

Ekwaro-Osire, Henry; Bode, Dennis; Ohlendorf, Jan-Hendrik; Thoben, Klaus-Dieter

Article — Published Version

Manufacturing process energy consumption modeling: a methodology to identify the most appropriate model

Journal of Intelligent Manufacturing

Provided in Cooperation with:

Springer Nature

Suggested Citation: Ekwaro-Osire, Henry; Bode, Dennis; Ohlendorf, Jan-Hendrik; Thoben, Klaus-Dieter (2024) : Manufacturing process energy consumption modeling: a methodology to identify the most appropriate model, Journal of Intelligent Manufacturing, ISSN 1572-8145, Springer US, New York, NY, Vol. 36, Iss. 8, pp. 5673-5693,
<https://doi.org/10.1007/s10845-024-02514-z>

This Version is available at:

<https://hdl.handle.net/10419/330888>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<https://creativecommons.org/licenses/by/4.0/>



Manufacturing process energy consumption modeling: a methodology to identify the most appropriate model

Henry Ekwaro-Osire^{1,2} · Dennis Bode^{1,2} · Jan-Hendrik Ohlendorf² · Klaus-Dieter Thoben^{1,2}

Received: 6 August 2023 / Accepted: 12 October 2024 / Published online: 20 November 2024
© The Author(s) 2024

Abstract

This paper investigates the appropriateness of machine learning (ML) and other techniques for modeling manufacturing processes energy consumption by developing a comparison methodology. Three research questions are posed: Firstly, how do prediction errors compare using different techniques with varying complexity and ML use? Secondly, how does performance vary with different amounts of data? Thirdly, how do different techniques compare in terms of required expertise, effort to build and interpretability of results? To answer these questions, the authors develop a structured approach, which is also envisioned to be useable by practicing engineers and manufacturers. Four modeling categories are defined, ranging from simple non-ML methods, such as linear regression, to complex ML methods, such as deep neural networks. The approach is evaluated using data from a compound feed manufacturing process. The results confirm the notion that non-ML models are better suited to understand and model manufacturing processes when few parameters are present, due to their high interpretability, while ML models are recommended for analyzing processes with many potentially relevant and interrelated parameters. Interestingly the approach finds that the complex ML category model does not outperform the simple ML category model in terms of prediction accuracy, and only has the drawback of requiring more expertise to build and having lower interpretability. The study concludes that the decision to use complex ML for modeling manufacturing process energy consumption should be critically questioned and that a simpler approach may be better suited, suggesting that the developed methodology would be of value to practicing engineers.

Keywords Manufacturing · Modeling · Data-driven sustainability · Energy efficiency · Resource efficiency · Machine learning · Optimization

Introduction

Motivation and objective

Machine learning (ML) is a powerful tool for identifying hidden energy savings potentials in manufacturing processes (Jamwal et al., 2021). Since the manufacturing sector is responsible for over a third of global energy consumption,

approaches for improving energy efficiency of manufacturing processes are greatly needed (IEA, 2020). The need for industry to reduce its carbon footprint is omnipresent in both academia and business, and is only growing in importance every year. The parallel growing adoption and capability of ML, logically raises the opportunity to bring these two topics together, and the need to do this systematically. The strength of ML lies in analyzing processes with many potentially relevant and interrelated parameters, and in providing recommendations or predictions on the outcome (e.g. energy consumption, resource consumption, product quality features, etc.) of the process. Thus, ML is especially well suited to model manufacturing processes and systems, to then optimize the operating parameters or the input materials with the objective of reducing energy consumption (Ekwaro-Osire et al., 2022; Samadiani et al., 2024; Surindra et al., 2024; Waltersmann et al., 2021). While ML models have proven to

✉ Henry Ekwaro-Osire
eko.biba@mailbox.org

¹ BIBA – Bremen Institute for Production and Logistics GmbH, Hochschulring 20, 28359 Bremen, Germany

² Faculty of Production Engineering, Institute for Integrated Product Development (BIK), University of Bremen, Badgasteiner Straße 1, 28359 Bremen, Germany

be valuable in recent years, they require more rare expertise than some non-ML modeling approaches. Similarly among the many different ML models, the amount of complexity and required expertise varies significantly (Engbers & Freitag, 2024). Additionally, ML models are typically “black box”, meaning it cannot be explained how the model resulted in specific prediction. Non-ML modeling methods, such as simulations or simple linear regressions can counter some of these disadvantages of ML methods, by typically having greater interpretability and requiring less expertise to build. Non-ML approaches differ from ML approaches in that they can be better suited when the main objective of the analysis is to mathematically represent the underlying physical phenomena of a process, or when processes with limited parameters are to be analyzed (whereas an ML model “only” provides a model for the relationship between process inputs and outputs, without providing a mathematical representation of inner workings of this relationship). The last decade has seen a significant increase in the application of ML in manufacturing industry, as demonstrated in a recent survey by Google Cloud where two thirds of manufacturers reported using AI in their regular operations and the mean share of IT spend on AI of all surveyed companies was over 30% (Google Cloud, 2021). The uptake of interest in AI and ML can also be seen in academia, where AI topics were present in 10% to as much as 25% of all manufacturing publications in recent years, as found in a study by the European Commission, Joint Research Centre (2022). The advantages and powerful capabilities of ML justify its popularity; however, this does not mean the applicability of ML should not be critically questioned, since there is often a trade-off of interpretability and dependence on ML experts when employing this technology. The step of investigating whether ML really is the most appropriate method for a given task is rarely carried out in academic research. Although the suitability of ML vs. non-ML methods have been compared for generic data scenarios as well as some manufacturing scenarios, this differentiation in the context of improving manufacturing energy and resource efficiency is poorly understood (Bzdok et al., 2018; Lago et al., 2018; Makridakis et al., 2018).

Thus, to summarize, the core problem this paper investigates is the difficulty to determine when to use ML, or different types of ML, for manufacturing energy consumption modeling. The objective of this paper is to develop a methodology to answer the following research questions resulting from the problem statement:

- (1) How do the prediction errors of different manufacturing energy consumption modeling techniques compare when varying the model complexity and use of ML?
- (2) How do the performances of different manufacturing energy consumption modeling techniques compare

when different amounts of data are used to train the models?

- (3) How do the different manufacturing energy consumption modeling techniques compare in terms of required expertise and effort to build the model, as well as in terms of interpretability of the results?

The questions will be investigated using a comparison method developed by the authors, applied to an evaluation scenario of data from a compound feed manufacturing process. The originality of this work lies in the methodology the authors develop, which allows for systematic identification of an appropriate modeling technique and data amount to select for a given manufacturing process. The methodology poses a valuable contribution in the field of industrial engineering, in that it provides guidance on the proper application of ML, to practitioners wanting to model the energy consumption of a process; where significant data science and modeling expertise is currently needed to discern the best approach, the proposed methodology can be applied by those with less experience, in order to obtain an indication on the data and model type best suited for their case. An additional unique aspect of the method is that unlike existing approaches, this one is designed to be applicable across different manufacturing domains and processes, as will be shown in the following chapters.

Theoretical background and existing literature

In this section the authors summarize the fundamental concepts of modeling for manufacturing energy efficiency. Next a classification of modeling approaches ideally suited for this comparative study is proposed. The approach is derived from existing classifications, a selection of which are mentioned in this publication. Finally selected studies that have compared ML and non-ML modeling approaches are presented. An extensive systematic literature review was conducted around this topic, resulting in 60+ papers, however, since this article is not a review paper, only selected referenced and findings from the full literature review are presented.

Modeling for manufacturing energy efficiency

The purpose of employing modeling to improve manufacturing energy efficiency is to predict how parameters of the system affect the energy consumption of the system. In this context, the system can be an entire manufacturing plant, a process or a single machine. This is done by creating a model of the relation among system parameters (such as machine settings, process settings and input material characteristics) and the energy consumption. Further other output characteristics which are constrained, such as product quality and production time, are typically also included. Various methods

can be used to create such a model. Mathematical regression models can be calculated, including: statistical regressions such as linear or polynomial, classical ML regressions such as decision tree, support vector machine (SVM), XGBoost or deep ML models such as convolutional neural networks (CNN) or long short-term memory networks (LSTM). For examples of some of these approaches see Alvela Nieto et al. or Zhang and Ji (Alvela Nieto et al., 2019; Zhang & Ji, 2020). Fuzzy logic can be used to build a representation of the manufacturing system, as demonstrated by Lau et al. (2008). Or, a simulation of the system can be created, as evaluated by Thiede and Dietmair and Verl (Dietmair & Verl, 2009; Thiede, 2012). In this paper the authors propose a categorization of the various modeling methods available, presented in the next section. Once a model of the system has been created, this model is used to predict how changes in system parameters will influence energy consumption. This can provide insights on opportunities to improve energy efficiency.

A final step that is often done is to use the model in an optimization algorithm to find a process parameter combination, given certain constraints, which results in the lowest possible energy consumption. Weichert et al. list several examples of such approaches in their literature review of ML for the optimization of production processes (Weichert et al., 2019). For such multi-objective optimization problems genetic algorithms are widely used (Zolpakar et al., 2021). For sake of completion, such an optimization will be demonstrated at the end of this paper, though not elaborated upon in detail.

Modeling categories

As seen in "Modeling for manufacturing energy efficiency" section, there are a variety of modeling approaches available. For the comparative investigation planned for this study, the authors require a categorization with the following characteristics: should differentiate between ML and non-ML approaches, should differentiate based on model complexity as well as effort to implement and should be applicable to manufacturing energy consumption modeling. The authors found several existing categorization of modeling approaches that fulfill some of their requirements, but did not find a classification that was fully appropriate for the planned analysis. Many papers on manufacturing energy consumption modeling mention that modeling approaches can be divided into two broad categories: the category called model-based or "white-box", in which models are created using formulas or simulations of the physical processes underlying the system, and the category called empirical data-based, in which models are created using statistics or ML on data collected from the manufacturing system (Abdoune et al., 2023; Ljung & Glad, 2016). Some papers which focus only on machine tool energy consumption modeling also mention a third category called state-based modeling which uses the different

operational states (e.g. ramp-up, processing, idle, etc.) of a machine tool to predict the energy consumption (Abdoune et al., 2023; Li et al., 2022).

