

Machine learning for optimization of energy and plastic consumption in the production of thermoplastic parts in SME

Willenbacher, Martina; Scholten, Jonas; Wohlgemuth, Volker

Published in:
Sustainability

DOI:
[10.3390/su13126800](https://doi.org/10.3390/su13126800)

Publication date:
2021

Document Version
Publisher's PDF, also known as Version of record

[Link to publication](#)

Citation for pulished version (APA):
Willenbacher, M., Scholten, J., & Wohlgemuth, V. (2021). Machine learning for optimization of energy and plastic consumption in the production of thermoplastic parts in SME. *Sustainability*, 13(12), Article 6800. <https://doi.org/10.3390/su13126800>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Article

Machine Learning for Optimization of Energy and Plastic Consumption in the Production of Thermoplastic Parts in SME

Martina Willenbacher ^{1,*}, Jonas Scholten ² and Volker Wohlgemuth ²

¹ Institute of Environmental Communication, Leuphana University Lüneburg, 21335 Lüneburg, Germany

² Industrial Environmental Informatics Unit, Department Engineering-Technology and Life, Hochschule für Technik und Wirtschaft Berlin (HTW), University of Applied Sciences Berlin, 10318 Berlin, Germany; jonas.scholten@posteo.de (J.S.); volker.wohlgemuth@htw-berlin.de (V.W.)

* Correspondence: martina.willenbacher@htw-berlin.de

Abstract: In manufacturing companies, especially in SMEs, the optimization of processes in terms of resource consumption, waste minimization, and pollutant emissions is becoming increasingly important. Another important driver is digitalization and the associated increase in the volume of data. These data, from a multitude of devices and systems, offer enormous potential, which increases the need for intelligent, dynamic analysis models even in smaller companies. This article presents the results of an investigation into whether and to what extent machine learning processes can contribute to optimizing energy consumption and reducing incorrectly produced plastic parts in plastic processing SMEs. For this purpose, the machine data were recorded in a plastics-producing company for the automotive industry and analyzed with regard to the material and energy flows. Machine learning methods were used to train these data in order to uncover optimization potential. Another problem that was addressed in the project was the analysis of manufacturing processes characterized by strong non-linearities and time-invariant behavior with Big Data methods and self-learning controls. Machine learning is suitable for this if sufficient training data are available. Due to the high material throughput in the production of the SMEs' plastic parts, these requirements for the development of suitable learning methods were met. In response to the increasing importance of current information technologies in industrial production processes, the project aimed to use these technologies for sustainable digitalization in order to reduce the industry's environmental impact and increase efficiency.

Keywords: machine learning; artificial intelligence; reduction of emissions and material; energy saving; sustainability



Citation: Willenbacher, M.; Scholten, J.; Wohlgemuth, V. Machine Learning for Optimization of Energy and Plastic Consumption in the Production of Thermoplastic Parts in SME. *Sustainability* **2021**, *13*, 6800. <https://doi.org/10.3390/su13126800>

Academic Editor: Silvia Fiore

Received: 5 March 2021

Accepted: 28 May 2021

Published: 16 June 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 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

In these times of industrialization and digitalization, it is becoming increasingly important to build a sustainable economy. The CO₂ balance of companies plays a major role in this. A further challenge is the environmental pollution caused by plastics, which many industrial companies use and therefore have to eliminate.

The challenges in building a sustainable economy are complex. They require appropriate solutions. One field of information technology is discussed particularly frequently: Artificial Intelligence, and Machine Learning. While large companies have entire departments for the development and implementation of machine learning, small- and medium-sized companies are hesitant to enter this field. SMEs face some challenges in the field of digitalization. SMEs perceive digitization as complex and expensive and in some cases do not see the necessity. In contrast to large companies, failed attempts can quickly lead to financial and personnel difficulties in small companies. The study "Potentials of artificial intelligence in the manufacturing industry in Germany", commissioned by the Federal Ministry of Economics and Energy, came to the conclusion that a lack of internal competence in the manufacturing industry is one of the biggest obstacles to the use of AI

technologies. Only 15% of SMEs in Germany use AI in their respective value-added stages to date. The degree of automation of process control based on AI is also still extremely low [1]. The professional qualifications of employees in the field of artificial intelligence are also only moderately developed. The University of Saarland studied 200 SMEs in 8 EU countries and found that at the present time only 15% of the companies enable their employees to acquire AI-relevant qualifications. The survey of the companies revealed that only 19% use AI applications in their companies [2]. A study by KfW Research even found a value of only 5% in 2020 for the use of AI applications in SMEs [3].

This paper describes the development of a prototype to monitor and optimize waste minimization and energy savings in an SME in the plastics industry using machine learning.

Artificial intelligence is a combination of engineering science and cognitive science. The goal is to translate human performance and perception in terms of cognitive abilities into a technical system. Thus, applications for problem cases are to be developed that could previously only be solved by human actions [4]. The basic requirements for research with AI are thus an understanding of algorithms, data structures, and combinatorics, which makes AI part of the field of theoretical computer science [5] and statistics. As a sub-item of AI, machine learning is one of the most advanced methods of data analysis and data processing [6]. The connection between computer (machine) and human (learning) already implies the basic idea of technology: to learn and predict by analogy with the brain. Machine learning has the goal of learning from data, creating a model, and using this model to make predictions [7]. One of the main applications of machine learning is the creation of adaptive systems and the generation of knowledge from large amounts of data to create a picture of reality [8] and to enable the system to solve problems on its own. Machine learning distinguishes 3 methods: Supervised learning, empowered learning, and unsupervised learning. In supervised learning, a model is created using trained data sets. Encouraging learning serves to develop a system that improves through interaction with its environment. Unsupervised learning is supposed to find patterns without being given precise guidelines. The choice of the learning procedure is based on the scenario and the problem [9].

Unsupervised learning processes data without labels. In this case, however, the entries are already categorized. This ignores information that is already known, which means that an inaccurate result is to be expected. In addition, the goal is already clearly defined: reducing misproduction and lowering energy consumption. A clustering analysis could possibly form groups that have similar properties but which do not contribute to answering the research question. With appropriate interpretation, it might be possible to obtain results, but by incorporating all the available information, especially the labels, the other types of machine learning promise more success. The application of reinforcement learning is one possible way to deliver results. However, one basic requirement was not met for this: interaction with the environment. In this project, already collected data was available. To use this method, the algorithm had to make direct changes to the machine by adjusting the parameters with the results [10]. Even if promising results could be obtained in theory, the financial effort is too high in comparison to the possible potential savings from the research findings. The prerequisites for supervised learning are fully met in this case.

Pre-marked data are available and it is defined how this information is to be analyzed and optimized. Based on this, models can be created and checked for accuracy. Consequently, the method of supervised learning was used.

2. Materials and Methods

2.1. Data Exploration and Data Validation

For the prototype development, the data of 4 different machines of a production plant with a total number of more than 60 machines were evaluated. The data were not only collected and trained offline, but also during ongoing production operations. In addition, indirect factors (e.g., maintenance intervals, mechanical faults) were also taken into account. The data were collected via OPC interfaces developed in the project and protocols of the

ISO standard-defined Modbus RTU. This required a systematic configuration of the IT department. These data are collected in a SQL database. In the first step, the data are viewed and pre-processed. After selecting a suitable machine learning algorithm, a model is generated and analyzed. Approximately 800,000 data sets are considered over a data acquisition period of 4 months. In the first step, the dataset is examined and the required parameters are selected.

Data without information gain are removed from the calculation to increase the calculation time of the model. In this case, these are the attributes tool, material and program. On the whole, the data can be used for machine learning, but the interruptions in the data acquisition lead to an increased effort in filtering and preparing the data. Table 1 contains the recorded machine parameters, which are collected in the database.

Table 1. Description of the database columns.

