Cleaner Environmental Systems 13 (2024) 100187

Contents lists available at ScienceDirect

Cleaner Environmental Systems

journal homepage: www.journals.elsevier.com/cleaner-environmental-systems

A simplified machine learning product carbon footprint evaluation tool

Silvio Lang [*], Bastian Engelmann, Andreas Schiffler, Jan Schmitt [**]

Technichal University of Applied Sciences Würzburg-Schweinfurt, Institute of Digital Engineering, Ignaz-Schön-Straße 11, Schweinfurt, 97421, Germany

ARTICLE INFO

Keywords:
Machine learning
Product carbon footprint sustainability

ABSTRACT

On the way to climate neutrality manufacturing companies need to assess the Carbon dioxide (CO_2) emissions of their products as a basis for emission reduction measures. The evaluate this so-called Product Carbon Footprint (PCF) life cycle analysis as a comprehensive method is applicable, but means great effort and requires inter-disciplinary knowledge. Nevertheless, assumptions must still be made to assess the entire supply chain. To lower these burdens and provide a digital tool to estimate the PCF with less input parameter and data, we make use of machine learning techniques and develop an editorial framework called MINDFUL. This contribution shows its realization by providing the software architecture, underlying CO_2 factors, calculations and Machine Learning approach as well as the principles of its user experience. Our tool is validated within an industrial case study.

1. Introduction

A key indicator that companies determine as a part of their way to net zero is the so-called product carbon footprint, or PCF. Determining the PCF is a major challenge, especially for Small and Medium Enterprises (SMEs), due to the heterogeneous data sources, distributed data and knowledge through the supply chain (Bianchini and Michalkova, 2019). Recently the PCF is normally evaluated by a comprehensive Life Cycle Analysis (LCA). Commercial LCA-tools are costly and need expert knowledge to model the material and energy streams, which is generally scarce in performing complete LCA studies (Arzoumanidis et al., 2017). Further, an entire data set to perform a LCA is often not available and assumptions have to be made e.g. to estimate material processing ways in the pre-chain or energy grid data, which can influence the LCA results critically (Anshassi et al., 2021). Due to the described nexus between the complexity and effort to put into an LCA and the necessity for a company to generate knowledge and data about CO_2 emissions, Machine Learning (ML) models are a promising approach to simplify the determination of the PCF (Xue et al., 2024; Algren et al., 2021). The core concept driving this tools functionality is the utilization of ML to identify similarities in products and processes between an established dataset of PCFs and a new product, akin to the initial training data space, requiring the calculation of its PCF. Using a limited number of comprehensive datasets containing analytically calculated CO_2-emissions and main features influencing the PCF for training, a self-optimizing machine learning model predicts the PCF. The features of our proposed PCF calculation model relate to four main factors:

- Material: weight, material type (including pre-chain emissions), pieces in product
- Processes: process technology, process time, energy type, energy consumption, energy mix at location
- Logistics in the factory from the supplier: transportation type, route from … to
- Logistics out from the factory to the customer: transportation type, route from … to

We create a **M**ach**I**ne Lear**N**ing Pro**D**uct Carbon **F**ootprint Eval**U**ation Too**L** (MINDFUL), where users can upload their data as CSV, link the columns (features) to the manda-tory PCF-input parameter in the editor (“Configurator*module*’). The software architecture allows conducting a standard analytical process (“*Analysismodule*’) as well as the derivation of a ML-model (“*Modelling-module*’). Within the *Applications-module* the results are visualized e.g. correlations or model accuracy. Due to this software characteristic as an editorial framework, the product and process features can be used and enhanced with a learning approach and expert knowledge for PCF determination. Here, ML techniques are helpful as the forecast model can be continuously adapted and improved via similarities and model adjustments through expert dialog. Companies provide an existing data set for model training. The trained model can then forecast a PCF based on product, process and logistics similarities. SMEs in particular often have a product portfolio

[*] Corresponding author.
[**] Principal corresponding author.
E-mail addresses: silvio.lang@thws.de (S. Lang), jan.schmitt@thws.de (J. Schmitt).

https://doi.org/10.1016/j.cesys.2024.100187
Received 8 January 2024; Received in revised form 18 March 2024; Accepted 12 April 2024
Available online 16 April 2024

characterized by the use of similar materials and processes. Therefore, the presented approach is promising to provide advances for sustainability management circularity as.

- different product or component design variants can be evaluated due to their PCF,
- the environmental influence of relocations for products of components production can be assessed,
- the impact of logistic routes to the PCF needs to be minimized,
- material process routes e.g. green vs. normal steel provide high optimization potential in terms of CO_2 reduction.

Our research contribution lies in the realization and proof-of-concept of the MINDFUL tool by a case study. Due to this, the paper is structured as follows:

Chapter 2 briefly discussed recent literature in the field of simplified LCA as well as PCF evaluation with machine learning. In Chapter 3 the editorial framework is presented by its requirements and methods and the software architecture and implementation. The validation of the tool is done in Section 4 by an industrial case study with a highly integrated mechatronic product. Due to the requirement of applicability of MINDFUL especially for SMEs, simplifications have to be done. Hence, limitations are present, which are discussed in Chapter 5. The contribution closes with a conclusion and further research issues (Section 6).

2. Brief literature review

Methods for calculating or assessing the PCF mostly reference the standard from ISO 14067. As mentioned, the derivation of data contributing to the PCF in Scope 1,2 and 3 emissions concerning material, logistic relations or manufacturing processes within the supply chain is costly and at the same time associated with many assumptions. Hence, the scientific literature deals with the utilization of intelligent methods and simplified LCA approaches to minimize this effort.

ML methods applied for PCF/LCA: Ghoroghi et al. (2022) provide a comprehensive review of leveraging ML methods to deliver LCA solutions. The authors conclude that ML can be a useful tool in certain aspects of the LCA, especially in optimization scenarios. The specific relation of ML and the derivation of carbon emissions is not explicitly done. Here, Xikai et al. (2019) apply a regression model to assess the carbon emissions of buildings by using its design features. On a high level for urban, street, building and household scale Milojevic-Dupont and Creutzig (2021) utilize ML to optimize urban planning for accelerating, improving and transforming urban infrastructure provision, the carbon emissions are on considered factor. The research of Ahmed et al. (2022) also provides insight into carbon emissions and the influencing factors on a high level. They analyze the impact of energy consumption, financial development, gross domestic product, population, and renewable energy on CO_2 emissions in order to build a LSTM-model to reduce CO_2.

A closer relation to product-related emissions is given in the area of the food industry. Milczarski et al. (2020) provides an ML approach to validate the quality of the pro-duction process and its parameters as the carbon emissions are subject to daily fluctuation due to humans influencing the production system and system errors itself. Also Abdella et al. (2020) investigate the food chain with ML techniques (logistic model) to discover the major environmental impacts to the total carbon footprint.

One ML approach contributing to the product development process for discrete goods is shown in Wisthoff et al. (2016) aims to quantify the general environmental impact potential design decisions have on a consumer product. The authors use a multi-layer perceptron neural network with backpropagation training to relate the LCA impact of 37 case study products to product attributes. The influence of logistical factors is excluded from this research.