Occasionally sub-categories are provided, though these are typically focused only on ML or lack differentiation based on model complexity. For example Máša et al. categorize different levels of mathematical models used for manufacturing energy saving measures, with recommendations for applications of each (Máša et al., 2018). The classifications are based partially on complexity, however focus on modeling analyses and use cases when the entire system is to be modeled. The 5-level pyramid they define contains basic balance models at level 1, regression and balance models of system components at level 2 (using ML algorithms such as artificial neural networks, as well as regression models using ordinary least squares), static simulation models of the entire manufacturing system at level 3, dynamic (linear) mathematical models for system components at level 4 and model optimization at level 5. Levels 1, 2 and 3 were incorporated by the authors of this study, however, more differentiation was needed within level two since complexity can vary significantly among different ML models. Studies found by the authors, which differentiate ML models do so at a very high granularity, and typically do not explicitly differentiate based on model complexity. For example Lago et al. compare a variety of statistical models including ones with and without regressors, and a large selection of ML models including, neural networks, support vector regressors and ensemble models, totaling 27 different models. Though the groupings of models used by Lago et al. provide a starting point for defining different categories of models, the high number of categories and lack of differentiation based on complexity and implementation effort make it inappropriate for the comparison planned for this study (Lago et al., 2018). In their literature review, Renna and Materi provided a thorough overview of the mathematical methods used to improve energy efficiency in manufacturing systems, grouping analysis into the categories of exact numerical, approximation strategies, heuristics/meta heuristics, experimental and real time data analysis and simulation (Renna & Materi, 2021). This classification, though thorough, does not provide a basis for comparing complexity and effort required for the different categories.

Thus, the authors of this paper created a categorization based on adaptations from published approaches, as well as their own experience and thinking, as will be presented in "Method" section.

Existing comparative studies

While multiple manufacturing studies exist in which different ML approaches are compared, the authors found no manufacturing studies comparing ML and non-ML approaches.

However, some papers from other disciplines were found which do such a comparison. A selection is presented in this section.

Bzdok et al. come from the field of biology and compared a statistical approach (differential expression analysis) and a ML approach (random forest) by attempting to identify certain genes in a given RNA sequence (Bzdok et al., 2018). They ran the analysis with both methods and then compared which identified more genes correctly over the course of 1000 simulations. They found that the statistical approach performed almost as well as the ML approach. They concluded that this was part due to the pre-existing knowledge they had which helped them design a good statistical model, and in other part due to the relatively small number of features in the data set. They predicted that the difference in performance would be greater if the data set were more complex and they did not have the preexisting knowledge.

Lago et al. did a similar comparison in the field of energy and economics, by forecasting spot electricity prices (Lago et al., 2018). They compared a variety of statistical models including ones with and without regressors, and a large selection of ML models including, neural networks, support vector regressors and ensemble models. As mentioned in "Modeling categories" section, in total the detailed study compared 27 models and found that the ML models mostly performed better. The study presented in this paper will focus less on comparing such a wide range of specific models, but rather will aim to compare models of different classes of implementation complexity, to quantify the trade-off in performance and effort. The groupings of models used by Lago et al. however present a starting point for defining the aforementioned classes of complexity.

Makridakis et al. also did a comparison of forecasting ability, but applied to the M3 forecasting competition data set, which contains time series from five domains of demographics, micro- and macroeconomics, industry and finance (Makridakis et al., 2018). They tested eight popular families of ML models and eight traditional statistical ones. Interestingly they found that the statistical models performed significantly better than the ML models, while also being computationally less complex. It should be noted that one reason they give is that the ML models might have been over-fitted, which is arguably an issue that could be reduced with further feature engineering. They also raise the consideration that the data set has relatively few features (only 6), and thus may be "too simple" for ML. This is something the authors of the study presented in this paper plan to investigate. Makridakis et al. end with a call to researchers to do more empirical testing of ML forecasting methods, rather than only publishing studies in which ML forecasts are claimed to be satisfactory without comparing to simple statistical methods or benchmarks. Though not confined to

forecasting, this paper aims to answer precisely this call to action.

It should be noted that the studies listed in this section only categorize ML models as supervised or unsupervised.

Materials and methods

Method

The authors have developed a structured approach to compare the effectiveness of manufacturing process energy consumption modeling techniques of different complexity levels, with and without ML. The approach consists of three dimensions, outlined as following:

Method dimension 1: model dimension

Different types of ML and non-ML modeling techniques of varying degrees of complexity are compared (see blue text in Fig. 1). Since as discussed in "Modeling categories" section existing modeling classifications are not ideal for the planned comparative analysis in this paper, the authors define four modeling categories, adapted from different classifications in literature.

The categorization contains four general categories of models, with the purpose of creating a collectively exhaustive categorization that will allow for systematic comparison of different approaches. Such a categorization is important as it will allow for a systematic comparison of modeling approaches, within the application of manufacturing energy efficiency, based on model complexity and use of ML. Table 1 summarizes the categorization. The authors adopted the common two high level classifications typically seen in literature in "Modeling categories" section, "model-based" and "empirical", into their classification. "Model-based" was used largely used to define the category called "Complex Non-ML" and various approaches within the "empirical" family were divided among the other three categories of the authors' classification.

Each category is further defined in the following text:

Model category: simple non-ML Literature contains various, partially conflicting definitions of when a regression algorithm is considered to be a ML algorithm and when it is not. This ambiguity prompts the need to establish a working definition to avoid any potential confusion. Thus, in the context of this paper, linear regression, also known as ordinary least squares, is categorized as the primary non-ML regression algorithm. This is because it involves fitting data to a straight line and does not inherently involve "learning". Other modeling approaches that are considered non-ML in this thesis are not algorithms that fit models to data, unlike all the other

Model Comparison Methodology

Expected Outcomes

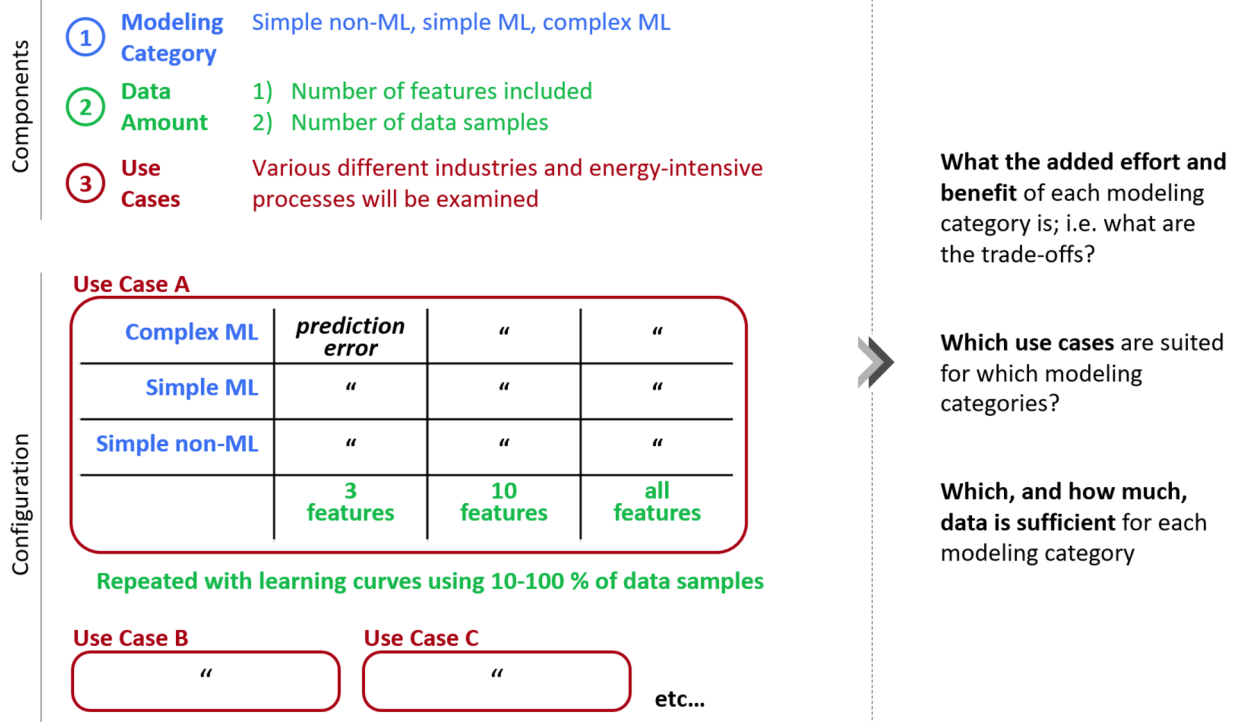


Fig. 1 Concept diagram of the model comparison approach to be used to answer the research questions (Color figure online)

Table 1 Categorization of modeling analyses, ranging in complexity and methodology, for improving energy efficiency of manufacturing systems

	Simple Non-ML	Complex Non-ML	Simple ML	Complex ML
Definition	Linear regression	Traditional manual modeling approaches that do not require ML	Classical ML models without hyperparameter tuning	Advanced/modern models with extensive hyperparameter tuning
Example analyses	<ul style="list-style-type: none"> Ordinary least squares linear regression 	<ul style="list-style-type: none"> Mechanistic process or system simulation Fuzzy logic modeling Physics-based model of the system 	<ul style="list-style-type: none"> Decision tree Random forest XGBoost SVM 	<ul style="list-style-type: none"> Deep learning models (multilayer perceptron, convolutional NN, recurrent NN, LSTM) Reinforcement learning models
Selected properties	<p>High interpretability and explainability; i.e. “white-box”</p> <p>Can be carried out in no/low code environments</p> <p>Requires less time to build than complex non-ML and ML</p> <p>Less data science expertise needed</p>	<p>High interpretability and explainability; i.e. “white-box”</p> <p>Can sometimes be carried out in no/low code environments</p> <p>Typically require most time to build out of all categories</p> <p>Little to no data science expertise needed, but high expertise in physics or simulation tools needed</p>	<p>Some interpretability and explainability; i.e. “grey-box”</p> <p>Can be carried out in no/low code environments</p> <p>Requires less time to build than complex non-ML and ML</p> <p>Some data science expertise needed</p>	<p>Almost no interpretability and explainability; i.e. “black-box”</p> <p>Cannot be effectively carried out in no/low code environments</p> <p>Requires more time to build than simple non-ML and ML</p> <p>Most data science expertise needed</p>

three categories in Table 1. Rather, they are manual modeling approaches that are created without the use of algorithmic methods. See model category “Complex Non-ML”. The definition of linear regression as utilized in this thesis can be found in the textbook of Hope, who notably also does not clearly state whether linear regression is ML or not. Linear regression is one of the simplest and most widely used method for assessing the relationships between continuous predictor and response variables (Hope, 2019). Its popularity can be attributed to its high interpretability and straightforward implementation.

Model category: complex non-ML Models in this category do not employ ML, but still demand a high level of expertise in the manufacturing system being modeled, as well as a deep understanding of complex modeling approaches. When it comes to modeling manufacturing systems and their energy efficiency without ML or linear regression, three modeling techniques are frequently employed. These techniques are mechanistic process or system simulation, fuzzy logic modeling, and physics-based modeling. It is important to note that all three of these modeling methods require customization and adaptation to suit the specific characteristics and requirements of the system being modeled. This customization is necessary to ensure that the resulting models accurately represent the intricacies of the given system.

