 1.73$			x	
Laser	D10600	1.44	$R^{III} = 1.63$			x	
Bending	D10216	0.62	$R^{II} = 0.78$		x		
Bending	D10607	0.62	$R^{II} = 0.78$		x		
Bending	D10605	0.62	$R^{II} = 0.78$		x		
Bending	D10600	0.62	$R^{II} = 0.83$	x			
Manufacturing CT models of type B items							
Steel plate	D10605	1.19	$R^{III} = 1.255$			x	
Steel plate	D10216	1.19	$R^{II} = 1.484$		x		
Steel plate	D10183	1.19	Excessive risk	x			
Steel plate	D10176	1.19	$R^I = 1.63$	x			
LT models of type B items							
Modeling was not applied.							

4. Discussion

In the experiments, various PAs, except NNs, and data samples were used to train the models. The best-performing models, based on the highest achieved metric scores, were selected for evaluation, as shown in Table 6. However, a model’s performance is only as good as the quality of the training data. In our research, we used the original dataset, assuming this would lead to greater acceptance by SMEs in applying ML. Therefore, at this stage, we chose not to apply data augmentation techniques to increase the dataset size. In some cases, no suitable ML model could be identified, suggesting that non-ML methods may be more appropriate, as shown in Tables 12 and 13.

The methodology was applied to model the MJ/CT, manufacturing CT, and LT time estimations of product types A and B. The application of the estimation risk assessment identified 22 of 37 models of time estimation that are suitable or conditionally suitable for implementation, as shown in Table 14.

The limited number of attributes combined with the low data rate resulted in inaccurate M/J CT estimation models. Estimating M/J CT based on the attribute count of the item alone is insufficient. Nevertheless, the choice of the decision tree algorithm is appropriate, the resulting models are easy to understand, and they provide an overview of the range in which the M/J CT varies for individual items.

The modeling of manufacturing CT estimation has found accurate models for type A jobs, but their further application is conditioned by accurate M/J CT estimates. By modeling the manufacturing CT of type B jobs, which was performed on increased data, significantly less accurate models were found. This points to the insufficient representativeness of the

data of the modeled phenomenon. The selection of inappropriate data causes this, as the similarity criterion was only the same item term.

Table 14. Overview of evaluation results.

Modeling	Number of Suitable and Conditionally Suitable Models
A type items	
M/J CT	6
CT manufacturing	6
LT	0
B type items	
M/J CT	8
CT manufacturing	2
LT	0

Based on the training data, no suitable models for LT were found in the experiments because the LT estimation attribute values and the learning values showed a low dependence. The low dependence of attributes is caused by the random method of job management by production in the company. For ML modeling of LT prediction, it is necessary to enter more attributes into information systems. Other methods than ML should be used for LT prediction. However, visualization of data with the integration of PAs for the representation of data boundaries depending on two attributes is helpful for the purpose of understanding the data behavior.

For type C and D jobs, the methodology could not be verified. The information strength of the data obtained from the systems is not sufficient to define the criterion of similarity of C and D jobs. In order to define the similarity criterion, it is necessary to go through the job specifications and technical drawings with the participation of an expert with in-depth knowledge of the company's products. This was not possible due to time constraints. However, the application of the method of creating the model for types C and D jobs is analogous to the presented creation for models of types A and B jobs. For type D jobs, implementing data augmentation techniques seems to make sense.

Data from a medium-sized company were used to apply the methodology. The product range of this company was so broad that, in some cases, we did not have enough data. Limited data is also typical for small companies. Otherwise, the range of products is not as wide as in the presented company and the similarities within the products are more present. Therefore, the idea of categorization based on similarity criterion also allows for an increase in the data available for modeling in small companies, if the data are available. This should be verified in further research. Additionally, comparison with other methodologies should be included in the scope of future studies.