Based on our research focus on manufactured discrete industrial goods, a generalized and customized framework for PCF estimation with ML is, to the best of the author's knowledge, not documented in scientific literature.

Simplified LCA: A part of the results of a LCA is the Global Warming Potential (GWP) expressed by the CO_2- Equivalent (CO_{2_e}) as a measure of carbon dioxide emissions e.g. for a product. To enable perform LCAs with a reduced effort, simplified LCA tools are present. A holistic review of their application in product design education is shown in Suppipat et al. (2021). The authors conclude, that the tools are valuable to foster environmentally friendly product re-(design). Present limitations lie in the data preparation process as well as the availability of the database related to new types of materials, manufacturing processes, and short-distance transportation modes.

In Beemsterboer et al. (2020) five simplification strategies for LCA are discussed and upcoming concerns are presented. An Exclusion strategy targets a decrease of impact categories and supplier stages to the cost of information losses. The *Inventory data substitution* aims to reduce pro-cess models by black-boxes. The risk is to adjust a model through the data available. The strategy of a *Qualitative expert judgment* is strongly based on the reliability of the evidence. The *Standardization* with pre-described product-specific required an exact fit of the product system. As a last simplification, the authors discuss the *Automation* strategy which utilizes computational and automated data integration methods. Here, concerns are about the balance of modeling time with expected gains as well as transparency issues. Also, the review of Gradin and Björklund (2021) identifies and analyzes simplification categories in LCA.

Molina-Besch and Pålsson (2020) present a tool to sup-port the food industry towards the developing process of product-packaging combinations that reduce the total environmental impact in food supply chains. Here, the LCA im-pact criteria are sorted into four areas that represent the life-cycle steps packaging material production, transport, house-hold and end-of-life. The authors constitute benefits in the parallel assessment of eco-efficiency and eco-effectiveness criteria from a life cycle perspective and the evaluation of the influence of packaging on food waste in households.

Based on the brief literature review for ML techniques in PCF/LCA and simplifications made in LCA, our contribution tends to close a gap in the aggregation of the PCF calculation of discrete products in four main categories due to the transparency issues and provide a flexible edi-torial framework for holistic data integration. Further, the implementation of a self-optimizing ML-modeling module offers the possibility the gain CO_2 values out of similarity measures rapidly.

3. CO_2 editorial framework

The previous paragraphs show the intention of our re-search to provide a flexible framework to compute the PCF for a broad range of products. Their emissions are mainly traced back to the material, processes as well as logistic impact.

3.1. Requirements and methods

The objective of the tool is to provide a quick and simple way to calculate the CO_2 emissions of products using ML. On one hand, it provides users and companies with a fast and straightforward way to estimate their product emissions. On the other hand, it serves as an educational tool, effectively illustrating the emissions associated with various products and their influential factors.

An essential requirement for the tool is simple and barrier-free handling for the user. Firstly this requires a user-friendly and self-explanatory interface of the tool. Further-more, the tool should be hardware-independent and acces-sible from any platform without the need for installation. Another main requirement is the configurability of the input data. It should be possible to compile the necessary data individually. This should be done without changing the source code of the tool via a user-interface-based configurator. In addition, diagrams and ML models created using the tool are to be downloaded and reused by the user. Finally, the tool was required to be comprehensible and illustrative to the user in order to promote knowledge and awareness of PCFs.

In order to best meet the previously described require-ments, a web application was chosen as the platform for MINDFUL. This allows users easy and hardware-independent access without the need to install software on their own devices. Python was picked as the programming language for the backend, as it provides tools for structuring data in the form of the pandas library and utilizes the scikit-learn library for essential ML components. The Flask framework was chosen to implement the frontend web components, as it seamlessly integrates with Python and allows the design of webpages using the HTML5 standard, ensuring compatibility and flexibility of the tool.

For the calculation of emission values and as a basis for ML, a 4-factor model is used that includes materials, processes, upstream and downstream logistics. The emission values of these four categories are calculated separately and added together to determine the total emissions of the respective product. Emission factors play a crucial role in this calculation process, aiding in estimating CO_2 emissions associated with each component of the 4-factor model. These factors delineate the CO_2 emissions of specific aspects or resources based on their respective quantities within the product. Sourced from databases of various independent research institutions, these factors ensure a thorough and unbiased calculation of the PCF, thereby enhancing the generalizability and independence of assessments.

The emissions of the individual sub-materials of the products are calculated in equation (1) by multiplying the respective mass (m) of each material by its corresponding emission factor.

$$Emissionmaterial = m \cdot factoremissions \tag{1}$$

Table 1
Emission factors material according to Umweltbundesamt (2023)[1], Bundesamt für Wirtschaft und Energie (2023)[2], Bradford et al. (2021)[3], open-co2net (2023)[4], Weisz et al. (2019)[5], City of Winnipeg (2012)[6], Yung et al. (2018)[7].

Material	Emission factor	Source
	(kg CO_2 per kg)	
Iron	1.344	[1]
Green iron	0.00228	[1]
Steel	1.569	[1]
Green steel	0.0116	[1]
Copper	2.56	[1]
Green copper	0.059	[1]
Aluminium	9.274	[1]
Green aluminium	0.0463	[1]
Steel sheet	2.38965	[2]
Steel sheet galvanized	2.49368	[2]
Steel sheet stainless	5.18	[2]
Aluminium sheet	10.65306	[2]
Copper sheet	7.18989	[2]
Lead	1.36275	[2]
Nickel	19.89619	[2]
Polyethylen (PE)	2.59	[1]
Polypropylen (PP)	3.53	[1]
Polyoxymethylene (POM)	2.50	[1]
Polyethylenterephthalat (PET)	3.08	[1]
Polyurethane (PU)	3.8841	[1]
Polyamide (PA6)	8.60	[3]
Viscose	14.00	[4]
Nitrile rubber	11.70	[5]
Kaliumperchlorat (KCLO4)	5.09	[6]
Natriumchlorid (NaCl)	0.20	[6]
Kaliumhydroxid (KOH)	1.94	[6]
Battery Li-ion	10.5	[6]
Battery alkaline	0.07	[6]
Battery NiCd	0.21	[6]
Battery NiMd	57.4	[6]
(kg CO_2 per cm2) Printed PCB	0.028	[7]

The emission factors relating to the different materials and their associated sources are listed in Table 1.

The emission values regarding the production processes are calculated using equation (2), which involves the respective energy consumption (P) and the duration of the process (t). These two parameters are then multiplied by the corresponding emission factor of the energy source used.

$$Emissionprocess = P \cdot t \cdot factoremissions \tag{2}$$

The emission factors in relation to the forms of energy used within the processes are provided below in Table 2.