Description of Database Columns
Date and time of shot
Unique identifier of DB
Name of the machine
Applied tool
Program name used during operation
Applied thermoplastic material
Quality per shot, 1: error-free piece 3: erroneous piece
Shots since program restart
Duration of production of the piece
Compensating mass for contraction on cooling of the piece
The process is switched over from this volume value (repressing)
Pressure for repressing
Maximum pressure of the process
Duration of filling in the mold
Duration of the melting process for the next process
Heating zones for granulate melting
Control of the temperature of the tool
Water temperature to control the tool heating circuit
Cumulative pressure over time
Energy meter reading of the corresponding machine

Figure 1 shows an overview of the available data per machine and parameter. Black markings stand for existing entries, and white markings symbolize data not entered. In the upper bar are the parameters, on the left are the corresponding machine designations, and on the right the number of data per machine. As you can see, all machines have completely white entries for some attributes. These are not missing data; rather, due to the different construction methods these measured values are not collected; for example, the cylinder heating zone does not exist. In general, a complete data set is available. Noticeable are the gaps in cylinder heating zone_K2_(3,4,5)] for machine M69. Due to the coherent, missing data, it can be assumed that in this case the production of the machine was changed over, where these areas are not used.

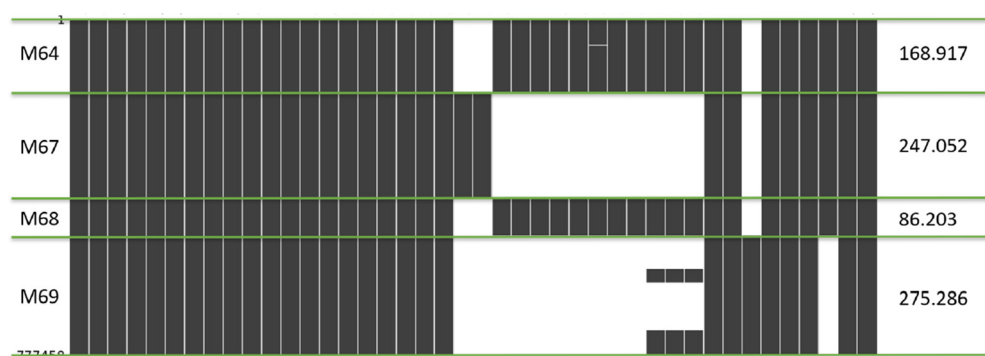


Figure 1. Data amount per machine.

2.2. Selection of the Appropriate Learning Method

At this point, the appropriate learning method for the goals of reducing error production and lowering energy consumption should now be determined. Unsupervised learning processes data without labels. In this case, however, the entries are already categorized. This means that information already known is ignored, which leads to an inaccurate result. Furthermore, the targets are already clearly defined. A clustering analysis could possibly form groups that have similar characteristics but do not contribute to answering the research questions. The application of reinforcement learning is one possible way to deliver results. However, a basic prerequisite for this is not given: interaction with the environment. Data already collected in the project are available. To use this method, the algorithm would have to make direct changes to the machine and use the results to adjust the parameters. The prerequisites for supervised learning are completely fulfilled in this case: pre-labelled data exist and they are defined by the way this information should be analyzed and optimized. Based on this data, models can be created and checked for accuracy. Consequently, the method of supervised learning is used.

2.3. Selection of the Appropriate Algorithm

Before individual algorithms are evaluated, the requirements must first be described. With about 600,000 complete data sets, there is a large amount of information that has to be processed. No assumptions can be made about the model; therefore, non-linear relationships should also be recognized when creating it. The problem requires that a method not only makes correct predictions, but is also interpretable and can explain the reasons why a decision is made. In the best case, the model should be created in as short a time as possible, since the analysis period is limited and costly calculations over several days can quickly impede the completion of tasks. The objective includes the explanation of quality in the form of a discrete classification, as well as a continuous estimation in the form of a regression for energy consumption. Optimally, the algorithm should be able to handle both types in order to reduce the development effort. In addition, a low risk of over-adaptation would be advantageous in order to obtain a general model.

2.3.1. Support Vector Machines

Support Vector Machines are able to perform both linear and non-linear classifications [11]. Furthermore, a regression and a classification are carried out [12]. The classifiers are only described by their support vector and are therefore less susceptible to over-customization. However, the kernel functions have this property, and therefore the problem is shifted from the hyper parameters to the kernel selection [13]. Furthermore, SVMs do not scale precisely enough for large data sets, because they are designed for smaller quantities. The reason for the decision not to use this algorithm for this work is its limited interpretability [14].

2.3.2. Artificial Neural Networks

Artificial neural networks are very promising. A large amount of data is even a basic requirement for this algorithm. In addition, a complex model can be created, and thus non-linear relationships can be mapped [15]. Neural networks can be used to represent both classifications and regressions. For satisfactory functioning, however, a large number of hyper parameters must be determined, which indicate how the basic structure of the neural network is constructed [16]. Compared to other machine learning algorithms, training the neural network requires significantly more computational effort [17].

2.3.3. Random Forest Algorithm

The Random Forest Algorithm offers the possibility to map a large number of data sets and parameters in one model [18]. What distinguishes its interpretability from other methods is its insight into the decisions that take place in the individual trees [19]. For a more comprehensive explanation, Random Forest is also supported by the LIME algorithm. Another specific feature is a low level of pre-processing. Data can be used directly without standardization or normalization. The algorithm is also robust against missing data. However, one of the greatest strengths is its tolerance for overfitting [20]. Due to the random selection of data sets, this circumstance only occurs in rare cases. The models work for both regression and classification [21]. Both the random forest and the neural networks are applicable to the present scenario. Considering its speed of model generation and its easier insight into the decision criteria for classifications, the random forest algorithm is therefore used.

2.4. Implementation

After selecting the machine learning algorithm, the implementation now follows. In the following, the influence of the parameters on the quality is examined first. In the database, a faultlessly produced plastic part is stored with 1, and an incorrect part with 3. For the random forest algorithm, there are therefore two labels to be classified: “quality good” and “quality bad”. For this purpose, the quality column is stored in isolation in a variable after the database has been read in.

2.4.1. Preprocessing

The first step is to consider which parameters should be excluded from the analysis. The shot_ID is a unique identifier of the database and does not contain any information of the machine. Excluding this column prevents the algorithm from remembering individual data records based on the ID. The logged time and date in the Machine_Timestamp column behave similarly. The size Shot_after_Restart derived from the Shot_ID is also neglected. These parameters are not useful for later statements about changes in machine settings. The Program, Tool, and Material columns contain only one value for each machine, which makes them superfluous. This also applies to all data fields that contain zeros.

2.4.2. Setup of Random Forest

In the next step the data set is divided into training and test data. As test data size, 25% is chosen. The algorithm divides the data randomly. Random numbers are usually generated by a deterministic algorithm. These are called pseudo-random numbers because, although they are statistical, they are not actually random due to their predictability. By knowing one generated number in sequence, all other values can be calculated [22]. Every generator needs a starting value, the so-called seed, which is set together with the random variable in this case. Predictability is advantageous in this environment because it guarantees that the same subdivision is made every time the program is started. Thus, later hyper parameter adjustments are comparable. Otherwise, changes in the result could also be due to the fact that other subdivisions have taken place. Now, the hyper parameters for the Random Forest Algorithm are created. For the first run, default values are used to verify the function of the script. Then, the model with the training data is created using the

fit method according to the algorithm from the theory part. Then, the algorithm predicts the labels of the test data. The accuracy score function compares the prediction with the correct results and returns the percentage detection rate.