5. Conclusions

The success of the application of DM is strongly conditioned by the quality and quantity of data. The presented methodology of DM was applied to historical data of concrete SMEs obtained from the company's database systems. The quality of the data was limited by a low dependency on learning data and a small number of attributes. This ultimately led to the creation of LT models with low accuracy. However, although the quality of the modeling data was limited, it was possible to present the application of the proposed methodology. The proposed categorization of job types is useful in the sense of increasing data and has led to the finding of suitable models. By defining the risk metric as a function of the accuracy of the model and the variation of the data, the entire modeling is evaluated as such, not just the model, as in previous research. The risk metric can also be used in production scheduling. To increase the success of the DM application in SMEs, it is recommended that specific information be added to the company's database, such as part

geometry, assembly structure, topology information, tolerances, material properties, etc. This can be achieved by linking CAD and ERP systems.

Author Contributions: Conceptualization, M.U. and F.K.; methodology, M.U., F.K. and R.M.; validation, M.U., F.K. and R.M.; resources, M.U., F.K. and R.M.; data curation, M.U.; writing—original draft preparation, M.U.; writing—review and editing, F.K.; supervision, R.M. All authors have read and agreed to the published version of the manuscript.

Funding: This publication was written at the Technical University of Liberec, Faculty of Mechanical Engineering with the support of the Institutional Endowment for the Long-Term Conceptual Development of Research Institutes, as provided by the Ministry of Education, Youth and Sports of the Czech Republic in the year 2024. The research reported in this paper was supported by institutional support for nonspecific university research.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. (The data are not publicly available due to privacy—company data).

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

References

1. Bokranz, R. *Arbeitswissenschaft, Zeitaufnahme und Weitere Techniken, Systeme Vorbestimmter Zeiten Multimomentaufnahme, Zeitberechnungsunterlagen*; Gabler: Wiesbaden, Germany, 1978; ISBN 3409385819.
2. Deutsche MTM-Vereinigung e.V.; Bokranz, R.; Landau, K.; Becks, C. *Produktivitätsmanagement von Arbeitssystemen*; Schäffer-Pöschel Verlag: Stuttgart, Germany, 2006; ISBN 3791021338.
3. REFA Methodenlehre der Betriebsorganisation/REFA—Verband für Arbeitsstudien und Betriebsorganisation e. V. Teil 3. *Planung und Steuerung*, 1st ed.; Carl Hansen: München, Germany, 1991; ISBN 3446163514.
4. REFA Methodenlehre des Arbeitsstudiums/REFA—Verband für Arbeitsstudien und Betriebsorganisation e. V. 1992, Teil 2. *Datenermittlung*, 7th ed.; Carl Hansen: München, Germany, 1992; ISBN 3446142355.
5. Ben Rabia, M.A.; Bellabdaoui, A. Simulation-Based Analytics: A Systematic Literature Review. *Simul. Model. Pract. Theory* **2022**, *117*, 102511. [\[CrossRef\]](#)
6. Kumar, V.; Garg, M.L. Predictive Analytics: A Review of Trends and Techniques. *Int. J. Comput. Appl.* **2018**, *182*, 31–37. [\[CrossRef\]](#)
7. Brühl, V. *Big Data, Data Mining, Machine Learning und Predictive Analytics: Ein konzeptioneller Überblick*; CFS Working Paper Series; Goethe University Frankfurt, Center for Financial Studies (CFS), Frankfurt a. M.: Frankfurt am Main, Germany, 2019.
8. Wu, S.; Wang, Y.; BolaBola, J.Z.; Qin, H.; Ding, W.; Wen, W.; Niu, J. Incorporating Motion Analysis Technology into Modular Arrangement of Predetermined Time Standard (MODAPTS). *Int. J. Ind. Ergon.* **2016**, *53*, 291–298. [\[CrossRef\]](#)
9. Golabchi, A.; Han, S.; AbouRizk, S.; Kanerva, J. Micro-Motion Level Simulation for Efficiency Analysis and Duration Estimation of Manual Operations. *Autom. Constr.* **2016**, *71*, 443–452. [\[CrossRef\]](#)
10. Turk, M.; Pipan, M.; Simic, M.; Herakovic, N. Simulation-Based Time Evaluation of Basic Manual Assembly Tasks. *Adv. Prod. Eng. Manag.* **2020**, *15*, 331–344. [\[CrossRef\]](#)
11. Karim, A.N.M.; Tuan, S.T.; Emrul Kays, H.M. Assembly Line Productivity Improvement as Re-Engineered by MOST. *Int. J. Product. Perform. Manag.* **2016**, *65*, 977–994. [\[CrossRef\]](#)
12. Razmi, J.; Shakh-Niyayee, M. Developing a Specific Predetermined Time Study Approach: An Empirical Study in a Car Industry. *Prod. Plan. Control* **2008**, *19*, 454–460. [\[CrossRef\]](#)
13. Zandin, K.B. *MOST Work Measurement Systems*, 4th ed.; CRC Press Taylor & Francis Group: Boca Raton, FL, USA, 2021; ISBN 978-0-367-34531-0.
14. Cho, H.; Lee, S.; Park, J. Time Estimation Method for Manual Assembly Using MODAPTS Technique in the Product Design Stage. *Int. J. Prod. Res.* **2014**, *52*, 3595–3613. [\[CrossRef\]](#)
15. Sullivan, B. *Sullivan Heyde's Modapts: A Language of Work*; Heyde Dynamics Pty Ltd.: Brisbane, Australia, 2001; ISBN 978-0-9596597-5-7.
16. Chen, J.; Zhou, D.; Kang, L.; Ma, L.; Ge, H. A Maintenance Time Estimation Method Based on Virtual Simulation and Improved Modular Arrangement of Predetermined Time Standards. *Int. J. Ind. Ergon.* **2020**, *80*, 103042. [\[CrossRef\]](#)
17. Cai, K.; Zhang, W.; Chen, W.; Zhao, H. A Study on Product Assembly and Disassembly Time Prediction Methodology Based on Virtual Maintenance. *Assem. Autom.* **2019**, *39*, 566–580. [\[CrossRef\]](#)
18. Assef, F.; Scarpin, C.T.; Steiner, M.T. Confrontation between Techniques of Time Measurement. *J. Manuf. Technol. Manag.* **2018**, *29*, 789–810. [\[CrossRef\]](#)
19. Bureš, M.; Pivodová, P. Comparison of Time Standardization Methods on the Basis of Real Experiment. *Procedia Eng.* **2015**, *100*, 466–474. [\[CrossRef\]](#)