The KNN regression algorithm stores all available cases and uses them to predict values based on a similarity measure (Patnaik et al., 2021). The algorithm estimates the value of a new data point by averaging the labels of its k-nearest neighbors, determined by a chosen distance metric. According to Müller and Guido (2018), the strengths of KNN are its simplicity and ability to provide reasonable performance without many adjustments, making it a recommended baseline method to try before exploring more advanced techniques. According to Géron (2022), RF regression combines multiple decision tree classifiers trained on different random subsets of the training set. The final prediction is made by aggregating the predictions of the individual trees within the ensemble. RF regression is described by Müller and Guido (2018) as currently among the most widely used machine learning methods due to its effectiveness and versatility, often delivering excellent performance without the need for extensive parameter tuning. Moreover, Li et al. (2018) elaborate that RF substantially mitigates the risk of overfitting.

With regard to upstream and downstream logistics, the emission values are calculated using equation (3), which considers the distance traveled (s) as well as the total weight (m) of the product. These two variables are multiplied by the emission factor for the associated means of transport.

$$\mathrm{Emissionlogistics} = \mathrm{m} \cdot \mathrm{s} \cdot \mathrm{f\ actoremissions} \tag{3}$$

The emission factors of the considered means of trans-port are listed in Table 3 due to its ensemble nature and the utilization of randomized subsets of the training data for each tree.

3.2. Machine learning algorithms

With the goal of MINDFUL serving as a web-based application that autonomously trains models without direct access to the source code, we prioritize aspects such as minimal required hyperparameters and robustness while still striving for the most precise results attainable. Therefore, K-Nearest Neighbor (KNN) regression as well as Random Forest (RF) regression were chosen.

3.3. Software architecture

Abstractly, the basic software architecture of MINDFUL consists of four modules. The development was based on an iterative growth model. Each of the four modules im-plements a partial aspect of MINDFULs

Table 2
Emission factors process according to Bundesamt für Wirtschaft und Energie (2023).

Energy source	Emission factor (kg CO_2 per kWh)
German electricity mix Lignite	0.435 1.053782
Hard coal	0.8733118
Natural gas	0.4325312
Oil	0.841253
Nuclear energy	0.0677879
Photovoltaic	0.06673
Onshore wind	0.010497
Offshore wind	0.005998
Renewable energy	0.00

Table 3
Emission factors traffic according to Schmied and Knörr (2011).

Vehicle	Emission factor (kg CO_2 per kg km)
LKW small	0.000229
LKW medium	0.000179
LKW large	0.000106
LKW extra large	0.000068
Train, electric	0.000018
Train, diesel	0.000026
Ship, inland	0.000034
Ship, oversea	0.000017
Airplane	0.000531

functionality built upon the previous module. The schematic structure of the software architecture as well as the functional scope and the relationships between each module are shown in Fig. 1.

The first module implements the necessary functions for the user to configure which data to use for the calculation and later on for the training of the ML-model (see Fig. 2). As previously discussed the calculation is based on a 4-factor model including materials, processes, upstream and downstream logistics. Furthermore, the user has the option between a default dataset or uploading a CSV file containing a personal dataset. The tool contains the calculation factors for a variety of forms of energy but not for mixtures of these. However, it offers the user the possibility to configure an individual energy mixture to customize the calculations even further. Each subsection of the configurator contains user elements to select the corresponding data from the uploaded file and link it to the corresponding parameter. Therefore the user is able to select the data that should be considered in the different categories without the need to change anything in the source code of the tool itself.

The function of the second module contains on one hand the structuring of the selected data into a dataframe and on the other hand the calculation of the corresponding CO_2 emissions. This process is handled automatically in the backend with no necessary interaction with the user.

The third model revolves around the training of the machine learning model. The values for CO_2 emission are the target values and the rest of the dataframe is used as inputs. Based on the data, a regression model is then trained. For the training, the user can choose between three options of configuration corresponding to his personal experience and knowledge of ML. These choices are discussed in more detail in the Section Implementation. Based on the selected option the model is trained in the backend and stored as the basis for further applications.

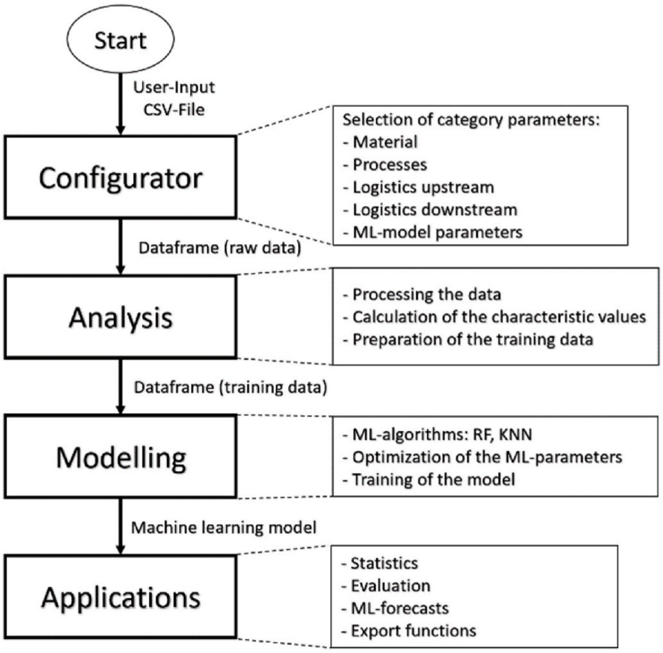
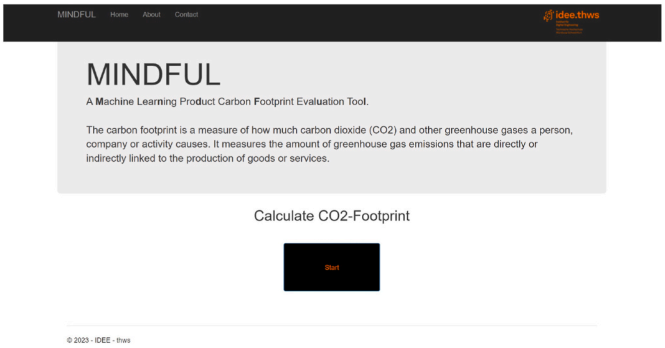


Fig. 1. Software architecture schematic.



MINDFUL

A Machine Learning Product Carbon Footprint Evaluation Tool.

The carbon footprint is a measure of how much carbon dioxide (CO2) and other greenhouse gases a person, company or activity causes. It measures the amount of greenhouse gas emissions that are directly or indirectly linked to the production of goods or services.

Calculate CO2-Footprint

© 2023 - IDEE - thws

Fig. 2. Start page of MINDFUL.

The final module implements a user interface with access to various applications. These functions include graphical and data analysis of the products. Furthermore, the users can view the results of the model training and have the option to download the model or a CSV file containing the uploaded data along with the calculated CO_2 values.

3.4. Implementation

The structure of the MINDFUL framework can be divided into three sub-areas. These include HTML5 tem-plates, Python code and static resources. The latter category includes images and files that are used within the pages of the web interfaces. The HTML5 templates are used for the graphic design of the web interface pages and for entering or selecting user data via forms. Some templates also include dynamically rendered components that can be adjusted at runtime.