In the next step, the model must be adapted for optimal detection. For this purpose, it is first of all necessary to find out which is a suitable evaluation method for a good model. The result from the last section calculated in Figure 2 is 99.2%. This represents the percentage of correctly classified parts. At first glance, this figure seems to be a good result. However, the value depends on the distribution of input data. In this case, the proportion of good parts is 98.4%. An algorithm that always predicts a good part would therefore have a classification accuracy of 98.4%. The aim is to reduce the number of faulty parts and their causes. Therefore, a good representation of bad parts in the random forest model is a high priority. For this purpose the confusion matrix is consulted. This describes how many good parts and how many bad parts are correctly identified.

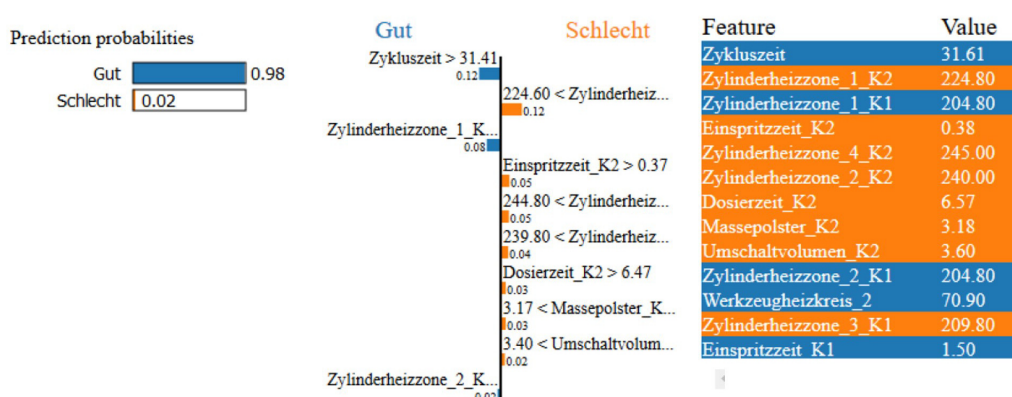


Figure 2. Explanation from LIME.

With this detailed insight into the result of the model, better statements about its quality can be made. While the good parts were detected almost 100% correctly, the bad parts are not yet within an acceptable range, with a detection rate of about 71%. A target of 90% is aimed for. Within this range it can be assumed that no overfitting has taken place and the model has generalized the data sets. The number of data sets (about 42,000) is the chosen 25% test portion of the total measured data, this can be seen in Table 2.

Table 2. Classification in dependency of the labels.

	Error-Free Pieces	Erroneous Pieces
Classification: Error-free pieces	41.510	188
Classification: Erroneous pieces	9	471

In further experiments, some hyper parameters will be adjusted for test purposes. If the number of trees is increased to 150, the detection rate even drops to 72% for bad parts, while good parts continue to be correctly classified at almost 100%. Increasing the tree depth from 6 to 10 with otherwise identical parameters results in a classification rate of 41,510/41,519 (approx. 100%) for good parts and 596/711 (83.4%) for bad parts. This indicates an increase. With the increased depth, however, indirect storage of the data records is more likely. A tree with as many leaves as records could have a separate Create_path. Another test evaluates the under sampling: Up to now, the ratio between error-free and error-prone data was about 99:1. However, a composition of 50:50 would be optimal. Therefore, in this test all bad parts are used and the same number of good parts are randomly read from the database, see this in Table 3. With this balanced data set, the Random Forest Model is again used with the parameters:

- min_samples_leaf: 3;

- `n_estimators`: 100;
- `min_samples_split`: 10;
- `max_features`: `sqrt`;
- `max_depth`: 6;
- `max_leaf_nodes`: `None`.

Table 3. Classification rate with balanced dataset.

	Error-Free Pieces	Erroneous Pieces
Classification: Error-free pieces	656	3
Classification: Erroneous pieces	3	655

Now, the classification rate is very high in both parts. It can be assumed that overfitting is taking place.

2.4.3. Selection of the Optimal Hyper Parameters

After randomly testing some hyperparameter combinations, a method is developed to determine the optimal hyperparameter in a structured way. Since each parameter combination takes about 5 min, batch processing is implemented to test different arrangements. For the program of the batch, the following parameters are permuted in all combinations:

- `balanced_dataset` = [True, False];
- `machine_options` = ['M64', 'M67', 'M68', 'M69'];
- `min_samples_leaf_options` = [3, 5, 7, 10, 30, 50];
- `n_estimators_options` = [10, 30, 50, 100];
- `min_samples_split_options` = [5, 10, 20, 30, 50];
- `max_depth_options` = [3, 4, 5, 10];
- `max_leaf_nodes_options` = [unlimited, 2, 5, 10].

This gives a total of 15,360 results, which are stored in the database. The detection rates of the good and bad parts are entered as absolute and percentage values, as well as the selection of the hyper parameters. Due to the precalculations, the results of the hyperparameter combination can now be output immediately. For this purpose, a graphical tool is created in which the suitable parameters are selected by mouse click and the results from the database can be displayed. With this tool, the hyper parameters can now be determined in such a way that the recognition for good and bad parts is about 90%.

2.4.4. Analysis of the Model

Now, a model exists which can classify data sets with a high probability. However, the desired research result is more complex. Thus, not only an answer for a single data set should be provided, but also an explanation of the model. It should give an exact indication of the area in which the best performance of the machines is achieved, so that waste is reduced, less plastic is produced, and less CO₂ emissions are produced.

For this purpose, the local Brutforce method is used with the LIME [23] explanation algorithm developed by Ribeiro, Singh, and Guestrin. The tool was written in Python and R, and stands for Local Interpretable Model-Agnostic Explanations. LIME is used to generate the explanation of the prediction of any classifier or regressor based on text, tables or images in machine learning, thereby making an approximate understanding of complex models possible. LIME is based on finding independent explanations locally and for each instance, and on fitting simple models locally to the predictions of the complex model. These simplified models make the complex data model interpretable. Instead of trying to create a global model, this method can provide explanations for the environment of a particular data set. For this purpose, the method generates random samples in the environment at equal intervals and weights them according to their distance from the original point. From this, the algorithm delivers a linear explanatory model. This can be displayed graphically for visualization.

On the left are the predictions of the random forest, on the right are the values that the individual parameters have, and the middle shows the explanations of LIME. This result can be used to analyze the decisions for a single data set. However, it is not yet possible to derive and formulate a general statement. The next step is therefore to provide an explanation for each test data set. In the first iteration, occurrences of the display formats specified in the above list are counted and evaluated. Since the generation of the explanations may take some time and memory errors may occur, the results are stored temporarily in a file. For reproducibility and dynamic processing, the hyper parameters, the machine, the LIME parameters, and the generation date are stored in the first line of the file in addition to the explanations. For this purpose, a dictionary is added to the Explainer class as a parameter. For the first test, however, only the most important labels are included. In the next step these are to be grouped and plotted. The search is performed using regular expressions and the required parameter values are extracted. In a dictionary, the arranged value ranges are counted and sorted. Now, more precise statements can be made.

First of all, it is noticeable that considerably more bad than good parts are listed in a diagram. The reason for this is the diagram's sorting according to importance. As described above, only the four most important labels for each data set are included in the evaluation. Thus, it can be interpreted that the injection time is less important for the classification of a faultless production than for a classification of a faulty part. Looking at the relative sizes of the graph within a classification, the result is that the good quality between 0.31 and 0.36 is a clear maximum. Conversely, a significant drop can be seen at this point with poor quality, whereas outside this area there is a high rate.

2.5. Adapting to Energy Analysis

Now, this procedure is transferred to the energy analysis. In addition to the pre-processing of the data, further aspects must be taken into account in the energy analysis. The absolute value of the energy meter is stored in the database after completion of a product. However, the variable to be analyzed is the energy consumption per plastic part produced. For this purpose, the difference of two data sets following each other in time is formed. For a first plausibility check, the energy consumption per time thus obtained is plotted in Figure 3.