20. Eraslan, E. The Estimation of Product Standard Time by Artificial Neural Networks in the Molding Industry. *Math. Probl. Eng.* **2009**, *2009*, e527452. [\[CrossRef\]](#)
21. Dağdeviren, M.; Eraslan, E.; Çelebi, F.V. An Alternative Work Measurement Method and Its Application to a Manufacturing Industry. *J. Loss Prev. Process Ind.* **2011**, *24*, 563–567. [\[CrossRef\]](#)
22. Polotski, V.; Beauregard, Y.; Franzoni, A. Combining Predetermined and Measured Assembly Time Techniques: Parameter Estimation, Regression and Case Study of Fenestration Industry. *Int. J. Prod. Res.* **2019**, *57*, 5499–5519. [\[CrossRef\]](#)
23. Kim, D.S.; Porter, J.D.; Buddhakulsomsiri, J. Task Time Estimation in a Multi-Product Manually Operated Workstation. *Int. J. Prod. Econ.* **2008**, *114*, 239–251. [\[CrossRef\]](#)
24. Al-Aomar, R.; El-Khasawneh, B.; Obaidat, S. Incorporating Time Standards into Generative CAPP: A Construction Steel Case Study. *J. Manuf. Technol. Manag.* **2013**, *24*, 95–112. [\[CrossRef\]](#)
25. Fang, W.; Guo, Y.; Liao, W.; Ramani, K.; Huang, S. Big Data Driven Jobs Remaining Time Prediction in Discrete Manufacturing System: A Deep Learning-Based Approach. *Int. J. Prod. Res.* **2020**, *58*, 2751–2766. [\[CrossRef\]](#)
26. Mohsen, O.; Mohamed, Y.; Al-Hussein, M. A Machine Learning Approach to Predict Production Time Using Real-Time RFID Data in Industrialized Building Construction. *Adv. Eng. Inform.* **2022**, *52*, 101631. [\[CrossRef\]](#)
27. Ruppert, T.; Abonyi, J. Software Sensor for Activity-Time Monitoring and Fault Detection in Production Lines. *Sensors* **2018**, *18*, 2346. [\[CrossRef\]](#)
28. Wang, C.; Jiang, P. Deep Neural Networks Based Order Completion Time Prediction by Using Real-Time Job Shop RFID Data. *J. Intell. Manuf.* **2019**, *30*, 1303–1318. [\[CrossRef\]](#)
29. Buzjak, D.; Kunica, Z. Towards Immersive Designing of Production Processes Using Virtual Reality Techniques. *INDECS* **2018**, *16*, 110–123. [\[CrossRef\]](#)
30. Bellarbi, A.; Jessel, J.-P.; Da Dalto, L. Towards Method Time Measurement Identification Using Virtual Reality and Gesture Recognition. In Proceedings of the 2019 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR), San Diego, CA, USA, 9–11 December 2019; pp. 191–1913.
31. Kunz, A.; Zank, M.; Nescher, T.; Wegener, K. Virtual Reality Based Time and Motion Study with Support for Real Walking. *Procedia CIRP* **2016**, *57*, 303–308. [\[CrossRef\]](#)
32. Armillotta, A.; Moroni, G.; Rasella, M. Computer-Aided Assembly Planning for the Diemaking Industry. *Robot. Comput. Integr. Manuf.* **2006**, *22*, 409–419. [\[CrossRef\]](#)
33. Eigner, M.; Roubanov, D.; Sindermann, S.; Ernst, J. Assembly Time Estimation Based on Product Assembly Information. In Proceedings of the DESIGN 2014 13th International Design Conference 2014, Windhoek, Namibia, 6–10 October 2014.