At this point the authors provide justification for why the modeling category “Complex Non-ML” is omitted from the comparison methodology. Regarding simulations, as Thiede highlights, there are few to no commercial simulation tools that are designed to simulate energy consumptions throughout a given manufacturing process. This would mean the authors would have to find a very specific software or develop one themselves, as Thiede did (2012). On top of this, a thorough simulation can be expected to cost significant time and resources to create (Banks, 2010; Chung, 2004). As Mourtzis et al. found in their literature review of simulations in manufacturing, most commercial software are designed for specific manufacturing processes, meaning different software would potentially need to be used when modeling different manufacturing processes, as the authors intend to do (Mourtzis et al., 2014). Physics-based models are suited for low-uncertainty low-complexity systems, rather than high-uncertainty high-complexity systems such as manufacturing processes with multiple steps (Wang et al., 2022). A physics-based model which does not account for all physical dynamics in a system will likely be inaccurate (Erge & van Oort, 2022). Most examples of studies in which physics-based models are created for improving manufacturing energy efficiency, a model is made only for a single process step, e.g. a cutting, rotating, or thermal process step, and is used to fine-tune the process step (Avram & Xirouchakis, 2011; Cubillo et al., 2016; Imani Asrai et al.,

2018; Osara, 2019). Expanding the scope of such a model, would require deriving and adding equations for multiple physical phenomena, as well as interactions among these. The resulting increase in complexity of the system most often then cannot be modeled with physics equations, or requires extensive computational resources, not to mention much time and expertise to derive. Moreover, integrating the stochastic nature of a machining process, such as tool wear, into physics-based models poses a significant challenge (Sealy et al., 2016). Rather physics and data-driven hybrid models are sometimes used (Wang et al., 2022). Referring to the classification proposed in this paper, such an approach would be a combination of “Complex Non-ML” with “Simple ML” or “Complex ML”, although the authors will not consider category combinations. Fuzzy logic modeling quickly becomes very time consuming, when more parameters are added to the model. Lau et al. demonstrated a fuzzy logic approach to forecast overall energy consumption of a clothing production factory, and included only three parameters (daily total mass of finished products, total labor hours of operators in the plant in one day and the total running time of equipment) (Lau et al., 2008). When applied in the context of manufacturing, fuzzy logic can only practically be applied to consider the overall manufacturing system, rather than individual energy consuming process steps, as Thiede states (2012). The above-mentioned limitations led the authors to conclude that the “Complex Non-ML” modeling approaches (simulation, physics-based and fuzzy logic) are not well suited for this comparative study and will thus be excluded.

Model category: simple ML ML models in this category consist of all those which do not rely on neural architectures or layers. They encompass a range of classical ML methods, including Support Vector Machines, Stochastic Gradient Descent, Nearest Neighbors, Naive Bayes, Decision Trees and ensemble methods like Gradient Boosting and Random Forests. These algorithms are widely accessible through popular ML libraries such as scikit-learn (Pedregosa et al., 2011). One notable advantage is their applicability even when one has limited prior knowledge in ML. They can be effectively utilized, especially when employing no/low code solutions like RapidMiner and KNIME, making them accessible to a broader audience for various applications (Berthold et al., 2007; Mierswa & Ralf, 2023).

Model category: complex ML Deep learning (DL), which has witnessed a surge in popularity, stands out as a fundamental departure from classical learning methods and serves as the differentiator in this classification system. It is characterized as complex due to the black-box nature of its algorithms, as opposed to classical ML models, which offer some interpretability when using methods such as feature importance analysis, recursive feature elimination and decision

tree visualization to name some interpretability techniques. Effectively applying DL algorithms necessitates at least some grasp of the algorithm's hyperparameters and architecture. Moreover, proper development of a DL algorithm typically requires significantly more time than development of a classical ML algorithm. Building a robust model often requires multiple iterations. While AutoML has streamlined some aspects, it has not yet made DL as quick and straightforward to apply as classical ML algorithms.

At this point a note should be made on DL and its application on non-tabular data, such as image, text and sound. DL has found widespread acclaim in these applications, where it consistently outperforms classical ML methods. However, it is important to clarify that this study does not directly consider image, text or sound recognition applications for manufacturing process modeling. Instead, these applications may be employed to contribute additional tabular data points. For instance, an image recognition algorithm may assess product quality and provide results such as “good quality” or “poor quality” for a given product batch or timestamp, which is then integrated into a data table containing other process measurements. This table is subsequently used to train a model of the manufacturing system. Thus, in this study, DL is applied only to tabular data, which is an application area where the superiority of DL over classical ML is not self-evident and has not been extensively explored in previous research.

For the “Complex ML” category two common types of neural networks are primarily considered: multilayer perceptron (MLP) and convolutional neural network (CNN), both taken from the Keras Python library (Chollet, 2015). See the theoretical background chapter for details on the algorithms. The hyperparameters of the CNN models will be tuned manually by trial and error for each dataset, using TensorBoard, while the hyperparameters of the MLP models will be tuned using a large grid search including the following hyperparameters: batch size, epochs, dropout rate, number of layers, number of nodes between layers and optimizer.

Method dimension 2: use case dimension

The specific objective of each analysis will depend on the use case, i.e. application scenario, under consideration. The methodology presented in this study is intended to be used on use case in which the goal is to model a manufacturing system in order to understand the relation between the system parameters and the energy consumption. The motivation for modeling in these use cases is typically to use the model to identify parameter changes that could improve the energy efficiency (see red text in Fig. 1). The authors chose to limit the analyses to modeling techniques in order to allow for a direct quantitative comparison of the different analyses, namely the prediction error. This provides a basis on which

analysis can be compared. The long-term plan of the authors is to do these comparisons for various manufacturing energy use cases, in order to gain insights on which analyses are suited for which use cases. The methodology can be applied to any manufacturing domain in which manufacturing system parameters can be used to predict energy consumption and other relevant system outputs. Examples of energy intensive use cases the authors are working on, in addition to the one which will be presented in this paper, include: water extraction during gelatin production, car body drying in automobile manufacturing and waste material balling in the plastics recycling industry. This paper will present the result of the method applied to one use case. Future work will summarize the findings from multiple use cases.

Method dimension 3: data dimension

A deciding factor for data-driven sustainable manufacturing is having a good data basis. However, the amount of data, both in terms of number of features and amount of data samples needed is often unknown and answered with “the more the better”. The authors do not plan to answer this question with a definitive number, but aim to gain insights to differentiate the relative required data amounts across the modeling categories. To investigate how well different analyses work with different amounts of data, the data amount per use case will be varied in two ways: varying the number of features and varying the amount of data samples.

Firstly, the number of features, i.e. columns of data, will be systematically varied by once using a selection of 3, 10 and finally all features. The total number of features varies depending on the use case, but is typically approximately 30 to 60, based on past modeling the authors have done and examples from literature. The subsets are selected using recursive feature elimination (RFE). RFE is a ML feature selection/reduction algorithm which attempts to find the most influential features by starting with all features in the training dataset and systematically eliminating them until the specified number is reached (Kuhn & Johnson, 2016). It is known as a “wrapper” selection method because it involves training a ML model for every possible combination of features. To do this, the RFE algorithm takes a specified ML algorithm and iteratively removes and replaces each feature in the model, checking the change in prediction error each time. Features are then ranked by greatest reduction in prediction error. The least important feature is discarded and the model is refitted with the remaining features. This cycle is repeated until the specified number of features remain. The RFE class from the scikit-learn package is used in this study (Pedregosa et al., 2011). The authors choose RFE to select the feature subsets since it is an objective and robust selection method. Supervised feature selection can be done with either wrapper, filter

or embedded methods. Filter methods use statistical techniques to score the importance of each feature to the target. Embedded methods are built-in feature importance methods of certain ML algorithms, and can be retrieved after a model is calculated. Certain filter methods such as Spearman's rank coefficient, could technically have been used, but would be less suited for this study since most filter methods are not intended to be used to select a specific number of features or are complex to implement for multivariate (multiple targets) regression models. Intrinsic/embedded methods were not used since they are only available for certain ML algorithms. An unsupervised feature selection method, which disregards the target variable, was not chosen because the target variables are specified for all modeling scenarios considered in this study. For more details on feature selection see Kuhn and Johnson (2016). Not only will this provide insight on the relation between number of features and model performance, but it may also reveal whether certain model categories perform better or worse with more or less features. In the context of modeling manufacturing process energy consumption, feature selection can also reveal which features have the most impact on the energy consumption of an analyzed process.

Secondly, learning curves will be used to compare the changes in model accuracy for different amounts of data samples, i.e. rows of data, from 10 to 100% of all available data. Learning curves consist of lines showing the training and validation errors of a model, for different amounts of training data. Analyzing these curves can provide various insights such as whether a model is overfitting or under fitting, whether more parameters may improve the performance or whether more training data could improve the performance. The authors will aim to have at least 6 months' worth of recorded data for each use case. The sampling rate will vary depending on the manufacturing system under investigation, so the absolute amount of data samples will also vary. The authors will decide on how to address this once multiple use cases have been investigated. Varying the amount of data samples will provide insights for the given use case on how much training data is needed for the different modeling approaches to obtain a given model (see green text in Fig. 1).

Calculation of prediction error comparison

The metric used for comparing the prediction error is the mean absolute error (MAE). Note that the term "accuracy" is purposely avoided since in the context of prediction, accuracy is the rate of correct predictions, which is relevant for classification problems, rather than regression problems, where the amount by which a prediction errs from the correct value is of interest. MAE is chosen because, after normalizing all data from zero to one, it allows for interpretation of the MAE as a percentage of the predicted data, which is more intuitive

to interpret; e.g. a manufacturer can better estimate what an acceptable percentage deviation in energy consumption of their process is, than an acceptable root squared error. The coefficient of determination (R^2) and root mean squared error (RMSE) however could also easily be calculated during execution of the methodology. Rather than simply stating that the error of one model is greater than that of the other, the difference in performance will be compared within the context of the labels. This is an important distinction to make when considering more than only the error of prediction models. In this study, the complexity of the models is also considered, so if for example, the error of a "Complex ML" model is not noteworthy lower than that of a "Simple ML" model, the conclusion could be made that the added complexity provides no prediction advantage and is only less interpretable and more cumbersome to create. Model prediction comparisons in all previous literature that the authors encountered either only check which model has a lower prediction error, or only test whether the difference in model performance is statistically significant by using the modified paired Student's test (aka t-test) combined with 2×5 cross validation, as proposed by (Dietterich, 1998). The Akaike and Bayesian Information Criteria have this same limitation, though they take the number of model parameters (i.e. features) into consideration. These approaches do not provide information on whether the prediction error difference in two models is large enough to warrant using one model over the other, rather only whether the difference is statistically significant. The authors of this paper use the logical assumption that the more the values to be predicted by the models vary, the less noteworthy a given difference in prediction error between two models is. Accordingly, in the so-called noteworthiness test presented in this paper the difference in MAE between models will be compared to the average standard deviation of the values being predicted (i.e. the labels). Note that the authors consciously avoid using the term "significant" since this has a specific meaning in the context of statistics; thus they use the term "noteworthy". The average of the standard deviation of all the labels is taken, since the labels vary among use cases; e.g. for one use case, electricity consumption and a quality measurement could be used as labels, while for another use case gas consumption and temperature could be used as labels. Arguably, the analysis could be made more accurate by giving different weights to the standard deviations of different classes of labels, e.g. energy-related labels are weighed more than quality-related labels, however, this would require individual evaluation of each use case, and would be difficult to implement in an objective manner. Thus, the averaging approach was chosen by the authors, to ensure the methodology can be applied consistently across different manufacturing energy scenarios. When the data is initially normalized to a range of 0 to 1, the average standard deviation and MAE are unit-less and can be divided by each

other to obtain a percentage. When the difference in MAE is above a threshold of 5% (defined by the authors) of the average standard deviation, it is considered noteworthy. 5% is selected since this is also the commonly chosen threshold value, α , for a t-test. A user of the methodology could however select any threshold value deemed appropriate. The noteworthiness test can be summarized as follows:

Assuming all data is initially normalized,

First calculate standard deviation $\sigma_{label\ n}$, of each label n :

$$\sigma_{label\ n} = \sqrt{\frac{\sum_{i=1}^I (y_{i,n} - \mu_n)^2}{I_n}},$$

where

I = total number of data instances of the label n $y_{i,n}$

= value of each instance of label n μ_n

= mean value of label n .

Then calculate the average standard deviation of all labels $\overline{\sigma_{labels}}$:

$$\overline{\sigma_{labels}} = \frac{\sum_{n=1}^N \sigma_{label\ n}}{N},$$

where

N = total number of labels.

Then calculate the absolute difference in MAE of two models being compared. Ensuring that the data the models were trained and tested on is the same data used to calculate the standard deviations.

$$|\Delta_{MAE}| = |MAE_{model\ A} - MAE_{model\ B}|.$$

Then calculate what fraction of the average label standard deviation the difference in MAE is, and whether this is above the threshold value T :

$$|\Delta_{MAE}| \text{ is considered } \dots \begin{cases} \text{not noteworthy, if } \frac{|\Delta_{MAE}|}{\overline{\sigma_{labels}}} < T, \\ \text{noteworthy, if } \frac{|\Delta_{MAE}|}{\overline{\sigma_{labels}}} \geq T, \end{cases}$$

Cumulatively the methodology described in "Method" section is expected to provide an empirical basis for answering the three research questions presented in this paper, and is summarized graphically in Fig. 1.

Evaluation use case

In this section the scenario from which data was taken to evaluate the methodology is described. The data was collected

over the course of the KIPro project (03ET1265A). The company for the scenario is a compound feed manufacturer, who produces feed for livestock. For an extensive explanation of the compound feed manufacturing process see Kersten et al. (2010). The company produces dozens of feed types, each with different recipes. Multiple processes with many parameters are involved to transform the raw input materials into finished feed. The process that this study investigated was the pelleting process. See Fig. 2. Here a mixed mass of flour and other ingredients is heated with steam and pressed through a matrix to bring the product into its final form as pellets, for easier transportation, storage and feeding. The pelleting process is one of the most energy consuming processes of feed manufacturing since heated steam is needed, as well as electrical energy for moving the press.

Sensors throughout the pelleting process provide rich data. The data set used for the evaluation consists of approx. 1100 data samples from the pelleting process, each representing one production batch. The data set consists of 67 features (input ingredients and process parameters) and 4 labels (maximum machine utilization, electricity consumption, steam consumption and production duration). The model with features and labels is illustrated in Fig. 3. The objective of the analysis in this scenario was to model the relation between energy consumption relevant metrics and production parameters, in order to:

- (1) Better understand the effect certain production parameters have on energy consumption.
- (2) Be able to perform an optimization, to minimize energy consumption while avoiding overutilization of the press machine.

Overview of the analysis conducted

This section details what models were created for each category, and how the data basis was varied. As justified in "Method" section, the category "Complex Non-ML" was omitted from the methodology due to lack of applicability.

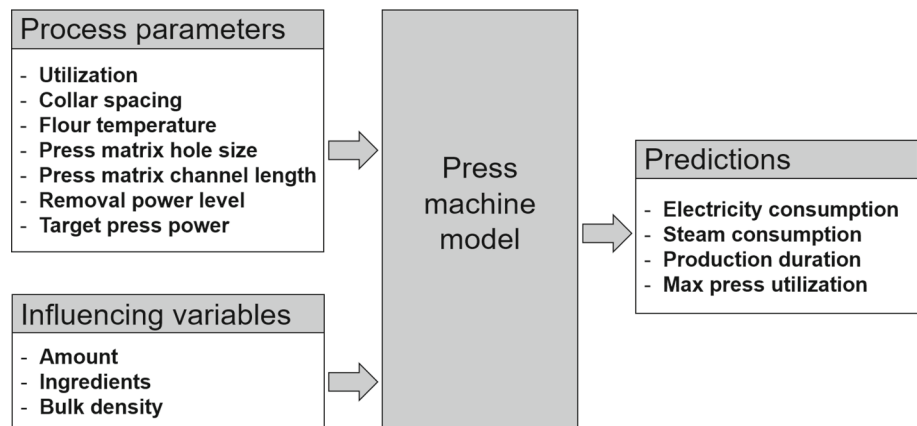
For the "Simple Non-ML" category, an ordinary least squares linear regression taken from the scikit-learn Python library was used (Pedregosa et al., 2011). For the "Simple ML" analysis the common classical ML algorithms decision tree regression, random forest regression and XGBoost regression were compared. Default configuration from the scikit-learn package were used. No hyperparameter tuning was done, and the models were directly fitted to training data. The XGBoost regression consistently resulted in the lowest prediction error and was thus selected as the model to use for this category. For the "Complex ML" category a multi-layer perceptron (MLP) model from the Keras Python library was used (Chollet, 2015). The hyperparameters of the model



Fig. 2 Images of the press machine modeled in this study. The machine is used in the pelleting process of compound feed production. **a** Shows the complete press machine, and **b** the inner matrix where the compound

feed mixture is being pressed through, in the form of dried pellets. Austing Mischfutterwerk GmbH

Fig. 3 Conceptual representation of the model built, including inputs and outputs



were tuned manually by trial and error, using TensorBoard. After tuning, the following hyperparameters were selected: 1000 epochs, batch size of 100, 6 network layers, each layer with 30 to 80 neurons, dropout rate of 0.1 between each layer and Adam as the optimizer.

All data was normalized to a value between 0 and 1. All models underwent fourfold shuffled cross validation. The features on which the models were trained was varied three times. First using three variable parameters selected with RFE: target machine utilization, collar spacing and flour temperature. Secondly using 10 features selected with RFE, which resulted in: the respective amounts of 4 specific ingredients, the overall amount of feed, flour temperature, output rate, press canal width, target machine power level and target machine utilization. Thirdly using all 67 features.

As defined in the method section, the data sample amount variation was done by creating learning curves. For the sklearn models, the learning curves were created using the scikit-learn learning curve function. For the MLP neural network the authors built a learning curve generation function.

The learning curves for all models included four fold shuffled cross validation.

Results and discussion

The results from the analysis described in Sect. 2.3 are presented in this section as follows: firstly, a comparison of the prediction error of the different model and number of features combinations; secondly, a comparison of how each model and number of features combination behaves with varying amounts of training data; thirdly, a qualitative comparison of the effort required for each of the three model categories; and lastly, a qualitative comparison of the interpretability of each of the three models. These results provide an example of the insights that can be obtained when applying the comparison methodology. When applied to their own manufacturing processes, the comparison results could be used by a decision maker at a manufacturing plant to better answer the following questions:

Table 2 Standard deviation of the four labels

Label	Standard deviation
$\sigma_{electricity\ consumption}$	0.114
$\sigma_{utilization}$	0.129
$\sigma_{steam\ consumption}$	0.096
$\sigma_{production\ duration}$	0.226
σ_{labels}	0.141

- (1) What class of model is worth implementing?
- (2) How many data points should be continued to be measured in the long term?
- (3) How many more data samples need to be collected whenever changes are made to the system in the future?

Prediction error comparison

First the prediction error of the models is compared, following the noteworthiness testing method described in "Method" section:

Calculate standard deviation of each label, and then the average (Table 2):

Rearranging the formula to provide a minimum MAE, which is still considered noteworthy

$$|\Delta_{MAE}| = \begin{cases} \text{not noteworthy, if } |\Delta_{MAE}| < T\overline{\sigma_{labels}}, \\ \text{noteworthy, if } |\Delta_{MAE}| \geq T\overline{\sigma_{labels}}. \end{cases}$$

Using the above formula, and the following values for the threshold and average standard deviation:

$$T = 0.0500, \overline{\sigma_{labels}} = 0.141,$$

$$|\Delta_{MAE}| = \begin{cases} \text{not noteworthy, if } |\Delta_{MAE}| < 0.00705, \\ \text{noteworthy, if } |\Delta_{MAE}| \geq 0.00705. \end{cases}$$

Thus, for all the predictions shown in this section, only a difference in MAE of more than 0.00705 should be considered significance. Figure 4 uses this value to indicate whether the prediction error of adjacent model and number of feature combinations are noteworthy:

Two conclusions can be made from the results above: firstly, the "Complex ML" model never has a noteworthy lower prediction error than the "Simple ML" model, and secondly, increasing the number of included features from the 10 most influential (as determined through RFE) to all 67 features never reduces the prediction error of any model by a noteworthy amount. Further prediction metrics can be found in "Appendix 1".

Data sample amount comparison

Next, the behavior of each model and number of features combination with different amounts of training data samples is investigated, using learning curves, as described in "Method" section. The results are summarized in Fig. 5. The individual learning curves can be found in "Appendix 2".

Reviewing the results one can see that as expected, as more features are included in the models, more data samples are needed to maximize the model performance. This is the case for all categories. The "Complex ML" model is the only one, which seems to be able to accommodate all 67 features without overfitting. This is in accordance with the typical assumption that more complex ML models are better suited for handling larger numbers of features. Interestingly, there is no major difference between the "Simple ML" and "Complex ML" models in the number of data samples needed to maximize performance; both the 10 feature and 67 feature scenarios the performance of the models plateau at approximately the same number of samples. This is counter to the typical assumption that more complex ML models require more data than simpler models.