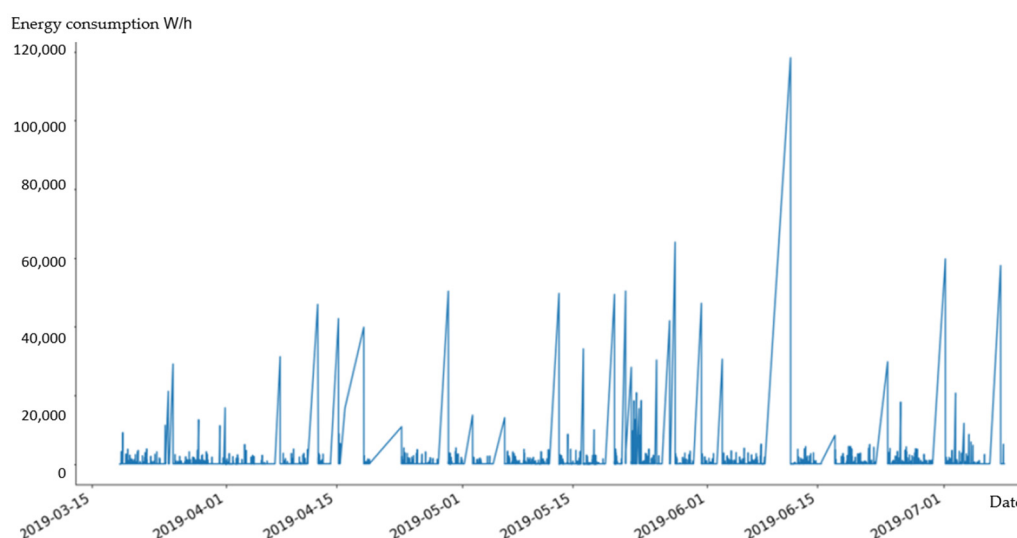


Figure 3. Energy consumption over time for machine 4.

It is obvious that further data filtering must take place. On closer inspection, it is noticeable that the peaks always occur after periods of time in which no data are available. From this it can be concluded that at these points in time, the complete recording of all parts was interrupted. The consequence in the calculation is that for the first part after a longer

break, the energy which was used for all parts in the interruption period is calculated. Two possible causes for the gaps in the data sets are connection problems in the infrastructure or the crash of the data acquisition script. The data records contain the cycle time information, which describes how long the machine needs to produce a part. The average cycle time for Machine 4 is 22.83 s. This time can be selected as an indicator for missing data records. A threshold value of 60 s between two data sets is defined. If the time difference is greater than this, the respective data set is ignored for the energy analysis. This already gives a more conclusive picture. There are also energy differences of 0 Wh, which are also excluded from the analysis for plausibility reasons. In addition, one is only interested in those production runs in which a proper plastic part was produced. Data sets with quality = 3 are therefore not taken into account.

Technical Implementation

Since the data acquisition is interrupted at some points, a time difference between two data sets must be created for the check. When working with databases, data integrity should be maintained at all times. This means that the time difference is not added to the table production log of the machines, but the separate table energy difference is created, with a reference to the ID in the production log. This has the advantage that in the case of errors, the original data in the production log remain unchanged, and changes can be undone by resetting the energy difference. In the first step, after connecting to the database, an empty array is initiated, which has the length of the data to be written. Since for the difference two consecutive values are always required, the first entry in the array must be initiated with a placeholder for an invalid value. In this case the common -1 is used. A missing hour in the data is conspicuous—followed by an apparently double data series, whose energy counter rises and falls. The reason for this is the time changeover on 31 March. The problem is solved by sorting by shot_ID. In general, it is noticeable that the energy series is largely constant. The noise in the signal can be attributed to the sensor's accuracy of 10 Wh. This has the consequence that the values alternate between two to three steps of 10. Some peaks are visible, but these are probably due to missing measurement data that the filtering from the previous section could not capture. At machine 1, on 17 May 2019, at 13:12, the energy consumption suddenly increases from about 170 Wh to 250 Wh. For analysis with Machine Learning, this device is most promising due to this significant difference, and can be used as a validator of the later algorithm. Figure 4 shows the energy consumption per plastic part over time.

From the basic structure, the algorithm of energy analysis is close to that of quality analysis. A model is created, which is interpreted with LIME, and finally, plotted. The only difference is in the labels: While the discrete values 1 and 3 were present for the quality analysis, a continuous range is possible for the energy analysis. This changes the evaluation of the hyper parameters used. When analyzing quality, a binary statement existed: correct or incorrect classification. This method could also be used for continuous values. However, to evaluate these as incorrect with minimal deviations from the original value leads to a seemingly bad algorithm in many data sets. A better evaluation is therefore to consider the relative deviation. This also improves the interpretation of individual predictions. It can be said not only whether a forecast is incorrect, but also how incorrect the forecast is. This also leads to the fact that—in contrast to the evaluation of quality—there is no unbalanced data set. Therefore, all entries for the machine can be used. Errors in production can lead to increased energy consumption. Therefore, only those data sets describing a faultless plastic part were considered in the energy analysis. In the next step, the hyper parameter combinations are tested again and finally the best Configuration is selected. Table 4 shows the optimal values for each machine.

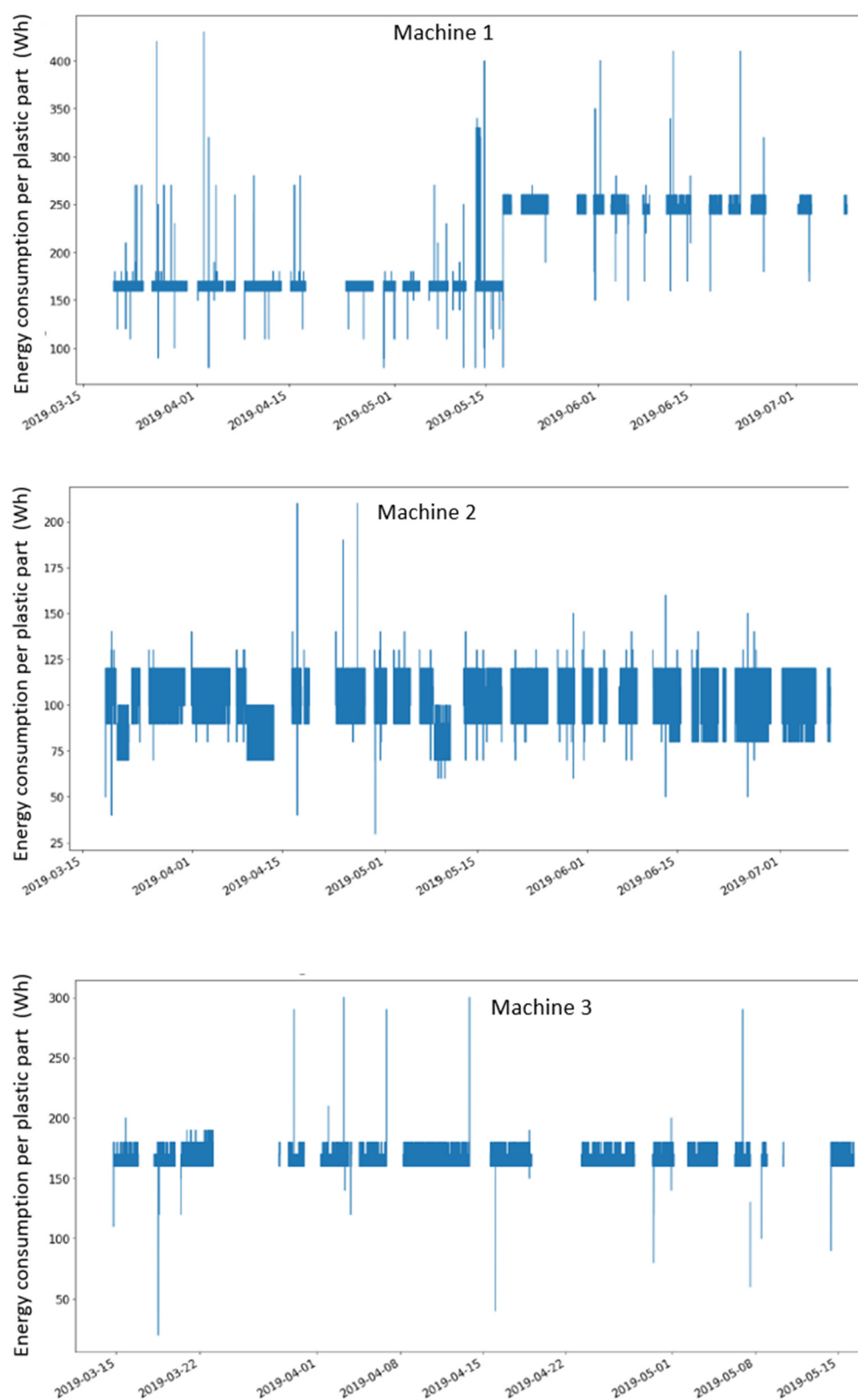


Figure 4. Cont.