34. Heo, E.-Y.; Kim, D.-W.; Kim, B.-H.; Frank Chen, F. Estimation of NC Machining Time Using NC Block Distribution for Sculptured Surface Machining. *Robot. Comput. Integr. Manuf.* **2006**, *22*, 437–446. [\[CrossRef\]](#)
35. So, B.S.; Jung, Y.H.; Park, J.W.; Lee, D.W. Five-Axis Machining Time Estimation Algorithm Based on Machine Characteristics. *J. Mater. Process. Technol.* **2007**, *187–188*, 37–40. [\[CrossRef\]](#)
36. Yamamoto, Y.; Aoyama, H.; Sano, N. Development of Accurate Estimation Method of Machining Time in Consideration of Characteristics of Machine Tool. *J. Adv. Mech. Des. Syst. Manuf.* **2017**, *11*, JAMDSM0049. [\[CrossRef\]](#)
37. Huang, J.; Chang, Q.; Arinez, J. Product Completion Time Prediction Using A Hybrid Approach Combining Deep Learning and System Model. *J. Manuf. Syst.* **2020**, *57*, 311–322. [\[CrossRef\]](#)
38. Chang, J.; Kong, X.; Yin, L. A Novel Approach for Product Makespan Prediction in Production Life Cycle. *Int. J. Adv. Manuf. Technol.* **2015**, *80*, 1433–1448. [\[CrossRef\]](#)
39. Hsu, S.Y.; Sha, D.Y. Due Date Assignment Using Artificial Neural Networks under Different Shop Floor Control Strategies. *Int. J. Prod. Res.* **2004**, *42*, 1727–1745. [\[CrossRef\]](#)
40. Sajko, N.; Kovacic, S.; Ficko, M.; Palcic, I.; Klancnik, S. Manufacturing lead time prediction for extrusion tools with the use of neural networks. *Eng. Rev.* **2020**, *11*, 48–55. [\[CrossRef\]](#)
41. Chen, T.; Wang, Y.-C. A Two-Stage Explainable Artificial Intelligence Approach for Classification-Based Job Cycle Time Prediction. *Int. J. Adv. Manuf. Technol.* **2022**, *123*, 2031–2042. [\[CrossRef\]](#)
42. Owensby, E.; Namouz, E.Z.; Shanthakumar, A.; Summers, J.D. *Representation: Extracting Mate Complexity from Assembly Models to Automatically Predict Assembly Times*; American Society of Mechanical Engineers: New York, NY, USA, 2012.
43. Miller, G.M.; Mathieson, J.L.; Summers, J.D.; Mocko, G.M. Representation: Structural Complexity Of Assemblies To Create Neural Network Based Assembly Time Estimation Models. In Proceedings of the International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, IDETC/CIE 2011, Chicago, IL, USA, 12–15 August 2012. [\[CrossRef\]](#)
44. Schneckenreither, M.; Haeussler, S.; Gerhold, C. Order Release Planning with Predictive Lead Times: A Machine Learning Approach. *Int. J. Prod. Res.* **2021**, *59*, 3285–3303. [\[CrossRef\]](#)
45. Chien, C.-F.; Hsiao, C.-W.; Meng, C.; Hong, K.-T.; Wang, S.-T. Cycle Time Prediction and Control Based on Production Line Status and Manufacturing Data Mining. In Proceedings of the ISSM 2005, IEEE International Symposium on Semiconductor Manufacturing, San Jose, CA, USA, 13–15 September 2005; pp. 327–330.
46. Öztürk, A.; Kayaligil, S.; Özdemirel, N.E. Manufacturing Lead Time Estimation Using Data Mining. *Eur. J. Oper. Res.* **2006**, *173*, 683–700. [\[CrossRef\]](#)
47. Quinlan, J.R. Induction of Decision Trees. *Mach. Learn.* **1986**, *1*, 81–106. [\[CrossRef\]](#)