The programming language Python was chosen for the implementation of the tool. The central advantages of the language are its widespread use and large number of avail-able libraries as well as its versatility and cross-platform compatibility. The language is well suited for structuring and managing data as well as the implementation of machine learning components. The Flask web framework is used to connect the HTML5-based components of the front end with the Python-based back end. The primary function of the framework is the implementation of routing and dynamic rendering of web pages as well as transferring data between the back and front end. Furthermore, it is possible to save user information in a "session' element based on the user's browser without the necessity to store and manage data on the webserver.

Upon launching MINDFUL, the initial start page is the first to load (cf. image 2). This page includes informative texts along with options for accessing "about' and "contact' details as well as a button to start the actual application.

After pressing the "start' button, the user enters the "configurator' module. First, the user can choose whether to use the default data set or upload his own personal data set from a CSV file. If a file is to be uploaded, the user is then redirected to a page that allows the file to be selected and uploaded based on input elements and the corresponding dialog window of the operating system. The next decision is whether the user wants to use only one form of energy for the production processes or assemble a personalized "energy mix'. If that is the case, the page "configuration energy mix' is loaded. On the one hand, this page consists of ten vertically arranged input fields for the most common forms of energy. The mix is configured by entering the percentages in the corresponding fields. In addition, a pie chart is generated employing JavaScript embedded in the page to graphically illustrate the energy mixture entered and its components (cf. image 3) (see Fig. 3).

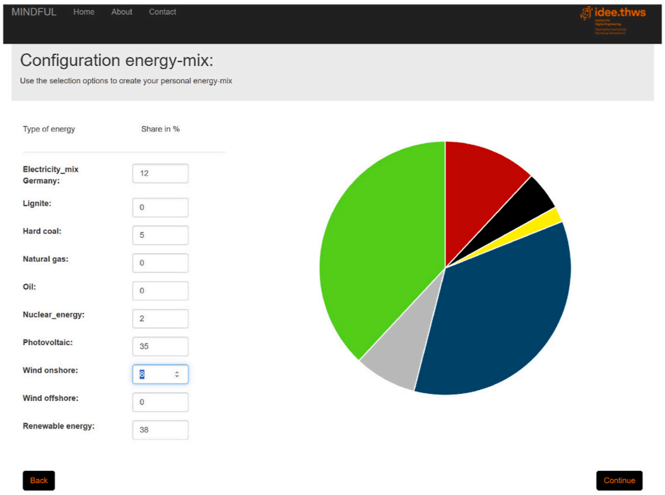
Based on the used CSV file, the selection options are provided in the following pages of the configurator in the form of column headings. This concludes the entry of the basic information and the configuration of the sub-aspects of the 4-factor model begins. Each of the four aspects, material, process and up- and downstream logistics, are configured one after another. Each with its specific configuration template including the necessary inputelements. Fig. 4 shows a partial page of the configurator using the example of material.

When users choose data within the corresponding inputelements, they establish associations with the different parameters. By pressing the “next' button, the data selected on the current page is first saved inside the session and afterward, the user is forwarded to the next page of the configurator. Once the configuration process is complete, the selected data is structured into a data frame. The configuration process of the four sub-aspects of the model is shown schematically in Fig. 5.

After all sections of the configurator have been run through, the “analysis' module is the next to follow. This module includes the creation of the dataframe based on the configuration as well as the calculation of the emission values. The information required for this is stored within the session. First, all the keys contained in the session are run through and these and their corresponding values are checked. A value is only transferred to the dataframe if it is not a control variable of the tool or a variable that is only used in the configurator for its customization. Furthermore, only values from fields in which something has been entered, i.e. the content does not correspond to “no selection', are transferred. For all elements where this is the case, the associated column is extracted from the CSV file and inserted into the data frame under the corresponding key.

In the next step, the user can select whether they already have emission values for their products. If so, they can select these values as the target value for training. If no values are available MINDFUL is able to calculate them accordingly. The last section of the configurator is to set the parameters for training the ML model. As described above, the user has three options for configuring the model. The first option is to train a default Random-Forest model without specify-ing any parameter. All hyperparameters of the model are automatically optimized by grid search based on the data set used. The ratio of test to training data is set at 20%. The second option grants access to edit the test-train-split for the training as well as the option to choose between training a KNN and a RF Model (cl. Fig. 6). Furthermore, it contains access to the third option to choose between optimizing the hyperparameter automatically with a grid-search or configuring them manually. The selected model parameters are stored in the dictionary “training_parameter' which in turn is stored in the session.

Thereby the data are automatically converted into a fitting data type (int or float). This is necessary to ensure that the calculations can be carried out correctly with the used data. Subsequently, the emission factors are calculated based on the formula and emission factors discussed in Section 3.1. The results are added afterward to the data frame.

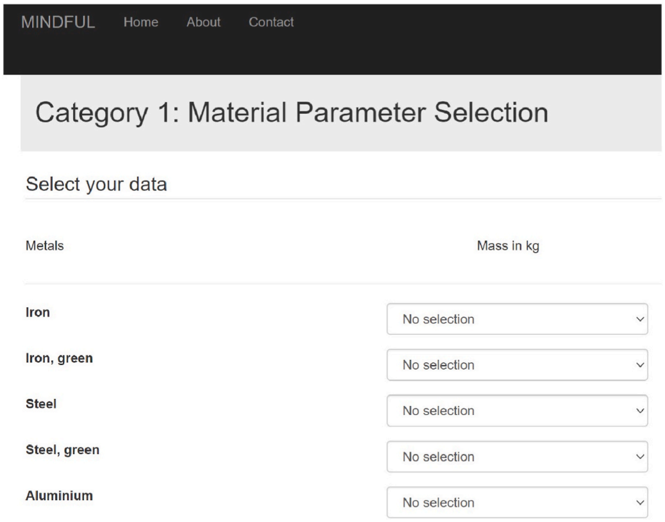


Configuration energy-mix:

Use the selection options to create your personal energy-mix

Type of energy	Share in %
Electricity_mix Germany:	12
Lignite:	0
Hard coal:	5
Natural gas:	0
Oil:	0
Nuclear_energy:	2
Photovoltaic:	35
Wind onshore:	8
Wind offshore:	0
Renewable energy:	38

Fig. 3. Configuration of personal “energy mix".



Category 1: Material Parameter Selection

Select your data

Metals	Mass in kg
Iron	No selection
Iron, green	No selection
Steel	No selection
Steel, green	No selection
Aluminium	No selection

Fig. 4. Extract from configurator page material.