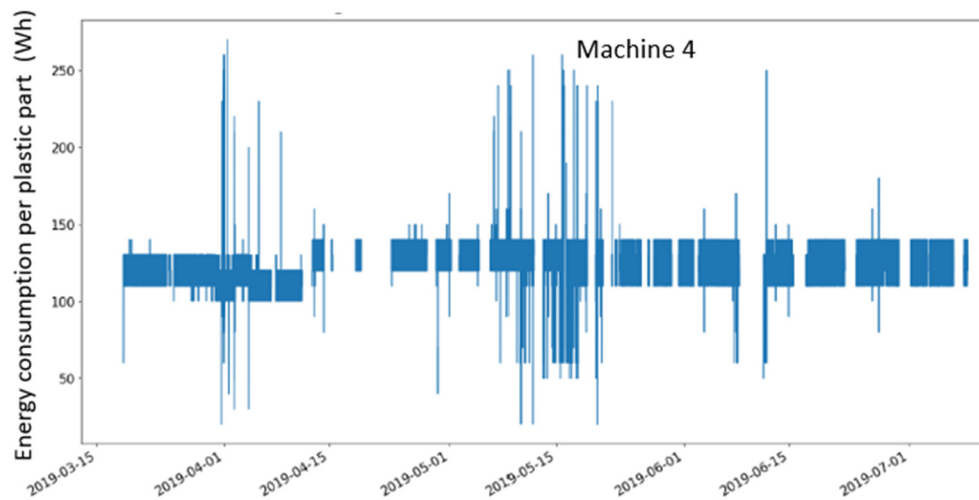


Figure 4. Energy consumption over time.

Table 4. Selected hyper parameters per machine.

Parameter	M 1	M 2	M 3	M 4
min_samples_leaf	100	80	200	100
n_estimators	20	70	50	40
min_samples_split	1000	800	1100	900
max_depth	3	3	6	4
max_leaf_nodes	20	30	30	20

Figure 5 shows the project procedure as an overview.

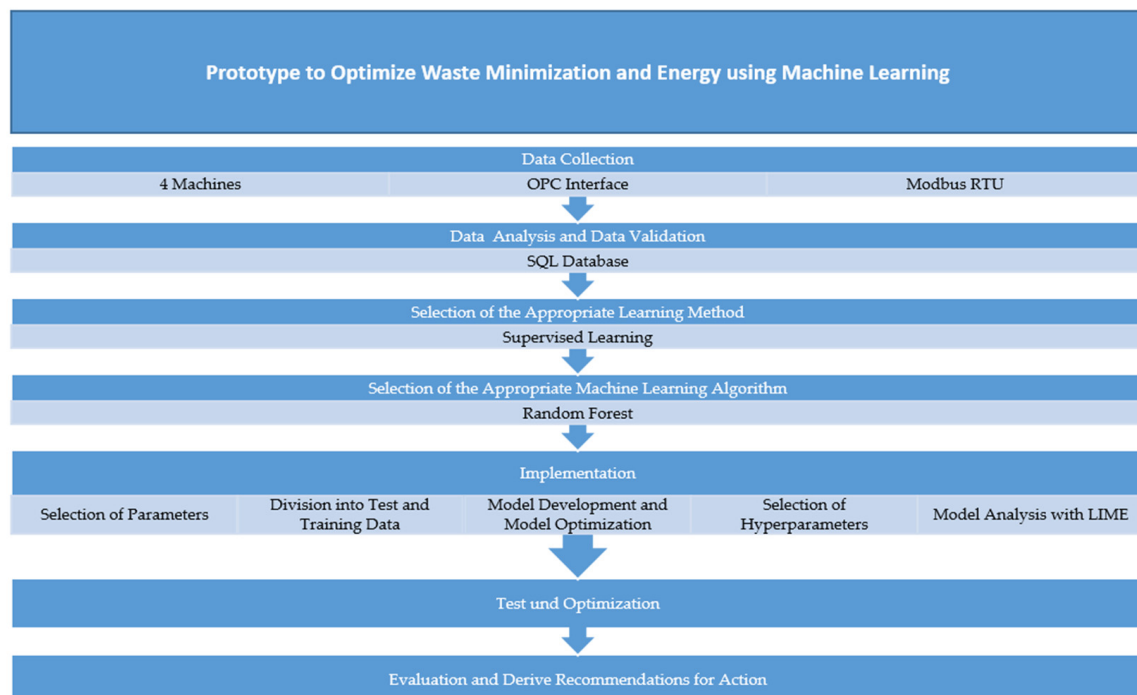


Figure 5. Project procedure.

3. Results

3.1. Interpretation of the Diagrams

When viewing the diagrams, different types of curves can be grouped together. During processing with LIME, premature filtering of the most important features was deliberately prevented, so that it is only during interpretation that a decision has to be made as to how relevant a feature is. As a result, a suitable selection must be made afterwards. It is therefore important to first consider the scale of the y -axis (probability or energy). A relevance of at least 1% for quality and 2 Wh for energy was chosen as a prerequisite for their respective inclusion in the evaluation. Figure 6a shows, for example, a clear increase in quality between 91.8 °C and 92.4 °C. However, the difference in probability (0.005) is too small to make a sound statement. Figure 6b is similar. However, there is the additional fact that no preferences can be determined from the curve shape. Here, a classical noise signal is present. Even filtering through the minimum number of data sets does not produce any improvement. With a few exceptions, only one data set per point is shown in this diagram. Occasionally, there are diagrams like Figure 6c in which very few x -values exist. Here, a small variance can already be seen in the raw data. Possible reasons for this could be a low resolution during data collection or a defective sensor.

Figure 6d shows the optimum case. Here, there is a relative difference of 50 Wh, the curves show a clear trend for optimum spray pressure above 740 mbar, and the noise of the values is very low. Another consideration should be taken into account when assessing the importance: In some diagrams the values are permanently below zero, thus always contributing to a bad part classification. This could lead to the conclusion that the feature is irrelevant, since no value contributes to good part detection. However, the relative value of two references is always decisive. Based on the diagrams, the results are presented for each machine. The results are read from the remaining diagrams with significant significance. For each characteristic, a recommendation is made as to which are the optimum machine settings. In addition, the size of the influence for classification into a good part is indicated. For the difference in relevance, the range between the smallest and largest value is shown (Figure 7a). The diagrams can be divided into three categories: Some charts cannot be classified into an optimal range. However, it is clearly visible at which points the result deteriorates. Therefore, the recommendation is given as a range above this range (Figure 7b). On the other hand, there is a peak in the diagram that can be clearly read (Figure 7c). There are also cases where noticeable peaks can be seen in two places (Figure 7d).