48. Quinlan, J.R. C4.5: Programs for Machine Learning. In *The Morgan Kaufmann series in Machine Learning*; Morgan Kaufmann Publishers: San Mateo, CA, USA, 2014; ISBN 978-0-08-050058-4.
49. Witten, I.H.; Eibe, F.; Hall, M.A.; Pal, C.J. *Data Mining: Practical Machine Learning Tools and Techniques (Fourth Edition)*, 4th ed.; The Morgan Kaufmann Series in Data Management Systems; Morgan Kaufmann: Cambridge, UK, 2017; ISBN 978-0-12-804291-5.
50. Zhu, H.; Woo, J.H. Hybrid NHPSO-JTVAC-SVM Model to Predict Production Lead Time. *Appl. Sci.* **2021**, *11*, 6369. [\[CrossRef\]](#)
51. Shao, Y.; Ji, X.; Zheng, M.; Chen, C. Prediction of Standard Time of the Sewing Process Using a Support Vector Machine with Particle Swarm Optimization. *Autex Res. J.* **2022**, *22*, 290–297. [\[CrossRef\]](#)
52. Ruschel, E.; Rocha Loures, E.D.F.; Santos, E.A.P. Performance Analysis and Time Prediction in Manufacturing Systems. *Comput. Ind. Eng.* **2021**, *151*, 106972. [\[CrossRef\]](#)
53. Choueiri, A.C.; Sato, D.M.V.; Scalabrin, E.E.; Santos, E.A.P. An Extended Model for Remaining Time Prediction in Manufacturing Systems Using Process Mining. *J. Manuf. Syst.* **2020**, *56*, 188–201. [\[CrossRef\]](#)
54. Ramirez, J.; Guaman, R.; Morles, E.C.; Siguenza-Guzman, L. Prediction of Standard Times in Assembly Lines Using Least Squares in Multivariable Linear Models. In *Applied Technologies, Proceedings of the ICAT 2019. Communications in Computer and Information Science, Quito, Ecuador, 3–5 December 2019*; Botto-Tobar, M., Zambrano Vizuete, M., Torres-Carrión, P., Montes León, S., Pizarro Vásquez, G., Durakovic, B., Eds.; Springer International Publishing: Cham, Switzerland, 2020; pp. 455–466.
55. Chen, T. A Fuzzy-Neural Knowledge-Based System for Job Completion Time Prediction and Internal Due Date Assignment in a Wafer Fabrication Plant. *Int. J. Syst. Sci.* **2009**, *40*, 889–902. [\[CrossRef\]](#)
56. Dogan, A.; Birant, D. Machine Learning and Data Mining in Manufacturing. *Expert Syst. Appl.* **2021**, *166*, 114060. [\[CrossRef\]](#)
57. James, G.; Witten, D.; Hastie, T.; Tibshirani, R. *An Introduction to Statistical Learning*; Springer Texts in Statistics; Springer: New York, NY, USA, 2013; Volume 103, ISBN 978-1-4614-7137-0.
58. Chen, T. Job Cycle Time Estimation in a Wafer Fabrication Factory with a Bi-Directional Classifying Fuzzy-Neural Approach. *Int. J. Adv. Manuf. Technol.* **2011**, *56*, 1007–1018. [\[CrossRef\]](#)
59. Šarić, T.; Šimunović, G.; Šimunović, K.; Svalina, I. Estimation of Machining Time for CNC Manufacturing Using Neural Computing. *Int. J. Simul. Model.* **2016**, *15*, 663–675. [\[CrossRef\]](#)