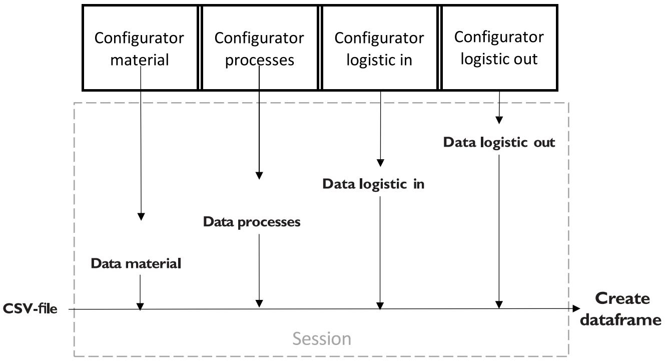
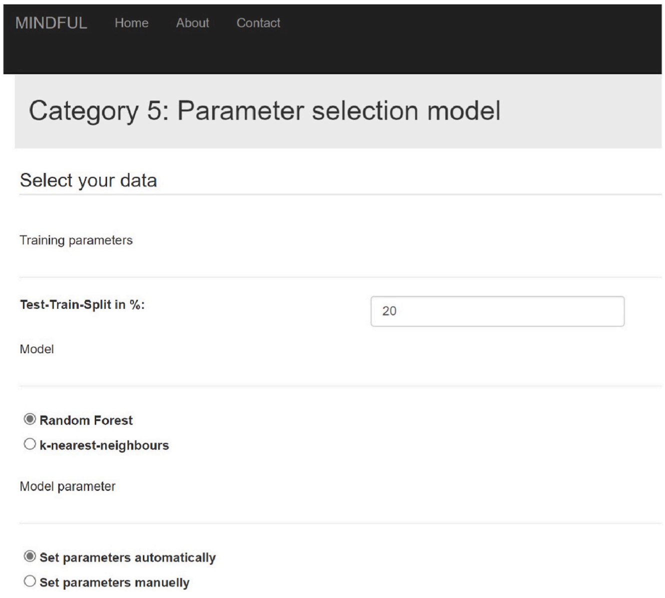


Fig. 5. Data acquisition and structuring process.



Category 5: Parameter selection model

Select your data

Training parameters

Test-Train-Split in %: 20

Model

● Random Forest
○ k-nearest-neighbours

Model parameter

● Set parameters automatically
○ Set parameters manuelly

Fig. 6. Selection options of ML model parameters.

The “modeling’ module implements the training of the ML model based on the parameters entered and the dataframe created in the previous module. If available the user-configured parameters are taken from the dictionary “training_parameter’ stored in the session. Depending on the configuration, either a RF or KNN regression model is trained based on the scikit-learn library. If hyperparameters are specified, they are used directly for the training. However, if this is not the case, they are automatically optimized utilizing a grid-search. Within the grid search for KNN, the parameters are varied within a range of 3–50 for the number of neighbors, and the weight metric is adjusted between ‘uniform’ and ‘distance’. In RF, parameter ranges span from 10 to 300 for the number of trees, 0 to 200 for maximum depth, and 2 to 10 for minimum samples split. After successfully completing the model training, the finished model is saved in ". pkl’ format using jobliblibrary as the basis for later applications.

The last module, “applications’, implements the various functionalities of MINDFUL based on the data provided and the ML model. The individual functions are arranged in tile-shaped buttons within the main menu and sorted by category (cl. Fig. 7).

The statistics functions enable users to visualize emission values and their components for individual products. The corresponding buttons can be used to select whether the total data or each of the four individual categories is to be visualized. They are depicted in the form of a horizontally arranged stacked bar chart. Within the diagram, the products are arranged in descending order depending on their respective emission values. All charts are generated using Python with the Matplotlib library, based on the configured dataframe. They are then loaded into a base64 buffer and transferred to the HTML template for display on the website. Additionally, users can download and save the charts in. png format using the corresponding button.

The “Evaluate Model’ function allows users to analyze the performance of the ML model in fitting the input data during training. The function provides a graphical comparison between the model predictions and the target values of the training data. Additionally, key metrics such as ac-curacy, mean squared error, and root mean squared error are displayed on the right side of the evaluation interface. Therefore, the user can validate the training results using the provided information and, if they are not satisfactory, repeat the training with different hyperparameters or select a different ML algorithm.

The “Similarity’ function is employed to assess the sim-ilarity between two individual data points. The similarity between the two products is computed using normalized euclidean distance, where a value of zero denotes no sim-ilarity, and a value of one signifies that the two data points as well as their respective components and emission values are identical. The results are presented as a heat map.

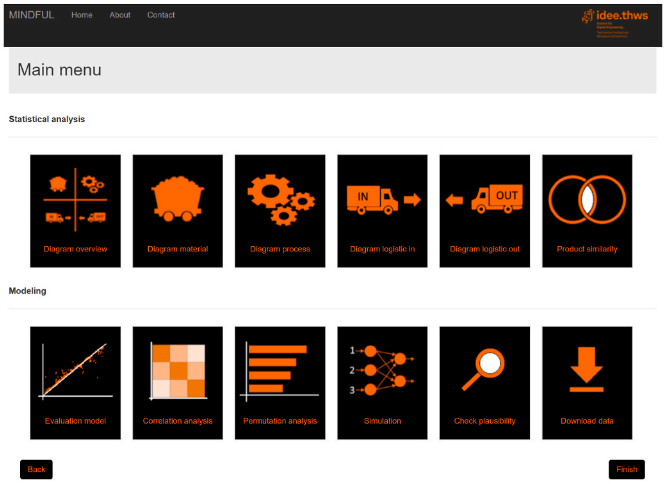


Fig. 7. Main menu of MINDFUL.

To explore correlations between the variables within the dataset, the ‘correlation’ function is applied. This involves calculating correlation coefficients between all data points within the dataset. Afterward, the results are then visualized graphically through a heat map.

The “Permutation’ function can be used to perform a per-mutation analysis on the trained model. This can be used to determine how great the respective influence of the input variables is on the output of the ML model. As a result, it is possible to evaluate the relevance of the individual inputs and, if necessary, to decide which should be disregarded. The shares of each input variable are then graphically displayed in the form of a vertically aligned bar chart, sorted descend-ing by importance.

Using the “Plausibility’ function, the individual CO_2 emis-sion values can be compared with each other to detect any deviations and thus determine their plausibility. For this purpose, the real PCF data, the calculated data and the data predicted by the ML model are displayed graphically in the form of a bar chart with three corresponding bars per product. This makes it possible to graphically determine the plausibility of existing values or to detect deviations.

The MINDFUL tool also offers the functionality to forecast emission values for new data sets, e.g. planned products, based on the previously trained model. The “Simulation’ function can be used for this purpose. The web interface dynamically generates the required input fields for every input variable of the trained model. The desired values can then be entered into these. After pressing the “calculate"- button the emission value is determined based on the ML model. The previously entered input values as well as the predicted PCF are afterward listed in a table.

A “Download’ function is also implemented in MINDFUL. Users can utilize this feature to download a zip folder containing both, the training data and the calculated values in the form of two CSV files as well as the trained machine learning model stored in the. plkformat.

4. Case study

To evaluate the functionality of MINDFUL, it is to be evaluated based on an industrial application test-dataset. The results of the calculation and the modeling by machine learning using the 4-factor model will be compared with the data of a professional and detailed analysis of CO_2 emissions. Based on this, the accuracy and effectiveness of the tool is evaluated.