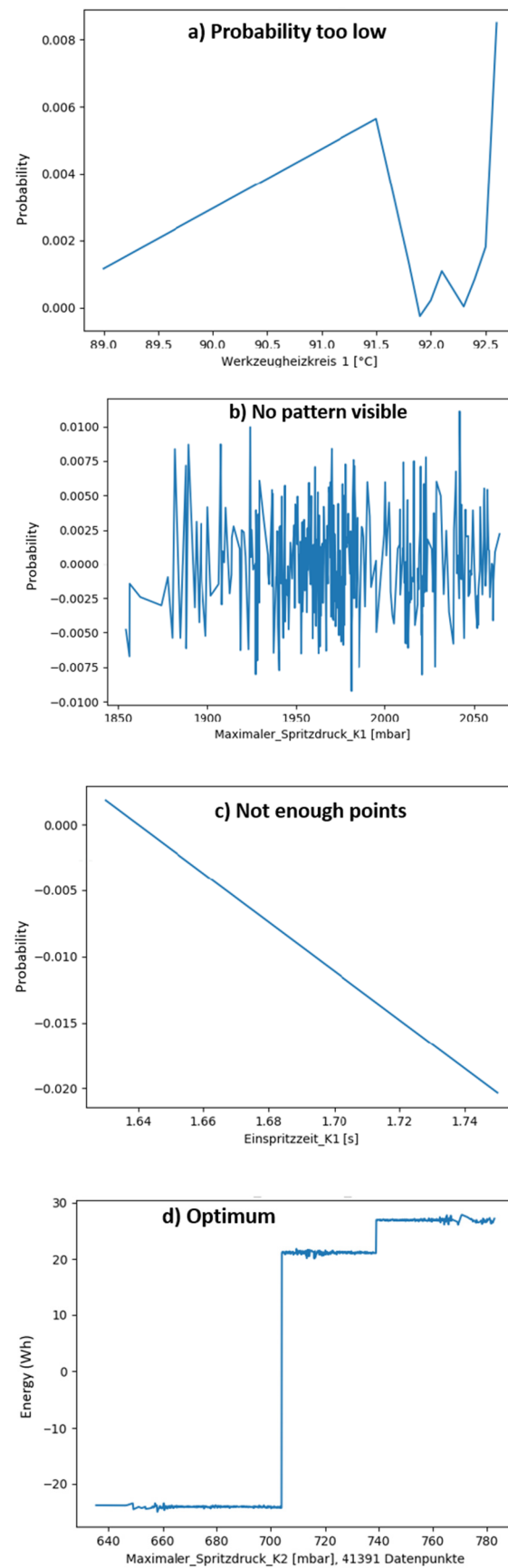
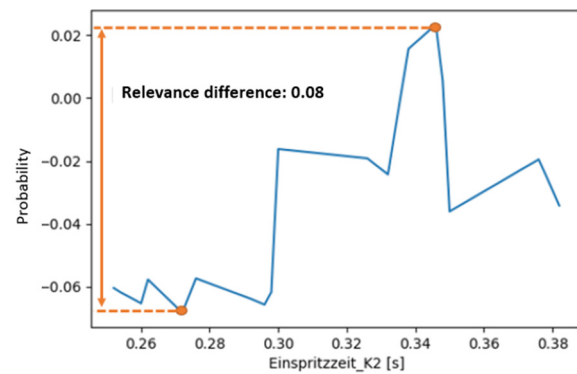
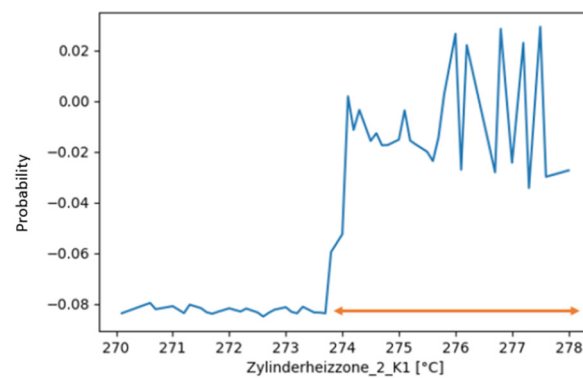


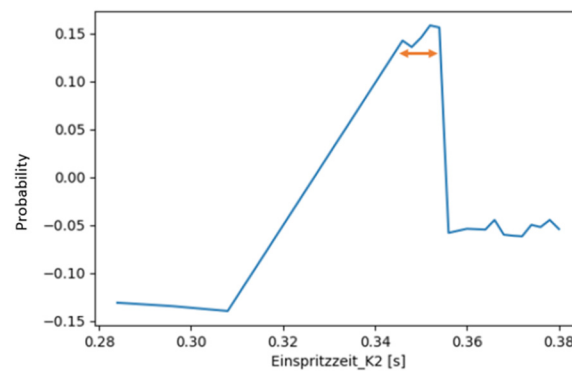
Figure 6. Evaluation of different diagrams (1).



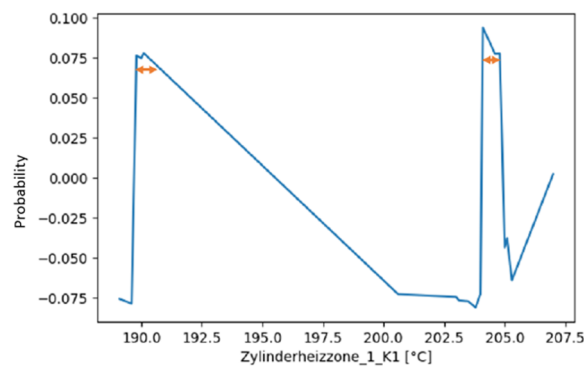
a) Specification of the relevance between the smallest and largest value



b) Range above a value is optimal.



c) Value in a small range is optimal.



d) Two ranges are optimal.

Figure 7. Evaluation of different diagrams (2).

3.2. Recommended Actions for the Machine Settings

3.2.1. Machine 1

For machine 1, a large amount of data was collected, with about 168,000 entries. The high ratio of 1.6% bad parts is important for the evaluation of the quality of the advantageous parts, see this in Table 5. Detailed results can therefore be expected. The same applies to the consideration of energy. In contrast to other machines, there was a jump here, which should lead to a good classification.

Table 5. Recommendation for quality.

Feature	Recommendation	Relevance Difference (Probability)
Duration of the melting process for the next process	5.75 s	0.02
Duration of filling in the mold	$0.34 \text{ s} < x < 0.35 \text{ s}$	0.29
Maximum pressure of the process	$685 \text{ mbar} < x$	0.04
Pressure for repressing	$685 \text{ mbar} < x$	0.06
The process is switched over from this volume value	$3.8 \text{ cm}^3 < x < 4.1 \text{ cm}^3$	0.04
Duration of production of the piece	$x > 31.42 \text{ s}$	0.15
Heating zones for granulate melting 1	$190 \text{ }^\circ\text{C}$ or $204 \text{ }^\circ\text{C}$	0.14
Heating zones for granulate melting 2	$219.5 \text{ }^\circ\text{C} < x < 220.3 \text{ }^\circ\text{C}$ or $x = 224.5 \text{ }^\circ\text{C}$	0.39
Heating zones for granulate melting 3	$204.5 \text{ }^\circ\text{C}$	0.05

Especially clear maxima can be calculated in the quality analysis for the characteristics' duration of filling in the mold and cylinder heating zone_1. With 29% and 39%, there is a considerable predictable reduction in the number of parts produced in error. Clear patterns have been established for energy. The highest optimization is used for the characteristic _maximum_ energy. Furthermore, there are savings of 25 Wh possible for injection time and 15.5 Wh for switch-over injection pressure. Under the assumption that the values influence each other, an overall estimate with 60 Wh/part is a very conservative forecast, which in practice can be significantly higher. See the recommendation for energy in Table 6.