Effort comparison

Next, the relative difference in effort needed to create the model for each category is compared. The "Simple ML and non-ML" analyses were completed in less than an hour. Compared to the "Simple ML and non-ML" analyses, the "Complex ML" analysis required significantly more effort. The same effort was needed for the data preprocessing, however the modeling process was significantly more extensive. Here the hyperparameters of a DL algorithm had to be tuned, which requires data science expertise and time. For the model used in this study circa one full working day was needed for the first version of the model, and then an additional day for further tuning. It should be noted that auto ML techniques are becoming increasingly sophisticated in being able to automate hyperparameter tuning, however are currently still not able to achieve this fully for "Complex ML" models (Karmaker et al., 2022).

The exact time and effort required is of course dependent on the experience of the analyst, so during this comparison focus should be placed on the relative difference in required effort among the model classes. In general, the greater the skill set of the analyst, the smaller the difference in effort among model classes is. The authors however assume that a difference in effort between "Complex ML" and "Simple ML and non-ML" will always be present, since more steps are needed in developing models in the former category.

		Performance (MAE)			Advantages	Disadvantages
Modeling Analyses	Complex ML	0.077	→ 0.044	0.042	• Highest performance for two of three data scenarios, though by a very small margin	• Most preparation needed • Little to no interpretability
	Simple ML	0.072	→ 0.048	0.043	• Close to best performance • Little preparation needed • Some interpretability through feature importance function	• Limited interpretability
	Simple Non-ML	0.083	→ 0.057	0.051	• High interpretability (can use beta coefficients to understand effects of each parameter) • Little preparation needed	• Lowest performance
		3 features	10 features	67 features		

Features Included in Model

Fig. 4 Mean absolute error (MAE) for each model and data amount combination, including indication whether difference in error is noteworthy

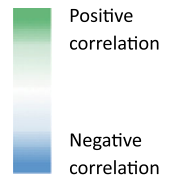
		Performance			Observations
Modeling Analyses	Complex ML	No improvement after ~500 data samples. High error. More features are needed, or a more complex model.	No improvement after ~600 data samples. Moderate error.	No improvement after ~900 data samples. Moderate error.	• 10 feature performance maximized with only ~600 data samples.
	Simple ML	No improvement after ~700 data samples. High error. More features are needed, or a more complex model.	No improvement after ~700 data samples. Model is overfitting. Moderate error.	Model is overfitting, but performance could improve slightly with > 900 data samples. Moderate error.	• Despite overfitting, model achieves good performance. • Model over fits regardless of amount of data samples.
	Simple k Non-ML	No improvement after ~200 data samples sufficient. High error. More features are needed, or a more complex model.	No improvement after ~500 data samples. Moderate error. More features are needed, or a more complex model	High error regardless of numbers of data samples. More complex model is needed.	• 10 feature performance maximized with only ~500 data samples, though performance is poor. • Model performance relatively poor for all data amounts suggesting model is inappropriate.
		3 features	10 features	67 features	

of Features Included in Model

Fig. 5 Description of behavior of each model and number of features combination for different amounts of training data samples

Table 3 β Coefficients of linear regression model with three features

	Max utilization	Electricity consumption	Steam consumption	Production duration
Utilization	0.04	0.04	0.02	0.01
Collar spacing	0.12	0.13	-0.05	-0.09
Flour temperature	-0.43	-0.05	0.36	-0.18



Positive correlation

Negative correlation

Interpretability comparison

The interpretability of each model is a differentiating factor in the classification system of complex and simple in this paper. The “Simple non-ML” model has the highest interpretability. By analyzing the β coefficients of the linear regression model, which indicate the incremental change each independent variable has on the dependent variables, one can directly see the weight each feature has on the model (if all data is normalized to the same scale). Though the analysis becomes more difficult to read as the number of features increases. As an example, the beta coefficients for the 3-feature scenario are shown in Table 3.

From Table 3, one can quickly interpret that reducing flour temperature significantly reduces steam consumption, while slightly increasing electricity consumption and significantly increasing max utilization. And that reducing collar spacing and utilization reduces electricity consumption, while slightly increasing steam consumption. If steam consumption were converted from mass to energy, then the tradeoff described above could be compared and a decision could be made whether to increase or decrease each parameter in order to save overall energy consumption. The constraints of max utilization and production duration would of course need to be considered. Dehning et al. demonstrate the same approach to assess the influence of different factors on the energy intensity of a car manufacturing plant (Dehning et al., 2017).

The XGBoost “Simple non-ML” model is based on a complex ensemble learning approach and thus does not provide an interpretable equation, such as the linear regression does. However the XGBoost Python library provides a built in function that provides a ranking of the influence of each feature in the model. By nature of how boosted decision tree algorithms, such as XGBoost, model data, this ranking is calculated during model training and is readily available. For details on how the ranking is calculated see (Hastie et al., 2009). For example, for the 67-feature scenario the 15 most influential features can be easily plotted with one function call. Thus, while it is not possible to easily interpret how the model makes certain predictions, it becomes less black box by understanding the feature importance.

The DL “Complex ML” model used in this study is black box and has little to no interpretability.

In summary, taking into consideration the type and quality of insights obtained from the different analysis, as well as the effort required for the analysis, the results of applying the comparison methodology suggests that the “Simple ML” approach using the 10 most influential features and ~900 data samples, strikes a good balance between prediction error, required effort and interpretability. This outcome would be valuable to a manufacturer, because they would have now obtained an indication that further developing and implementing a complex DL ML model, is not necessary, and that they could use a simple XGBoost model instead. This would reduce the manufacturer’s dependency upon data science expertise, and allow them to use a more interpretable model. Additionally, the manufacturer would know that it is sufficient to use only 10 of the 67 data points to model the energy consumption, and that data from approximately 700 production batches would be sufficient for retraining the model in the future, should the production process change.

Outlook: using the model for process parameters optimization

When a model of a manufacturing process is available, it provides valuable insight into the relationship between process parameters and energy consumption. However, a model can add more value when used in an optimization to identify the specific parameter settings that would generate the most energy savings.

Although the topic of optimization is not the focus of this paper, for completeness, an optimization is presented in this section to demonstrate the added value of a well-built model for achieving energy savings in production. Additionally this novel optimization approach developed by the authors can be applied to other energy efficiency modeling use cases. In this section, the previously discussed “Simple ML” model (XGBoost) with 10 features is used as the optimization function within the optimization algorithm. That is, the optimization algorithm uses the model to predict the energy consumption for a certain combination of parameters,

then makes small adjustments to the parameters and then uses the model to again make a prediction using the changed parameters. These steps are repeated by the algorithm until the prediction converges to a minimum or maximum value. “Differential evolution” from the library “scipy.optimize” was used as the optimization algorithm (Pedregosa et al., 2011).

Since the ML model models more than one output, and since further constraints besides a quantity to be minimized are relevant for the optimization objective, a rating function was defined based on the model outputs, which calculates a rating for each instance. The goal of the optimization is to minimize this rating. For the use case in this paper, the following simplified approach was used, which consists of minimizing the energy costs during pelletizing and considers a maximum press machine utilization. Energy costs are calculated as follows:

$$\begin{aligned} \text{Energy costs} = & (\text{electricity per ton})(\text{electricity price}) \\ & + (\text{steam consumption per ton}) \\ & (\text{gas required to produce steam})(\text{gas price}). \end{aligned}$$

The evaluation function is defined as follows:

$$\text{Rating} = \begin{cases} ((\max \text{ utilization} - \text{limit}) + 1)(\text{energy costs}), & \text{if } \max \text{ utilization} > \text{limit}, \\ \text{energy costs}, & \text{if } \max \text{ utilization} \leq \text{limit}. \end{cases}$$

In the case presented in this paper, the limit was set to 100. This ensured that when the maximum permitted utilization of 100 was exceeded, the penalty term provided for an increase in the rating. For utilization values smaller than 100, the rating corresponded to the energy costs that are to be minimized.

In addition, another penalty term was added, which aimed to limit the available parameter combinations. This is to prevent parameter combinations from being proposed by the optimizer, where the model cannot make a reliable prediction due to a lack of training data. These parameter combinations would in reality likely not result in the values that the model predicts. In the optimization function used in this paper, it is only possible to specify a left and a right limit value for the parameters to be optimized. Thus a rectangular (with two parameters) or cuboid (with three parameters) parameter search space is spanned. One can adjust these limits to the minimum and maximum values of the model training data set to avoid that the optimizer works on an extrapolation of the model and suggests values that are far from those occurring in the training data set. However, after initial testing, it became clear that not all conceivable parameter combinations were represented. The result was that parameter combinations far away from “known” parameter combinations of the training

data set were often suggested by the optimizer. The reason is that the model predicted low energy costs for these combinations. This is an issue however, since the model can likely not make accurate predictions with these unknown parameter combinations.

To avoid this, the authors created a novel approach. The approach uses a search tree and a prior scaling of the parameter dimensions by a maximum allowed distance between known parameter values. This is then used as an additional penalty term in the optimization. The scikit-learn “BallTree” library is used for this purpose (Pedregosa et al., 2011). In the approach, at least 10 training examples must be within an acceptable defined distance from the current parameter combination. During the optimization process, in addition to the rating predicted by the model, the distances between the given parameter combination and the nearest 10 parameter combinations from the (scaled) training data set are calculated. If one of the 10 nearest neighbors has a distance > 1 , another penalty term is added to the optimization function. This further narrows the available parameter space following the training data distribution and reduces the likelihood of an unreliable model prediction when calculating the optimal parameter combination.

The parameters used for optimization were press target utilization, press collar spacing and flour temperature, which are adjustable on the pelleting machine. After performing the RFE to select the most important 10 features, these parameters were included. This optimization was performed the same test data set, of 239 production batches, as used to test the trained model. As a result, the following plots of energy costs, machine utilization and energy consumption each show the model estimates for all 239 production batches.

From Fig. 6, it is clear that due to the parameters chosen by the optimizer, the energy costs are expected to be lower than the expected costs for the original parameters.

Average cost decrease per batch = 5.3%.

Figure 7 shows that the utilization boundary condition is adhered to after the optimization; the expected values for the maximum utilization for the optimized parameters are below the limit of 100% in almost all cases. Press utilization is the maximum measured power load of the press during a batch production as a percentage of its rated maximum load. Utilization of the press beyond 100% reduces the machine life span and causes malfunctions requiring production halts, which in turn waste energy due to otherwise unnecessary process ramp-ups.

Overutilization ($> 100\%$) occurrences decrease for the 239 batches = 21.0%.

Figure 8 shows that the energy consumption is reduced for almost every batch.

Average energy decrease per batch = 6.0%.

Overall, the optimized parameters achieved an expected average saving of 5.3% in energy costs compared with the original parameters, and a 6.0% average decrease in total energy consumption. Additionally, overutilization of the press machine is reduced by 21.0%.