4.1. Case study setup

The dataset provided for this case study consists of four products including their respective assemblies. Each product contains a minimum of three and a maximum of five assemblies. The dataset consists of a total of 18 data points. For data protection reasons, the product names and their components are anonymized and referred to below as main products (MP), assembly (ASS) and components (C) and numbered in ascending numerical order. For reference, Fig. 8 provides a graphical representation of the basic structure of the dataset used.

The products primarily comprise various types of plas-tic materials, encompassing thermoplastics such as PA6, POM, PP, and PE. Additionally, the composition includes viscose and nitrile rubber. Furthermore, metal components are included, comprising stainless steel sheets, as well as electrical elements like a battery and a printed circuit board (PCB). In addition, the dataset contains the characteristic data regarding energy consumption, duration and form of energy from both the suppliers’ processes and the company’s own manufacturing and assembly processes. Moreover, the individual components were delivered to the company by truck (12 - 24t) and the corresponding routes were documented. Finally, the CO_2 emission values determined by the company are given for each product or assembly. These are used as reference values to evaluate the accuracy of both the calculation (cf. Chapter 3.1) and the modeling of the dataset using ML. When conducting the study the dataset is loaded into MINDFUL, the necessary data is selected and

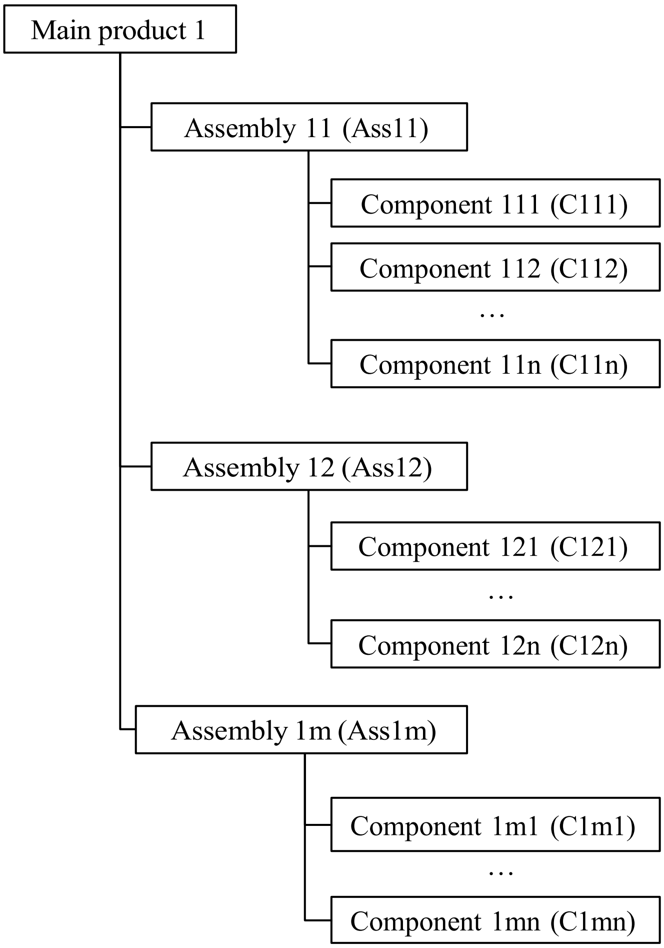


Fig. 8. Product structure of the used dataset.

both the analysis and the training of a random forest model are carried out. The model undergoes training using a test-train-split of 20:80, alongside the automatic optimization of model parameters through grid search. The resulting three emission values per product are then compared with each other and the respective deviation from the PCF determined by the company is determined.

4.2. Case study results

To evaluate the outcomes of the conducted test, along-side the recorded CO_2 emission values for each product in the dataset, the corresponding emission values generated through the analysis and predictions of the machine learning model were also documented. In addition, the deviations between real and calculated values and between real and ML predictions are calculated. These are each given as a percentage deviation from the real PCF values. The determined parameters are listed in Table 4.

The analysis data exhibited a mean of 17.04% with a standard deviation of ±41.87% and a median of 1.16%. The predicted values of the random forest model, on the other hand, resulted in a mean value of −3.36% with a standard deviation of ±10.11%. The median value here was 0.00%. Fig. 9 illustrates the case study outcomes graphically. The data is presented in the form of a bar chart, where the 3 bars of each product are grouped together.

During the calculation of the PCF using MINDFUL, it is observed that the majority of values fall within a range of $\leq |5\%|$, except for four noticeable spikes. Further examination of these particular datapoints reveals that these spikes cor-respond to assemblies containing the printed circuit boards (PCBs) of the respective products. The disparities between the company's emission values and those calculated by the tool arise from variations in the CO_2 factors assigned to raw materials. In the case of PCBs, the company applied a lower CO_2 factor than used by the tool during their calculations. Furthermore, MINDFUL only utilizes the data from the 4- factor model for the calculation, whereas the company's data is analyzed with more comprehensive factors and a higher number of parameters.

Table 4
Results of the case study on the provided dataset.

Product	Real PCF	Calculated PCF	ML predicted PCF	Δ real - calculation in %	Δ real - ML in %
MP1	1.87	1.91	1.86	2.14	−0.53
MP2	1.59	1.63	1.59	2.52	0.00
MP3	1.58	1.59	1.59	0.63	0.63
MP4	1.78	1.81	1.75	1.69	−1.69
ASS1	0.45	0.45	0.44	0.00	−2.22
ASS2	0.30	0.38	0.23	26.67	−23.33
ASS3	0.15	0.15	0.11	0.00	−26.67
ASS4	0.07	0.07	0.07	0.00	0.00
ASS5	0.33	0.32	0.34	−3.03	3.03
ASS6	0.65	0.68	0.67	4.62	3.08
ASS7	0.09	0.18	0.09	100	0.00
ASS8	0.04	0.04	0.04	0.00	0.00
ASS9	0.07	0.07	0.07	0.00	0.00
ASS10	0.37	0.37	0.38	0.00	2.70
ASS11	0.60	0.69	0.62	15.00	3.33
ASS12	0.33	0.32	0.34	−3.03	3.03
ASS13	0.65	0.68	0.67	4.62	3.08
ASS14	0.20	0.51	0.15	155.00	−25.00
Mean value				17.04	−3.36
Standard deviation				41.87	10.11
Median value				1.16	0.00

The variation in values observed in the ML model is attributed to the limited size of the available data set. The initial set consisted of 18 data points, which was expanded to a training data set of 50 by using random resampling. As a result, the outcomes exhibit three prominent peaks within the range of −27% to −23%, while the rest of the data points display a deviation of $\leq |3.08\%|$. A larger data set would likely enhance the modeling capabilities of the RF model and thus generate better results. Finally, it should also be mentioned that when using MINDFUL, only 5 min, including input and computation time, were required to evaluate the PCF for the dataset used.

5. Limitations and discussion

Azarkamand et al. (2020) present a detailed overview of their development of a standardized tool, based on Excel and Visual Basic software, to calculate greenhouse gas emissions in ports, aiming to address the inconsistency in carbon footprint calculation methods across ports.