Table 6. Recommendation for energy.

Feature	Recommendation	Energy Difference (Wh)
Duration of the melting process for the next process	5.75 s	5
Duration of filling in the mold	$x > 0.37 \text{ s}$	25
Compensating mass for contraction	$x > 3 \text{ cm}^3$	10
Maximum pressure of the process	$x > 742 \text{ mbar}$	45
Pressure for repressing	$x > 742 \text{ mbar}$	15.5
The process is switched over from this volume value	$x > 3.45 \text{ cm}^3$	13
Duration of the melting process for the next process	5.75 s	5

3.2.2. Machine 2

Machine 2 has about 267,000 entries; the second largest data set of the considered machines. During the period under consideration, a low rate of faulty production of 0.4% was achieved. With 1070 bad parts, the analysis nevertheless shows good results are possible, see this in Table 7. The energy consumption over time shows a constant distribution, with individual periods of reduced energy. It was therefore to be expected that patterns for an optimization can be recognized. A clear result, as with machine 1, is not to be expected.

Table 7. Recommendation for quality.

Feature	Recommendation	Relevance Difference (Probability)
Duration of the melting process for the next process	5.15 s	0.03
Compensating mass for contraction on cooling	8.4 cm ³	0.04
Pressure for repressing	x > 7112 mbar	0.02
Control of the temperature of the tool	79.5 °C	0.09
Heating zones for granulate melting 1	267.1 °C < x < 69.3 °C	0.08
Heating zones for granulate melting 2	x > 274.2 °C	0.08
Heating zones for granulate melting 3	x > 279.2 °C	0.02
Heating zones for granulate melting 4	285 °C	0.12
Duration of the melting process for the next process	5.15 s	0.03

The evaluation of quality shows several possibilities for improvement, which are between 8% and 12%. The cylinder heating zones are particularly present. The mold heating circuit shows a clear maximum at 79.5 °C, at which point the probability is 9% higher than the minimum. For energy optimization, there is only a significant difference in the switching volume (10 Wh). Nevertheless, the results of all characteristics allow a saving of 20 Wh/part forecast. See this recommendation for energy in Table 8.

Table 8. Recommendation for energy.

Feature	Recommendation	Energy Difference [Wh]
Pressure over time	x > 870 mbar/s	3
Compensating mass for contraction on cooling	x > 11 cm ³	5
The process is switched over from this volume value	x < 22 cm ³	10
Heating zones for granulate melting	x > 275 °C	2

3.2.3. Machine 3

Machine 3 has the smallest amount of information, with about 86,000 data records. In addition, less than 1000 data were classified as bad parts. In terms of energy consumption over time, the scatter of values is minimal on this machine, and therefore, expectations regarding the results are low. With regard to energy, only correlations with the cycle time were detected, and this is 2.5 Wh at a low level. No savings are therefore made in terms of energy consumption, as predicted by changed attitudes. The situation is different for quality. There, a clear peak can be seen in the mold heating circuit, which shows a difference of 20% for the quality prognosis, check Table 9.

Table 9. Recommendation for quality.

Feature	Recommendation	Relevance Difference (Probability)
Duration of filling in the mold	0.342 s	0.08
Control of the temperature of the tool	72.1 °C	0.2
Heating zones for granulate melting 1	194.5 °C	0.08

With 8% in each case, the change of the injection time and, for the cylinder heating zone_1, a reduction in the number of missing parts can be predicted. All in all, there is little potential for optimization in practical tests on machine 3, see Table 10 for the recommendation for energy. Few data have their origin in the lower production, compared to the other machines. This means that this device, anyway, is less relevant for the total savings.

Table 10. Recommendation for energy.

Feature	Recommendation	Energy Difference [Wh]
Duration of production of the piece	$x > 33.5$ s	2.5

3.2.4. Machine 4

For machine 4, the most entries in the database are available, with about 275,000. The high rate of missing parts of 1.6% provides a good data basis for the patterns behind the good and bad quality classifications. The energy consumption over time shows a constant straight line, which, however, reduces energy consumption section by section. In the results for quality, the cylinder heating zones, with values between 8% and 18%, have a significant impact, see Table 11.

Table 11. Recommendation for quality.

Feature	Recommendation	Relevance Difference (Probability)
Duration of the melting process for the next process	5.26 s	0.05
Compensating mass for contraction on cooling	$6.1 < x < 6.6$ cm ³	0.15
Water temperature to control the tool heating circuit	25.8 °C	0.03
Heating zones for granulate melting 1	$x > 243.3$ °C	0.18
Heating zones for granulate melting 2	244.1 °C	0.16
Heating zones for granulate melting 3	245.6 °C	0.08
Duration of the melting process for the next process	5.26 s	0.05

The mass cushion shows a large noise component in the diagram. In the recommended range between 6.1 cm³ and 6.6 cm³, however, a clear straight line and an improved probability for good parts can be found. With regard to energy consumption, there are two characteristics that influence energy: the tool heating circuit and the cycle time. In both cases there is a point where the energy level drops rapidly. If the machine settings are adjusted accordingly, energy savings of 10 Wh/part can be expected, see Table 12.

Table 12. Recommendation for energy.

Feature	Recommendation	Energy Difference (Wh)
Control of the temperature of the tool	210 °C	8
Duration of production of the piece	$x > 22.8$ s	9

In terms of quality, it is often the characteristics that control the temperature that are decisive. The cylinder heating zone in particular has an influence on the quality of every machine without exception. When looking at the origin of the data, a correlation is conclusive: If the granulate is not at the right temperature for the plastic, it may dry too quickly or too slowly. If the produced part falls out of the injection mold into the output, deformations can occur afterwards if the heat is too high, which can damage the product. If the temperature is too low, the liquefied granulate may be too tough and may not fill the mold completely, which also leads to rejection. In terms of energy, no similarities between the machines can be seen. The analysis must therefore be carried out again for each machine. Nevertheless, significant differences in some of its characteristics are evident.

For energy savings, the measurement period is 112 days. This value is interpolated to 12 months in order to make statements about energy savings per year, see Table 13.

Table 13. Prediction of energy savings per machine.

	Potential Savings Energy per Piece (Wh)	Produced Pieces in Measuring Period	Interpolated for a Year	Energy Saving per Year (kWh)
M 1	60	168,917	550,488.44	33,029.31
M 2	20	247,052	805,124.82	16,102.5
M 3	0	86,203	280,929.42	0
M 4	10	275,286	897,137.41	8971.37
				58,103.18

In total, for the four machines, savings of about 58,100 kWh are predicted. Based on the Federal Environment Agency's recommendations, a value of 474 g CO₂ per kWh is assumed for CO₂ reduction. This results in a potential reduction of 27,540 kg CO₂ per year for the four machines.

4. Discussion

By using the Random Forest Algorithm, the information could be used efficiently and precisely to create a model. With the help of the LIME method, diagrams were created from which patterns could be interpreted. The resulting findings provide answers to the research questions: Machine learning methods can show that the individual parameters of the machines influence quality and energy consumption. The potential electricity reduction is 27,540 kg CO₂ per year for the four machines under consideration. The next step is to test the results obtained on the machines and verify their effects in practical use. This procedure is of elementary importance. The collected data only represent a section of the real environment, both in the spatial and temporal components. A correlation between two parameters gives only an indication, but cannot provide a basis for proof between cause and effect. An interesting investigation for future work is a global approach, where the individual machines are not considered, but their entirety. For example, external influences such as temperature fluctuations in the production hall could be detected, and new optimization possibilities could be found. On the software side, there are also approaches that can still be pursued. For example, the application of a neural network could lead to further results. For the interpretation of the models, it has to be evaluated whether other methods can be used to achieve more meaningfulness. Presently, the parameters are independent of each other. A multi-dimensional view, which analyses combinations of settings, has the potential to reveal further energy savings. However, current work shows that the use of machine learning in SMEs can provide hidden avenues for energy savings and a reduction in faulty production.