As a final step before accepting the results, it is necessary to implement the proposed process parameters in the real-life process and measure the actual reduction in energy consumption. This is a necessary step since both the optimization and the evaluation are based on the ML model. Thus, the calculated improvements may in part be due to inaccuracies in the model prediction and may be less or greater in practice. This scenario is a possible application of energy consumption modeling to increase energy efficiency of a

manufacturing process, and illustrates the efficiency gains that can be achieved using an optimization approach.

Conclusion

Summary

This paper demonstrates a novel method for investigating how complex of a prediction model is needed to obtain reliable predictions of the energy consumption of a manufacturing process. The authors define four categories of analysis ranging in complexity and usage of ML. A comparison method is constructed around these categories so that a systematic evaluation on the effort, model performance and required data for different modeling approaches can be made. This method is tested on a use case from a compound feed manufacturer and results in the conclusion that a “Simple ML” model using the 10 most influential features and ~ 700 data samples, is sufficient to achieve a useable prediction

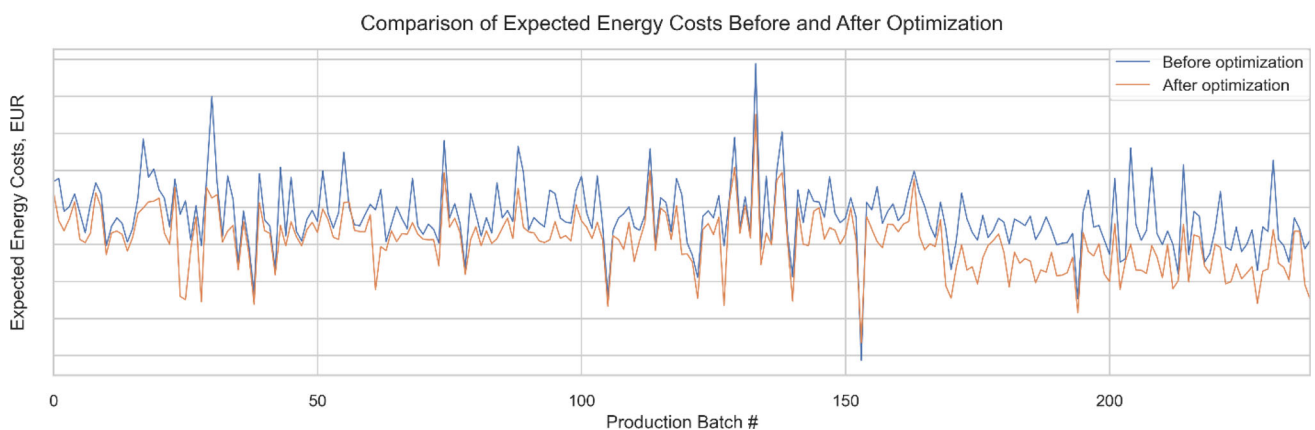


Fig. 6 Comparison of predicted energy costs before and after optimization. Prediction based on the actual parameter combinations in blue, and the prediction based on the optimized parameters in orange. Y-axis values omitted due to confidentiality (Color figure online)

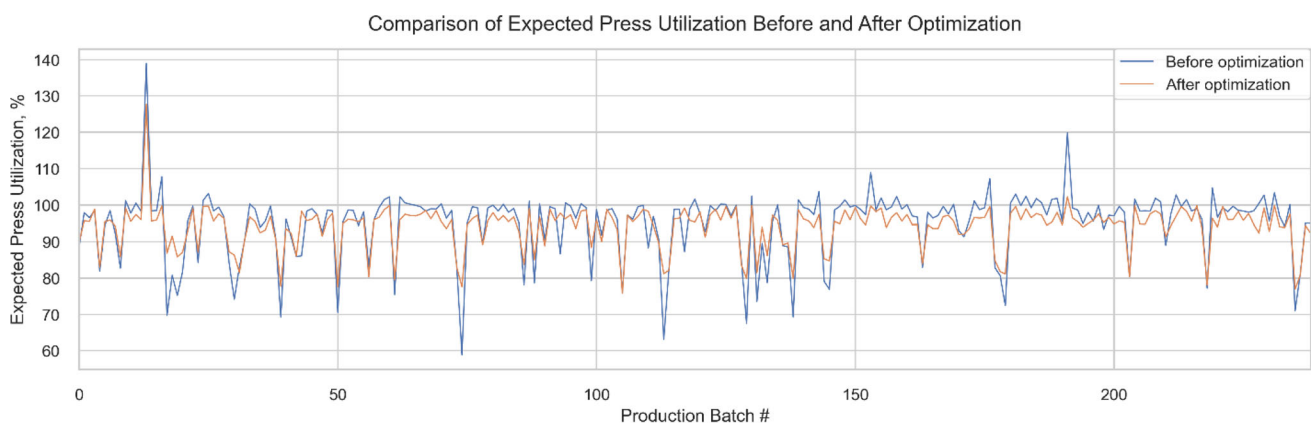


Fig. 7 Comparison of predicted machine utilization before and after optimization. Prediction based on the actual parameter combinations in blue, and the prediction based on the optimized parameters in orange (Color figure online)

while requiring less effort and providing more interpretability than a “Complex ML” model. To demonstrate the practical value of modeling the energy consumption a manufacturing process, a novel optimization approach is applied using the identified model. Energy consumption savings of 6% are predicted to be achievable if the optimal process parameter settings are implemented.

Upon revisiting, it is clear that the developed methodology has the capability to answer the research questions posed at the start of the study, for a given manufacturing process. For the manufacturing process used in the evaluation, and answers to the questions can be summarized as follows:

- (1) How do the prediction errors of different manufacturing energy consumption modeling techniques compare when varying the model complexity and use of ML?
By defining different categories of modeling techniques based on complexity and use of ML, the authors were able to make a structured comparison. For the conducted experiment, increasing model complexity resulted in lower prediction errors, however with significantly diminishing returns.

- (2) How do the performances of different manufacturing energy consumption modeling techniques compare when different amounts of data are used to train the models?

The authors were able to investigate the effects of data amount on the different models, by systematically varying the number of features, as well as the number of data samples, used to build the models. For the investigated use case, adding more features improved the performance of all models, but with diminishing returns; only the increase from 3 to 10 features reduced prediction error by a noteworthy amount. Models with

more features required more data samples. As expected, the “Simple non-ML” model required the least data samples. Unexpectedly, for the 10 and 67 feature scenarios, the “Simple-” and “Complex ML” models required approximately the same number of data samples to minimize prediction error.

- (3) How do the different manufacturing energy consumption modeling techniques compare in terms of required expertise and effort to build the model, as well as in terms of interpretability of the results?

The two “Simple” models required a similar effort to build, and significantly less than the “Complex ML” model. The “Simple non-ML” model (linear regression) was the most interpretable. The two ML models were less interpretable, though the “Simple ML” model (XGBoost) had at some interpretability, as opposed to the “Complex ML” model (MLP).

Limitations

One limitation of the work presented here is that the “Complex non-ML” category was omitted in the analysis, due to applicability. Including this category would provide a more comprehensive comparison; however the comparison would then only be applicable to a smaller selection of use cases. The methodology is purposely constructed to be as simple as possible, while still providing useful insights. With that being said, a second limitation is that the methodology does not differentiate modeling approaches beyond the four categories. There are dozens of different models within each of the four categories, which may differ slightly from each other in terms of complexity, interpretability and required

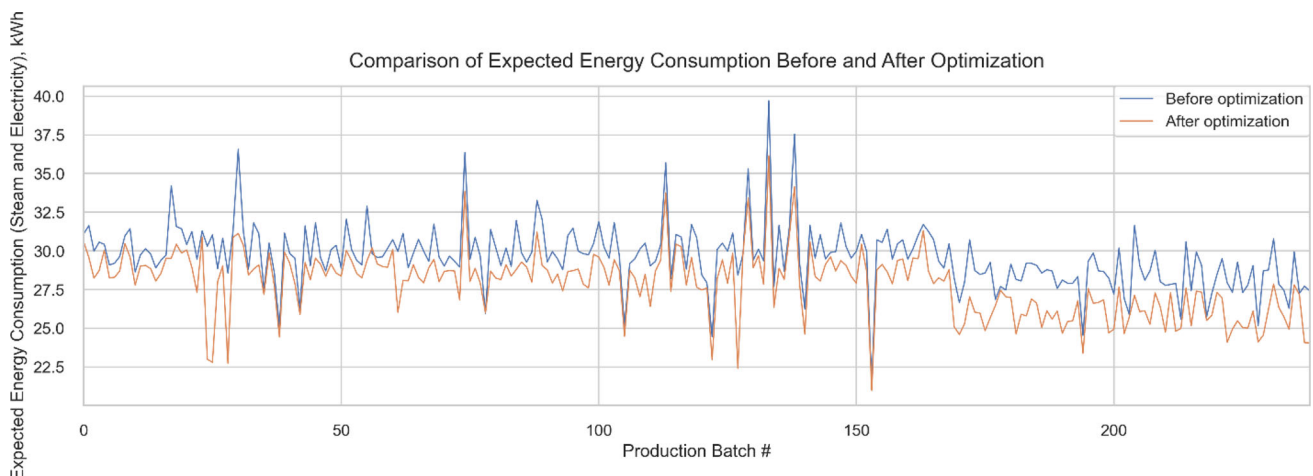


Fig. 8 Comparison of predicted total energy consumption before and after optimization. Prediction based on the actual parameter combinations in blue, and the prediction based on the optimized parameters in orange (Color figure online)

effort to build. An improvement could be made by defining sub-categories in order to differentiate even more among modeling approaches.

Contribution and future research

The primary contribution of this paper is a structured replicable model comparison methodology, which enables the systematic identification of an appropriate model and amount of data for a given manufacturing energy consumption modeling scenario. The originality of the methodology lies in its manufacturer-oriented design, applicability to various manufacturing scenarios and systematic structure. The novel tree based optimization approach is a further contribution of this paper. The comparison methodology can provide valuable insights in the fields of industrial engineering and data analytics because such differentiation can help manufacturers and practitioners at the start of a modeling initiative decide which approach to take and to estimate the resources to invest. Thereby facilitating manufacturers in improving the energy efficiency of their processes. This contribution relates to Sustainability Development Goal #9, “Build resilient infrastructure, promote inclusive and sustainable industrialization

and foster innovation” (UN General Assembly, 2015). The comparison methodology is designed to be generic and applicable to a variety of manufacturing processes and industries, rather than being developed for only a specific process. Once the comparison methodology has been applied to further use cases, the authors expect to be able to generalize which model category and data amounts are most appropriate for different types of manufacturing energy modeling use cases. Then it would not be necessary to execute the full comparison methodology for a given use case each time, in order to obtain an indication which model class and data amount would likely be sufficient. These insights would be even more widely accessible among practitioners, than if they would first need to be calculated each time.

Appendix

Appendix 1: Model prediction performance, with further metrics

See Table 4.

