PREDICTION OF WINDOW HANDLE STATE USING MACHINE LEARNING

M. Vollmer[1], M. Langer[1], F. Banihashemi[1], H. Harter[1],
D. Kierdorf[1] and W. Lang[1]
[1]Institute of Energy Efficient and Sustainable Design and Building, TU Munich, Munich, Germany

ABSTRACT

The project described in this paper investigates the energy-relevant behavior of window control actions of the occupants of an office building in Regensburg, Germany as a case study. The extensive data monitoring regarding energy consumption, indoor as well as outdoor climate, and window control actions (state of the window handle) started in 2017.

Different machine learning classification algorithms are used together with the measured data to train models for the prediction of window openings and closings. The procedure is designed to identify the potentials and limitations of the realistic forecasting of occupant behavior based on the available data.

INTRODUCTION

The building sector plays a major role in today's strive for a more sustainable future. It is responsible for more than one-third of global resource consumption. Additionally, the energy consumption taking place in buildings represents approximately one-third of the global end-use energy consumption (Abergel et al. 2017).

Beyond the properties of the construction and the building technology, the occupants' interactions with the building control elements are relevant for the resulting energy demand. Especially the window control behavior impacts the amount of energy needed to maintain a comfortable indoor climate (Gaetani et al. 2018).

The anticipation of the window control behavior is crucial for the two application cases of thermal building simulations and building automation.

In thermal building simulations, the window control behavior is modeled according to appropriate standards or through rule-based functions e.g. according to DIN V 18599-10 (DIN, V. 18599-10: 2018-19) or DIN EN 15251 (DIN, EN 15251:2012-12). The underlying assumptions are unrealistic (e.g. natural ventilation is not considered in DIN V 18599), which leads to an inadequate representation of the true occupant indoor behavior. Thermal building simulations are often used to determine the best measures to increase energy efficiency, reduce resource consumption, and increase thermal comfort. Unrealistic modeling of the window control behavior generates a performance gap between the estimated energy demand and the real energy consumption and there is thus no guarantee that the best possible measures are identified (Moeller et al. 2020).

In the field of building automation, the aim is to control the building technology in such a way that the occupant feels comfortable and that the energy consumption is minimized. In case of a need for fresh air at some point during the cold season, it is more efficient to use the ventilation system with heat recovery than to open the window. To prevent the occupants from opening the window, the building automation must anticipate the occupant behavior as early as practicable and adapt the indoor climatic conditions in time.

A new approach for the more realistic prediction of occupant behavior is offered by the application of machine learning algorithms. The following factors are known to influence the window opening and closing behavior: air temperature, mean radiant temperature, air velocity, relative humidity, air quality (measured by the indicator of CO_2-concentration), outdoor conditions (noise, temperature, humidity, and wind), current window state (closed, open, tilted), clothing insulation, level of activity / metabolic rate, routines, habits and mental states (Fabi et al. 2012). These aspects strongly depend on the respective person, location, and situation.

PROBLEM DEFINITION, GOAL, AND SCOPE

The goal of this study is to identify the potentials and limitations of using machine learning classifiers to detect patterns and predict the occupants' window control behavior. In accordance with the principle of Occam's Razor, it is common practice in the field of machine learning to first try the simplest approach and evaluate its performance (Sammut et al. 2010). It is therefore investigated how well machine learning models perform when the only available input variables are frequently monitored standard measurements even though it is known that a high number of additional factors influence the window control behavior. For the application of a predictive model in building automation, predictions must be made long enough in advance to ensure sufficient reaction time for the building technology. This

 DOI: 10.3217/978-3-85125-786-1-31

study, however, does not focus on a specific application but rather on drawing conclusions about general potentials and limitations. Therefore, the simplest case of predicting the behavior directly after the current minute is defined as a first approach. The machine learning models are trained on data of one zone and thus predict the behavior of the occupants of that zone.