Eleftheriadis and Anagnostopoulou (2024) developed a car-bon footprint calculation tool using the LCA methodology, aimed at evaluating the environmental impact of SMEs products or services. Tested in a greek cheese factory, the tool showed potential for cost and environmental impact reductions, indicating its potential applicability across SMEs and plans for broader testing.

Mariette et al. (2022) developed a web-based tool named GES 1point5, employing VueJS and Django frameworks, to estimate greenhouse gas emissions in research settings, enabling carbon footprint analysis across different scales and providing options for anonymous or authenticated usage, while processing input data into CO_2 using emission factors and storing entries for output display. This tool employs a mathematical approach to calculate CO_2 emissions but it is primarily designed to build a database for comprehensive analysis.

The previously examined tools exclusively employ mathematical or

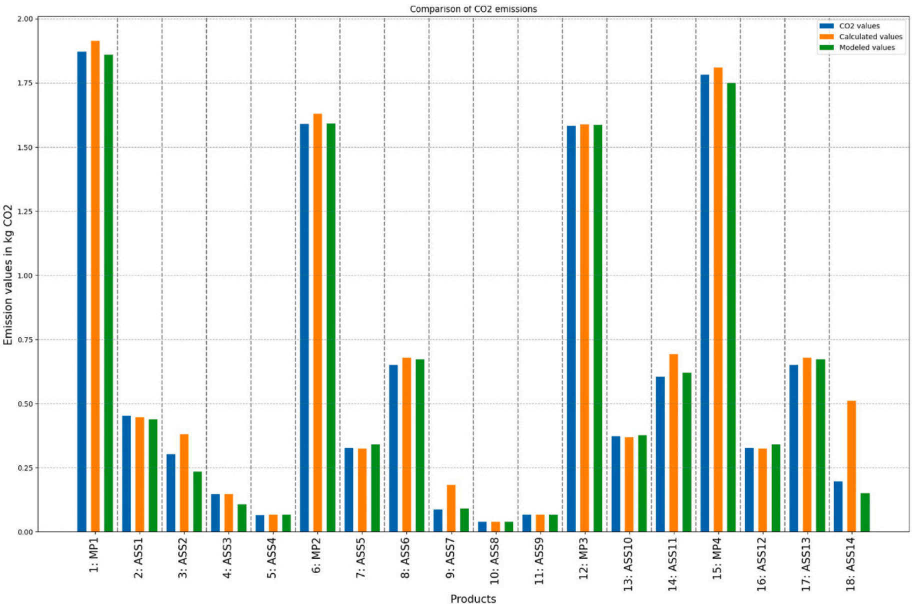


Fig. 9. Graphical representation of the results of the case study.

analytical methodologies for emissions estimation, without integration of machine learning techniques. These tools are specialized for distinct applications and depend on predetermined input parameters. Furthermore, two of these tools operate as web-based platforms, similar to MINDFUL. However, their focus remains solely on CO_2 emissions calculation, lacking validation against empirical data from real-world scenarios. The presented MINDFUL tool provides an application-oriented solution for manufacturing companies to estimate a PCF. With the aim to simplify the amount of necessary information and data, we try to establish a four-factor method. Based on our industrial experience and strengthened by the scientific literature e.g. Flysjö et al. (2014), Laurent et al. (2010), Gaussin et al. (2013) the utilization of the named four factors the majority of emissions are considered in our tool. Nevertheless, incomplete data can influence the PCF result, whereby it must also be taken into account, that also LCA with a high number of input parameters and their uncertainties that contribute to the full life cycle also contain error potential Ghoroghi et al. (2022). Establishing Machine Learning methods will not contribute to make the results more accurate, but our aim is to get a more useable way for reliable results with less expert knowledge. Especially ML can provide (once deployed) an easy-to-use method to estimate the PCF and compare product variants, production scenarios, process alternatives or supply chains by a minimum of input parameters. A further limitation is related to the depth of the value chain, as the material pre-chain is only represented by a fixed emission factor, knowing that specific processing of raw-material can influence the result. We have taken this into account by integrating alternative values for sustainable materials of steel and copper, as these were at least available.

Our MINDFUL framework is designed with the requirement to be an editor, where the user can decide the strategy to derive the PCF. Concerning the named aspects of uncertainty, expert knowledge or input parameter, we see the following options to use the tool to get the most valuable output in terms of reliability:

1. Gain fast analytical results: The user sets up a CSV file with the recommended input parameter without CO_2 emission and the tool calculates the PCF based on the analytical formula shown in Section 3.1
2. Model Training with LCA-data and predict: product portfolio: Hence LCA provide the most detailed data about the product, its processing and logistics, the user provides a small data-set to our tool and trains a machine learning model (normal, advanced or expert model) which predicts the PCF for other product in the portfolio.
3. Self-optimizing model: This option is similar the point
2. If the user has a new dataset calculated by a LCA, the machine learning model is re-trained, which im-proves the model quality.
4. PCF scenario design: The user puts the data for products or component variants in the CSV file, which contain e.g. different logistically relevant locations, material routes (green steel) or design alternatives.

6. Conclusion and further research

Our contribution presents the requirements of the simplified editorial tool to estimate the PCF with ML. Based on this, the underlying method is presented as well as the software architecture to provide an easy-to-use software tool, which is broadly applicable for PCF assessments. To show the functionality a case study is conducted from real data of industrially manufactured products.

Further research activities will focus on building more default training data-sets from a wide range of industrially manufactured goods to provide an initial ML model with a high applicability. Additionally, the user experience needs to be evaluated to improve the software. Also, more pre-defined emission factors especially for different pre-chains of material need to be integrated, as currently sustainable methods for raw material processing e.g. green hydrogen as a primary energy source emerge. This enables the evaluation of different scenarios for the design of the value chain of a product.

CRediT authorship contribution statement

Silvio Lang: Writing – original draft, Validation, Software, Conceptualization. **Bastian Engelmann:** Supervision. **Andreas Schiffler:** Writing – review & editing, Resources, Funding acquisition. **Jan Schmitt:** Writing – review & editing, Project administration, Methodology, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Silvio Lang reports a relationship with Technical University of

Applied Sciences Würzburg-Schweinfurt that includes: employment. Jan Schmitt reports a relationship with Technical University of Applied Sciences Würzburg-Schweinfurt that includes: employment. Bastian Engelmann reports a relationship with Technical University of Applied Sciences Würzburg-Schweinfurt that includes: employment. Andreas Schiffler reports a relationship with Technical University of Applied Sciences Würzburg-Schweinfurt that includes: employment. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data that has been used is confidential.

Acronym Index

CO_2	Carbon dioxide
KNN	K-Nearest Neighbor
LCA	Life Cycle Analysis
MINDFULM	ach**I**ne Lear**N**ing Pro**D**uct Carbon **F**ootprint Eval**U**ation Too**L**
ML	Machine Learning
PCF	Product Carbon Footprint
RF	Random Forest
SME	Small and Medium Enterprise

References

Abdella, G.M., Kucukvar, M., Onat, N.C., Al-Yafay, H.M., Bulak, M.E., 2020. Sustainability assessment and modeling based on su- pervised machine learning techniques: the case for food consumption. J. Clean. Prod. 251, 119661.