Table 4 Model prediction performance, with further metrics

Modeling Analyses	Performance		
	Complex ML		
	$R^2 = 0.43$ MAE = 0.077 RMSE = 0.114	$R^2 = 0.70$ MAE = 0.044 RMSE = 0.072	$R^2 = 0.72$ MAE = 0.042 RMSE = 0.072
	Simple ML		
Simple Non-ML	$R^2 = 0.47$ MAE = 0.072 RMSE = 0.104	$R^2 = 0.62$ MAE = 0.048 RMSE = 0.080	$R^2 = 0.68$ MAE = 0.043 RMSE = 0.073
	$R^2 = 0.36$ MAE = 0.083 RMSE = 0.114	$R^2 = 0.63$ MAE = 0.057 RMSE = 0.083	$R^2 = 0.66$ MAE = 0.051 RMSE = 0.080
# of Features Included in Model			
	3 features	10 features (RFE)	67 features

Appendix 2: Learning curves

See Figs. 9, 10, and 11.

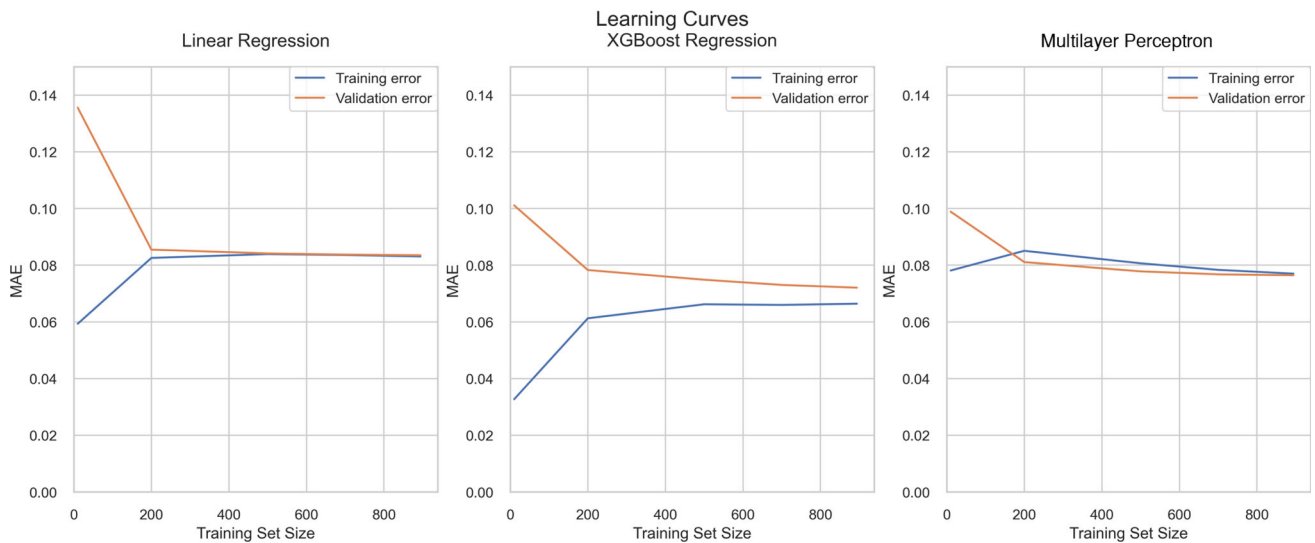


Fig. 9 Learning curves for three features scenario. ~ 200 Data samples sufficient for linear regression. > 900 for XGBoost. ~ 500 for CNN. MAE > 5% should be considered high. Small gap indicates low variance. High training error indicates high bias for all models. Thus, more

data samples will not help, but rather more features are needed, or a more complex model, since current models are under fitting the data

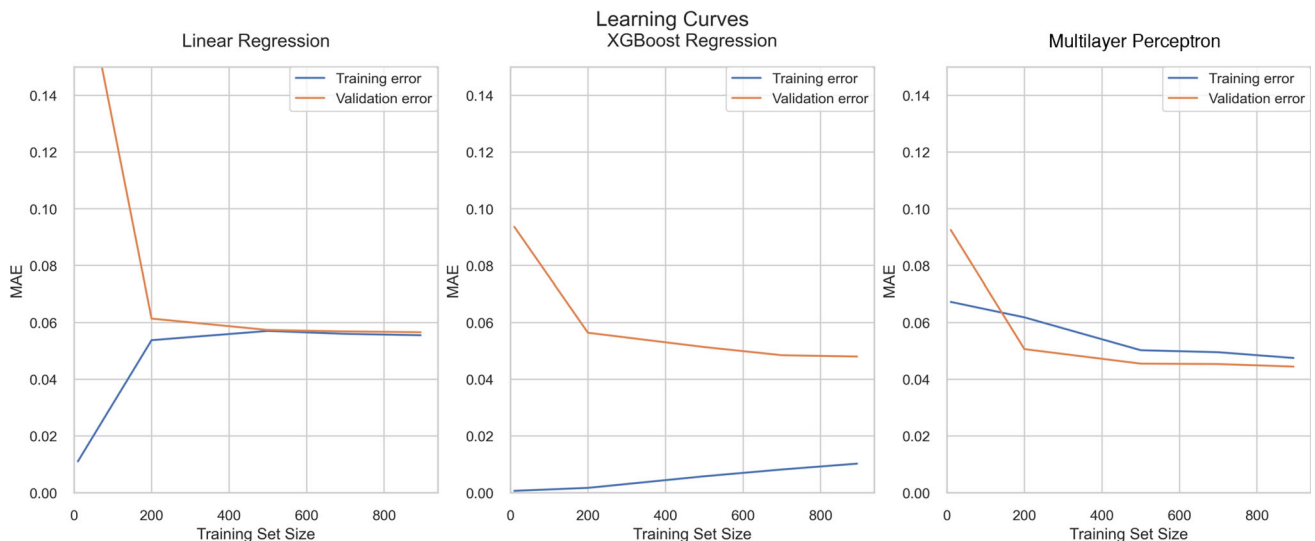


Fig. 10 Learning curves for 10 features scenario. Linear regression model has high bias and low variance, and thus if the irreducible error has not been reached, the model either requires more features or is not complex enough a model for the process. After 500 data samples the model does not improve. Low training error for XGBoost indicates low

bias. Large gap indicates high variance, and that the XGBoost model is overfitting. Validation curve is plateaued, so adding more data samples would likely not improve the model. CNN error lines have also converged, and plateaued, so either the model has reached the irreducible error, or more features or complexity is needed

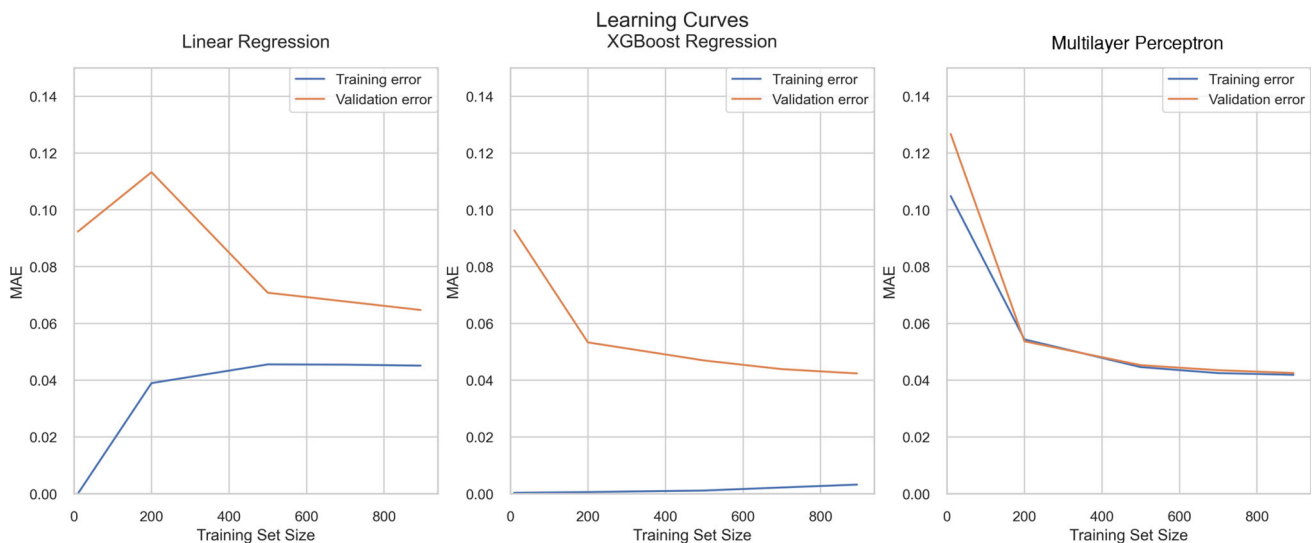


Fig. 11 Learning curves for all 67 features scenario. Linear regression model has high bias and moderate variance, and thus the model is unsuitable for the data set (more features have only increased the variance of the model). For XGBoost, low training error indicates low bias. Large

gap indicates high variance, and that the model is overfitting. Validation curve is not quite plateaued, thus adding more data samples may improve the model. CNN has converged and just plateaued, so more data samples would likely not improve the model

Author contributions Conceptualization: Henry Ekwaro-Osire; Methodology: Henry Ekwaro-Osire, Dennis Bode; Formal analysis and investigation: Henry Ekwaro-Osire, Dennis Bode; Writing—original draft preparation: Henry Ekwaro-Osire; Writing—review and editing: Dennis Bode, Jan-Hendrik Ohlendorf, Klaus-Dieter Thoben, Henry Ekwaro-Osire; Funding acquisition: Dennis Bode; Supervision: Klaus-Dieter Thoben.

Funding Open Access funding enabled and organized by Projekt DEAL. This research has been funded by the German Federal Ministry for Economic Affairs and Climate Action (BMWK) through the Project “ecoKI” [03EN2047A]. The authors wish to acknowledge the funding agency and all project partners for their contribution.

Data availability The datasets generated during and analyzed during the current study are not publicly available due to a confidentiality agreement with the company from whose manufacturing facility the data was collected. Datasets are available from the corresponding author on reasonable request.