5. Conclusions

In the plastics processing company studied, several scenarios are conceivable for the further expansion of digitalization. If the material flow management system and the necessary data sources cover the entire operation, various possibilities will be created. It is conceivable that through the entire intelligent monitoring of the production cycle, further potential optimization opportunities can be found in the areas of resource efficiency, machine utilization, and the use of operating resources. Opportunities for savings and optimization potential can also be assumed outside of production. One area that could be worthwhile is material procurement: Procurement can be optimized by predicting the required raw materials in combination with a market analysis. Maintenance intervals and necessary repairs can also be better planned and carried out at an early stage before damage occurs. Likewise, logistics (fuel or electricity costs) and production (utilization of machines or planning of personnel, etc.) can lead to higher efficiency. With regard to the operational environmental area, it will be important to substitute and allocate resources, or, alternatively, to avoid them and to find and use new savings avenues. Machine learning can play a key role here, as it is very applicable to the data requirements of the environmental sector—large, heterogeneous data volumes, different formats and

sources. Conceivable scenarios range from the analysis and prediction of environmental data during operations and the resulting control and monitoring of production, to self-learning production processes. Machine learning is a forward-looking technology and is currently also gaining ground in the environmental sector. However, digitalization presents SMEs with a number of challenges. More than half of all businesses see themselves as laggards in this area [24]. SMEs perceive digitalization as complex and expensive and in some cases do not see the need for it. Likewise, in most companies, the human resources required for implementation are not available. In contrast to large companies, failed attempts in small firms can also quickly lead to financial difficulties [25]. At this point in time, it can be concluded that digitalization and especially the use of artificial intelligence in SMEs is still too uncertain and that they are largely clinging to their old structures.

Author Contributions: Conceptualization, M.W. and V.W.; methodology, M.W.; software, M.W. and J.S.; validation, M.W. and J.S.; formal analysis, M.W.; investigation, M.W. and J.S.; resources, M.W.; data curation, M.W.; writing—original draft preparation, M.W. and J.S.; writing—review and editing, M.W., J.S. and V.W.; visualization, M.W. and J.S.; supervision, V.W. and M.W.; project administration, M.W.; funding acquisition, V.W. All authors have read and agreed to the published version of the manuscript.

Funding: In cooperation with Novapax Kunststofftechnik Steiner GmbH & Co. KG, the University of Applied Sciences Berlin is working on the implementation of a prototype in the Nova [26] research project to monitor and optimize waste minimization and energy savings in an SME in the plastics industry using machine learning. This research was funded by Deutsche Bundesstiftung Umwelt, grant number 34589/10.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data in this study are available on reasonable request from the corresponding author.

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

References

1. Gabriel, P. *Potenziale der Künstlichen Intelligenz im Produzierenden Gewerbe in Deutschland*; Begleitforschung PAiCE, iit-Institut für Innovation und Technik in der VDI/VDE Innovation + Technik GmbH: Berlin, Germany, 2018; p. 5.
2. Kaul, A.; Schieler, M. *Künstliche Intelligenz im Europäischen Mittelstand: Status quo, Perspektiven und was Jetzt zu tun ist*; University of Saarland: Saarbrücken, Germany, 2019.
3. Zimmermann, V. *Künstliche Intelligenz: HOHE Wachstumschancen; Aber Geringe Verbreitung im Mittelstand*, KfW Research Fokus Volkswirtschaft. Nr. 318; KfW: Frankfurt/Main, Germany, 2021.
4. Russell, S.; Norvig, P. *Artificial Intelligence: A Modern Approach*, 3rd ed.; Pearson: Upper Saddle River, NJ, USA, 2009; Chapter 1.
5. Poole, D.; Mackworth, A.K.; Goebel, R. *Computational Intelligence: A Logical Approach*; Oxford University Press: New York, NY, USA, 1998; Chapter 1.
6. Bönisch, M. Konzept zur Einführung von Maschinellem Lernen im Kontext der Betrieblichen Nachhaltigkeit in Mittelständischen Unternehmen. Master's Thesis, University of Applied Sciences Berlin, Berlin, Germany, 2019.
7. Alpaydin, E. *Maschinelles Lernen*; Oldenbourg: Oldenburg, Germany, 2004.
8. Görz, G.; Rollinger, C.-R.; Schneeberger, J. (Eds.) *Handbuch der Künstlichen Intelligenz*, 5th ed.; Oldenbourg Verlag: Oldenburg, Germany, 2014; Chapter 11.
9. EMC Education Services. *EMC Services, Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data*; Wiley: Indianapolis, IN, USA, 2015; Chapter 7.
10. Maimon, O.; Rokach, L. *Data Mining and Knowledge Discovery Handbook*; Series in Solid-State Sciences; Springer: Boston, MA, USA, 2010; Chapter 12.
11. EMC Services, *Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data*; Wiley: Indianapolis, IN, USA, 2015; Chapter 5.
12. Dean, J. *Big Data, Data Mining, and Machine Learning: Value Creation for Business Leaders and Practitioners*; Wiley and SAS Business Series; Wiley: Indianapolis, USA, 2014; Chapter 5.
13. Willenbacher, M.; Wohlgemuth, V. Machine Learning for Sustainable Logistics. In Proceedings of the EnviroInfo 2019, Kassel, Germany, 23–26 September 2019.

14. Willenbacher, M.; Wohlgemuth, V. Einsatzmöglichkeiten von Methoden der Künstlichen Intelligenz zur Optimierung von Stoff- und Energieströmen und prototypische Umsetzung auf der Basis von Stoffstromnetzen. In *Nachhaltige Betriebliche Umweltinformationssysteme*; Springer: Wiesbaden, Germany, 2018.
15. Kotu, V.; Deshpande, B. *Data Science: Concepts and Practice*; Elsevier Science: Amsterdam, The Netherlands, 2018.
16. Nisbet, R.; Miner, G.; Yale, K. *Handbook of Statistical Analysis and Data Mining Applications*; Elsevier Science: Amsterdam, The Netherlands, 2017.
17. Raschka, S. *Python Machine Learning*; Packt Publishing: Birmingham, UK, 2015; p. 90.
18. Lokuciejewski, P.; Marwedel, P. *Worst-Case Execution Time Aware Compilation Techniques for Real-Time Systems, Embedded Systems*; Springer: Dordrecht, The Netherlands, 2010; p. 171.
19. EMC Services, *Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data*; Wiley: Indianapolis, IN, USA, 2015; Chapter 6.
20. Love, R. *Linux-Kernel-Handbuch: Leitfaden zu Design und Implementierung von Kernel 2.6*; Open Source Library; Addison-Wesley: San Francisco, CA, USA, 2005; p. 433.
21. Ribeiro, M.T.; Singh, S.; Guestrri, C. "Why Should I Trust You?": Explaining the Predictions of Any Classifier. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 1135–1144.
22. Leiser, D.C. *Digitalisierung in KMU kompakt: Compliance und IT-Security*; IT Kompakt; Springer: Berlin/Heidelberg, Germany, 2019.
23. Lindner, D. *KMU im digitalen Wandel: Ergebnisse empirischer Studien zu Arbeit, Führung und Organisation*; Essentials; Springer: Wiesbaden, Germany, 2018.
24. Research Project "NOVA". Available online: <http://nova.f2.htw-berlin.de/> (accessed on 22 April 2021).
25. Muhammad, I.; Yan, Z. Supervised Machine Learning Approaches: A Survey. *ICTACT J. Soft Comput.* **2015**, *5*, 947.
26. Rao, D.; Visin, F.; Rusu, A.A.; Teh, Y.W.; Pascanu, R.; Hadsell, R. Continual unsupervised representation learning. In Proceedings of the 33rd Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, BC, Canada, 8–14 December 2019.