Ahmed, M., Shuai, C., Ahmed, M., 2022. Influencing factors of carbon emissions and their trends in China and India: a machine learning method. Environ. Sci. Pollut. Control Ser. 29 (32), 48424–48437.

Algren, M., Fisher, W., Landis, A.E., 2021. Machine learning in life cycle assessment. In: Data Science Applied to Sustainability Analysis. Publisher.

Anshassi, M., Sackles, H., Townsend, T.G., 2021. A review of lca assumptions impacting whether landfilling or incineration results in less greenhouse gas emissions. Resour. Conserv. Recycl. 174, 105810.

Arzoumanidis, I., Salomone, R., Petti, L., Mondello, G., Raggi, A., 2017. Is there a simplified lca tool suitable for the agri-food indus- try? an assessment of selected tools. J. Clean. Prod. 149, 406–425.

Azarkamand, S., Ferré, G., Darbra, R., 2020. Calculating the carbon footprint in ports by using a standardized tool. Sci. Total Environ. 734, 139407.

Beemsterboer, S., Baumann, H., Wallbaum, H., 2020. Ways to get work done: a review and systematisation of simplification practices in the lca literature. Int. J. Life Cycle Assess. 25, 2154–2168.

Bianchini, M., Michalkova, V., 2019. Data Analytics in SMEs: Trends and Policies. OECD.

Bradford, S., Rupf, R., Stucki, M., 2021. Climbing ropes—environmental hotspots in their life cycle and potentials for optimization. Sustainability 13 (2), 707.

Bundesamt für Wirtschaft und Energie, 2023. Informationsblatt CO2- Faktoren. Bundesamt für Wirtschaft und Energie.

City of Winnipeg, 2012. WSTP South End Plant Process Selection Report. City of Winnipeg.

Eleftheriadis, I., Anagnostopoulou, E., 2024. Developing a tool for calculating the carbon footprint in smes. Sustainability 16 (5), 1905.

Flysjö, A., Thrane, M., Hermansen, J.E., 2014. Method to assess the carbon footprint at product level in the dairy industry. Int. Dairy J. 34 (1), 86–92.

Gaussin, M., Hu, G., Abolghasem, S., Basu, S., Shankar, M., Bidanda, B., 2013. Assessing the environmental footprint of manufactured products: a survey of current literature. Int. J. Prod. Econ. 146 (2), 515–523.

Ghoroghi, A., Rezgui, Y., Petri, I., Beach, T., 2022. Advances in application of machine learning to life cycle assessment: a literature review. Int. J. Life Cycle Assess. 27 (3), 433–456.

Gradin, K.T., Björklund, A., 2021. The common understanding of simplification approaches in published lca studies—a review and mapping. Int. J. Life Cycle Assess. 26, 50–63.

Géron, A., 2022. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow. O'Reilly Media, Incorporated.

Laurent, A., Olsen, S.I., Hauschild, M.Z., 2010. Carbon footprint as environmental performance indicator for the manufacturing industry. CIRP annals 59 (1), 37–40.

Li, Y., Zou, C., Berecibar, M., Nanini-Maury, E., Chan, J.C.-W., van den Bossche, P., Van Mierlo, J., Omar, N., 2018. Random forest re- gression for online capacity estimation of lithium-ion batteries. Appl. Energy 232, 197–210.

Mariette, J., Blanchard, O., Berné, O., Aumont, O., Carrey, J., Ligozat, A.-L., Lellouch, E., Roche, P.-E., Guennebaud, G., Thanwerdas, J., Bardou, P., Salin, G., Maigne, E., Servan, S., Ben-Ari, T., 2022. An open- source tool to assess the carbon footprint of research. Environ. Res.: Infrastruct.Sustain. 2, 035008.

Milczarski, P., Zieliński, B., Stawska, Z., Hłobaż, A., Maślanka, P., Kosiński, P., 2020. Machine learning application in energy con- sumption calculation and assessment in food processing industry. In: International Conference on Artificial Intelligence and Soft Computing. Springer, pp. 369–379.

Milojevic-Dupont, N., Creutzig, F., 2021. Machine learning for geographically differentiated climate change mitigation in urban areas. Sustain. Cities Soc. 64, 102526.

Molina-Besch, K., Pålsson, H., 2020. A simplified environmental eval- uation tool for food packaging to support decision-making in packaging development. Packag. Technol. Sci. 33 (4–5), 141–157.

Müller, A.C., Guido, S., 2018. Introduction to Machine Learning with Python: A Guide for Data Scientists. O'Reilly Media.

openco2net, 2023. Viscose/rayon (clothing fabric production). https://www.openco2.net/en/emission-factors/product/viscose-rayon-clothing-fabric-production/129.

Patnaik, S., Yang, X.-S., Sethi, I.K., 2021. Advances in machine learning and computational intelligence. In: Proceedings of ICMLCI 2019. Springer, Singapore.

Schmied, M., Knörr, W., 2011. Carbon Footprint - Teilgutachten "Monitoring für den CO2-Ausstoß in der Logistikkette" -. Öko-Inst. e.V.

Suppipat, S., Teachavorasinskun, K., Hu, A.H., 2021. Challenges of applying simplified lca tools in sustainable design pedagogy. Sustainability 13 (4), 2406.

Umweltbundesamt, 2023. Probas - prozessorientierte basisdaten für umweltmanagementsysteme. https://www.probas.umweltbundesamt.de/php/index.php.

Weisz, U., Pichler, P.-P., Jaccard, I.S., Haas, W., Matej, S., Nowak, P., Bachner, F., Lepuschütz, L., Windsperger, A., Windsperger, B., Weisz, H., 2019. Der carbon fußabdruck des österreichischen gesund- heitssektors. Endbericht, Klima- und Energiefonds. Austrian Climate Research Programme, Wien.

Wisthoff, A., Ferrero, V., Huynh, T., DuPont, B., 2016. Quantifying the impact of sustainable product design decisions in the early design phase through machine learning. In: International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, vol. 50145. American Society of Mechanical Engineers. V004T05A043.

Xikai, M., Lixiong, W., Jiwei, L., Xiaoli, Q., Tongyao, W., 2019. Com- parison of regression models for estimation of carbon emissions during building's lifecycle using designing factors: a case study of residential buildings in tianjin, China. Energy Build. 204, 109519.

Xue, X., Zhang, X., Pang, Y., Gao, F., Xu, M., Lin, S., 2024. A review of machine learning applications in life cycle assessment studies. Sci. Total Environ. 912, 168969. Available online 28 November 2023.

Yung, W.K., Muthu, S.S., Subramanian, K., 2018. Carbon Footprint Analysis of Printed Circuit Board. Elsevier Inc., Kowloon, Hong Kong SAR.