Declarations

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

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the

permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Abdoune, F., Ragazzini, L., Nouiri, M., Negri, E., & Cardin, O. (2023). Toward digital twin for sustainable manufacturing: A data-driven approach for energy consumption behavior model generation. *Computers in Industry*, 150, 103949. <https://doi.org/10.1016/j.compind.2023.103949>
- Alvela Nieto, M. T., Nabati, E. G., Bode, D., Redecker, M. A., Decker, A., & Thoben, K.-D. (2019). Enabling energy efficiency in manufacturing environments through deep learning approaches: Lessons learned. In F. Ameri, K. E. Steck, G. von Cieminski & D. Kiritsis (Eds.), *IFIP advances in information and communication technology, 1868-4238: Advances in production management systems: Production management for the factory of the future. APMS conference*, 2019 (Vol. 567, pp. 567–574). Springer. https://doi.org/10.1007/978-3-030-29996-5_65
- Avram, O. I., & Xirouchakis, P. (2011). Evaluating the use phase energy requirements of a machine tool system. *Journal of Cleaner Production*, 19(6–7), 699–711. <https://doi.org/10.1016/j.jclepro.2010.10.010>
- Banks, J. (2010). *Discrete-event system simulation* (5th ed.). Prentice Hall.
- Berthold, M. R., Cebon, N., Dill, F., Gabriel, T. R., Kötter, T., Meinl, T., Ohl, P., Sieb, C., Thiel, K., & Wiswedel, B. (2007). KNIME: The Konstanz Information Miner. In *Studies in classification, data analysis, and knowledge organization (GfKL 2007)*. Springer.
- Bzdok, D., Altman, N., & Krzywinski, M. (2018). Statistics versus machine learning. *Nature Methods*, 15(4), 233–234. <https://doi.org/10.1038/nmeth.4642>
- Chollet, F. (2015). *Keras* (computer software). GitHub. <https://github.com/fchollet/keras>

- Chung, C. A. (2004). In C. A. Chung (Ed.), *Simulation modeling handbook: A practical approach. Industrial and manufacturing engineering series*. CRC Press.
- Cubillo, A., Perinpanayagam, S., & Esperon-Miguez, M. (2016). A review of physics-based models in prognostics: Application to gears and bearings of rotating machinery. *Advances in Mechanical Engineering*, 8(8), 168781401666466. <https://doi.org/10.1177/1687814016664660>
- Dehning, P., Thiede, S., Mennenga, M., & Herrmann, C. (2017). Factors influencing the energy intensity of automotive manufacturing plants. *Journal of Cleaner Production*, 142, 2305–2314. <https://doi.org/10.1016/j.jclepro.2016.11.046>
- Dietmar, A., & Verl, A. (2009). A generic energy consumption model for decision making and energy efficiency optimisation in manufacturing. *International Journal of Sustainable Engineering*, 2(2), 123–133. <https://doi.org/10.1080/19397030902947041>
- Dietterich, T. G. (1998). Approximate statistical tests for comparing supervised classification learning algorithms. *Neural Computation*, 10(7), 1895–1923. <https://doi.org/10.1162/089976698300017197>
- Ekwaro-Osire, H., Bode, D., Thoben, K.-D., & Ohlendorf, J.-H. (2022). Identification of machine learning relevant energy and resource manufacturing efficiency levers. *Sustainability*, 14(23), 15618. <https://doi.org/10.3390/su142315618>
- Engbers, H., & Freitag, M. (2024). Automated model selection for multivariate anomaly detection in manufacturing systems. *Journal of Intelligent Manufacturing*. <https://doi.org/10.1007/s10845-024-02479-z>. Advance online publication.
- Erge, O., & van Oort, E. (2022). Combining physics-based and data-driven modeling in well construction: Hybrid fluid dynamics modeling. *Journal of Natural Gas Science and Engineering*, 97, 104348. <https://doi.org/10.1016/j.jngse.2021.104348>
- European Commission, Joint Research Centre. (2022). *AI watch: AI uptake in manufacturing*. European Commission, Joint Research Centre. <https://doi.org/10.2760/267198>
- Google Cloud. (2021). *Artificial Intelligence acceleration among manufacturers*. Google Cloud. <https://cloud.google.com/blog/products/ai-machine-learning/research-on-ai-trends-in-manufacturing>
- Hastie, T., Tibshirani, R., & Friedman, J. H. (2009). In T. Hastie, R. Tibshirani & J. Friedman (Eds.), *The elements of statistical learning: Data mining, inference, and prediction*. Springer series in statistics (2nd ed.). Springer.
- Hope, T. M. (2019). Linear regression. In A. Mechelli & S. Vieira (Eds.), *Machine learning: Methods and applications to brain disorders* (pp. 67–81). Academic. <https://doi.org/10.1016/B978-0-12-815739-8.00004-3>
- IEA. (2020). *Tracking industry 2020*. IEA. <https://www.iea.org/reports/tracking-industry-2020>
- Imani Asrai, R., Newman, S. T., & Nassehi, A. (2018). A mechanistic model of energy consumption in milling. *International Journal of Production Research*, 56(1–2), 642–659. <https://doi.org/10.1080/00207543.2017.1404160>
- Jamwal, A., Agrawal, R., Sharma, M., Kumar, A., Kumar, V., & Garza-Reyes, J. A. A. (2021). Machine learning applications for sustainable manufacturing: A bibliometric-based review for future research. *Journal of Enterprise Information Management*. <https://doi.org/10.1108/JEIM-09-2020-0361>. ahead-of-print.
- Karmaker, S. K., Hassan, M. M., Smith, M. J., Xu, L., Zhai, C., & Veeramachaneni, K. (2022). AutoML to date and beyond: Challenges and opportunities. *ACM Computing Surveys*, 54(8), 1–36. <https://doi.org/10.1145/3470918>
- Kersten, J., Rohde, H.-R., Nef, E., & Almann, H. (2010). *Mischfutterherstellung: Rohware, Prozesse, Technologie* (3. Aufl.). *Mischfutter-Bibliothek*. Agrimedia.
- Kuhn, M., & Johnson, K. (2016). *Applied predictive modeling* (corrected at 5th printing). Springer. <https://doi.org/10.1007/978-1-4614-6849-3>
- Lago, J., de Ridder, F., & de Schutter, B. (2018). Forecasting spot electricity prices: Deep learning approaches and empirical comparison of traditional algorithms. *Applied Energy*, 221, 386–405. <https://doi.org/10.1016/j.apenergy.2018.02.069>
- Lau, H., Cheng, E., Lee, C., & Ho, G. (2008). A fuzzy logic approach to forecast energy consumption change in a manufacturing system. *Expert Systems with Applications*, 34(3), 1813–1824. <https://doi.org/10.1016/j.eswa.2007.02.015>
- Li, H., Yang, D., Cao, H., Ge, W., Chen, E., Wen, X., & Li, C. (2022). Data-driven hybrid Petri-Net based energy consumption behaviour modelling for digital twin of energy-efficient manufacturing system. *Energy*, 239, 122178. <https://doi.org/10.1016/j.energy.2021.122178>
- Ljung, L., & Glad, T. (2016). *Modeling and identification of dynamic systems*. Studentlitteratur AB.
- Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). Statistical and Machine Learning forecasting methods: Concerns and ways forward. *PLoS ONE*, 13(3), e0194889. <https://doi.org/10.1371/journal.pone.0194889>
- Máša, V., Stehlík, P., Touš, M., & Vondra, M. (2018). Key pillars of successful energy saving projects in small and medium industrial enterprises. *Energy*, 158, 293–304. <https://doi.org/10.1016/j.energy.2018.06.018>
- Mierswa, I., & Ralf, K. (2023). *RapidMiner Studio* (version 10.3) (computer software). <https://www.rapidminer.com>
- Mourtzis, D., Doukas, M., & Bernidaki, D. (2014). Simulation in manufacturing: Review and challenges. *Procedia CIRP*, 25, 213–229. <https://doi.org/10.1016/j.procir.2014.10.032>
- Osara, J. A. (2019). Thermodynamics of manufacturing processes—The workpiece and the machinery. *Inventions*, 4(2), 28. <https://doi.org/10.3390/inventions4020028>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12(85), 2825–2830. <https://jmlr.csail.mit.edu/papers/v12/pedregosa11a.html>
- Renna, P., & Materi, S. (2021). A literature review of energy efficiency and sustainability in manufacturing systems. *Applied Sciences*, 11(16), 7366. <https://doi.org/10.3390/app11167366>
- Samadiani, N., Barnard, A. S., Gunasegaram, D., & Fayyazifar, N. (2024). Best practices for machine learning strategies aimed at process parameter development in powder bed fusion additive manufacturing. *Journal of Intelligent Manufacturing*. <https://doi.org/10.1007/s10845-024-02490-4>. Advance online publication.
- Sealy, M. P., Liu, Z. Y., Zhang, D., Guo, Y. B., & Liu, Z. Q. (2016). Energy consumption and modeling in precision hard milling. *Journal of Cleaner Production*, 135, 1591–1601. <https://doi.org/10.1016/j.jclepro.2015.10.094>
- Surindra, M. D., Alfariy, G. A. F., Caesarendra, W., Petra, M. I., Praseetyo, T., Tjahjowidodo, T., Królczuk, G. M., Glowacz, A., & Gupta, M. K. (2024). Use of machine learning models in condition monitoring of abrasive belt in robotic arm grinding process. *Journal of Intelligent Manufacturing*. <https://doi.org/10.1007/s10845-024-02410-6>. Advance online publication.
- Thiede, S. (2012). Energy efficiency in manufacturing systems. In Zugl.: Braunschweig, Technical University, Dissertation, 2011 (2012nd ed.) *Sustainable production, life cycle engineering and management*. Springer. <http://www.loc.gov/catdir/enhancements/fy1409/2012935578-d.html>

- UN General Assembly. (2015). *Transforming our world: The 2030 Agenda for Sustainable Development* (Resolution 25). UN General Assembly. <https://sustainabledevelopment.un.org/post2015/transformingourworld/publication>
- Waltersmann, L., Kiemel, S., Stuhlsatz, J., Sauer, A., & Mieke, R. (2021). Artificial intelligence applications for increasing resource efficiency in manufacturing companies—A comprehensive review. *Sustainability*, 13(12), 6689. <https://doi.org/10.3390/su13126689>
- Wang, J., Li, Y., Gao, R. X., & Zhang, F. (2022). Hybrid physics-based and data-driven models for smart manufacturing: Modelling, simulation, and explainability. *Journal of Manufacturing Systems*, 63, 381–391. <https://doi.org/10.1016/j.jmsy.2022.04.004>
- Weichert, D., Link, P., Stoll, A., Rüping, S., Ihlenfeldt, S., & Wrobel, S. (2019). A review of machine learning for the optimization of production processes. *The International Journal of Advanced Manufacturing Technology*, 104(5–8), 1889–1902. <https://doi.org/10.1007/s00170-019-03988-5>
- Zhang, C., & Ji, W. (2020). Edge computing enabled production anomalies detection and energy-efficient production decision approach for discrete manufacturing workshops. *IEEE Access*, 8, 158197–158207. <https://doi.org/10.1109/ACCESS.2020.3020136>
- Zolpakar, N. A., Yasak, M. F., & Pathak, S. (2021). A review: Use of evolutionary algorithm for optimisation of machining parameters. *The International Journal of Advanced Manufacturing Technology*, 115(1–2), 31–47. <https://doi.org/10.1007/s00170-021-07155-7>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.