60. Wu, Y.; Hou, F.; Cheng, X. Real-Time Prediction of Styrene Production Volume Based on Machine Learning Algorithms. In *Advances in Data Mining. Applications and Theoretical Aspects*; Perner, P., Ed.; Springer International Publishing: Cham, Switzerland, 2017; pp. 301–312. [\[CrossRef\]](#)
61. Witten, I.H.; Eibe, F. *Data Mining—Praktische Werkzeuge Und Techniken für das Maschinelle Lernen*; Carl Hanser Verlag: München, Germany; Wien, Austria, 2001; ISBN 3-446-21533-6.
62. Chen, T. Incorporating Fuzzy C-Means and a Back-Propagation Network Ensemble to Job Completion Time Prediction in a Semiconductor Fabrication Factory. *Fuzzy Sets Syst.* **2007**, *158*, 2153–2168. [\[CrossRef\]](#)
63. Ji, J.; Pannakkong, W.; Buddhakulsomsiri, J. A Computer Vision-Based Model for Automatic Motion Time Study. *Comput. Mater. Contin.* **2022**, *73*, 3557–3574. [\[CrossRef\]](#)
64. Cleve, J.; Lämmel, U. *Data Mining: Datenanalyse für Künstliche Intelligenz*; De Gruyter Oldenbourg: Berlin, Germany, 2024; ISBN 978-3-11-138770-3.
65. Döbel, I.; Leis, M.; Molina Vogelsang, M.; Neustroev, D.; Petzka, H.; Rüping, S.; Voss, A.; Wegele, M.; Welz, J. Maschinelles Lernen—Kompetenzen, Anwendungen Und Forschungsbedarf 2018. Available online: <https://www.bigdata-ai.fraunhofer.de/de/publikationen/ml-studie.html> (accessed on 9 September 2024).
66. Bauer, M.; van Dinther, C.; Kiefer, D. Machine Learning in SME: An Empirical Study on Enablers and Success Factors. Americas Conference on Information Systems 2020. Available online: <https://core.ac.uk/download/pdf/326836032.pdf> (accessed on 6 May 2024).
67. Burggräf, P.; Steinberg, F.; Sauer, C.R.; Nettesheim, P. Machine Learning Implementation in Small and Medium-Sized Enterprises: Insights and Recommendations from a Quantitative Study. *Prod. Eng. Res. Devel.* **2024**. [\[CrossRef\]](#)
68. Wuest, T.; Weimer, D.; Irgens, C.; Thoben, K.-D. Machine Learning in Manufacturing: Advantages, Challenges, and Applications. *Prod. Manuf. Res.* **2016**, *4*, 23–45. [\[CrossRef\]](#)
69. Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning*; Springer Series in Statistics; Springer: New York, NY, USA, 2009; ISBN 978-0-387-84857-0.
70. Mumuni, A.; Mumuni, F. Data Augmentation: A Comprehensive Survey of Modern Approaches. *Array* **2022**, *16*, 100258. [\[CrossRef\]](#)
71. Shalev-Shwartz, S.; Ben-David, S. *Understanding Machine Learning: From Theory to Algorithms*; Singapore Cambridge University Press: Cambridge, UK; New York, NY, USA; Port Melbourne, Australia; Delhi, India, 2014; ISBN 978-1-107-05713-5.
72. Patel, N.; Upadhyay, S. Study of Various Decision Tree Pruning Methods with Their Empirical Comparison in WEKA. *Int. J. Comput. Appl.* **2012**, *60*, 20–25. [\[CrossRef\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.