PROJECT AND DATA DESCRIPTION

The data used for this study is collected within the scope of the research project 'Ferdinand Tausendpfund -- Lebenszyklusanalyse und Gebäudemonitoring' of the Institute of Energy Efficient and Sustainable Design and Building of TU Munich in cooperation with the construction company Ferdinand Tausendpfund GmbH & Co. KG (Vollmer et al. 2019). In the scope of this project, an innovative three-story office building is monitored over a time span of four years. The outer walls of the three stories are realized using different solid construction methods. The ground floor is constructed out of reinforced concrete, the 1st floor out of heat-insulating bricks, and the 2nd floor out of sand-lime bricks. All outer walls have an identical U-value of $U = 0.18W/m^2K$. Characteristic parameters of the building are:

- area thermal envelope:
  $A = 1{,}720m^3$
- gross volume:
  $V_e = 4{,}246m^3$
- air volume:
  $V = 3{,}397m^3$
- net floor area:
  $A_{NF} = 1{,}097m^2$

The monitoring concept of the research project provides one monitored reference room per story with the following indoor climatic parameters being recorded:

- operative temperature (°C)
- air temperature (°C)
- relative humidity (%RH)

For this study, data of a room on the second floor with one desk and one window is used. The room is climatically isolated from the rest of the building except for the door leading to the hallway. The positions of the window handles (closed, tilted, or open) are recorded through EnOcean contacts (EnOcean GmbH). These contacts send a signal to a data recorder, which documents the position of the window handle and the time stamp of each transition.

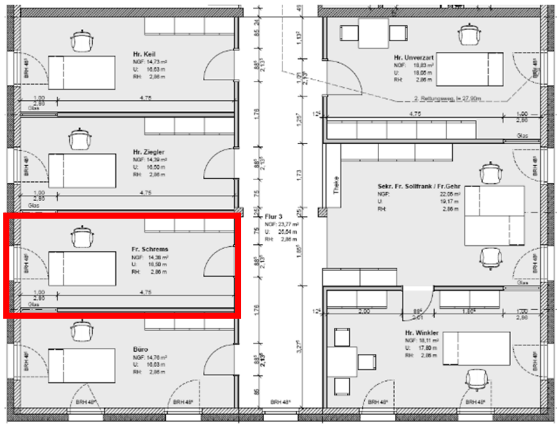

Figure 1: North oriented ground floor plan of the 2nd story

Figure 1 illustrates a north-oriented floor plan of the 2nd floor of the object of investigation where the examined zone is marked in red. Characteristic parameters of the zone are:

- window area:
  $A_W = 2.40m^3$
- window orientation:
  West
- total energy transmittance:
  $g = 0.50$
- heat transmission coefficient:
  $U = 0.87W/m^2K$
- net floor area:
  $A_{NF} = 14.4m^2$

The outdoor climatic parameters are recorded by a weather station on the roof of the building. The following data is recorded by the weather station:

- outdoor air temperature (°C)
- relative humidity (%RH)
- global radiation (W/m^2)
- wind speed (m/s)
- precipitation (mm)

The measured data is centrally combined and stored in a Beckhoff measurement system and uploaded to an SQL database on an external server at regular intervals of one minute (Beckhoff Information System). The monitoring started at the beginning of 2017 and is running until the end of 2020.

RELATED RESEARCH

In a study carried out by Markovic et al. (Markovic et al. 2017), machine learning classification models are trained to determine the window state during a ten-minute time step, given the corresponding values of the input variables (indoor and outdoor climate, occupancy). The trained models can be applied in cases where only the measured variables are known and the goal is to determine the window state for each ten-minute time step. To predict future window control actions, the focus must shift away from the classification of current window states and towards

 DOI: 10.3217/978-3-85125-786-1-31

the prediction of window state transitions (Fabi et al. 2012).

Only by identifying future window state transitions, energetic precautions (e.g. regulation of heating and ventilation systems) can be implemented. The present study focuses on the prediction of window openings and closings.

CONVERSION INTO A MACHINE LEARNING TASK

The nature of the described problem can be assigned to the field of supervised machine learning. It is the subfield of machine learning that learns from data where the output is already known and the goal is to predict the output for unseen data (Burkov 2019). The target outcome is the action on the window handle after each minute. When a window is only slightly open, the ventilating effect may be comparable to the tilted state. For the sake of simplicity, the window state is binarized by only differentiating between the open and closed state. The possible actions after each time step are 'no action', 'window opening', and 'window closing'. Each action can be interpreted as a category, which means that the problem has to be solved by the approach of classification (Burkov 2019). Among the variables measured within the scope of the project, only the following variables are known to have an influence on the ventilation behavior and are therefore selected as input features:

- interior operative temperature (ºC)
- interior air temperature (ºC)
- interior relative humidity (%)
- exterior ambient relative humidity (%)
- exterior air temperature (ºC)
- exterior wind speed (m/s)

In addition to the measured values, the current state of the window is included as a categorical feature. The day of the year and the minute of the day are added to cover the seasonal and daily behavioral patterns. To represent the cyclicality of these variables accordingly, they are transformed into two features respectively by applying the sine and cosine functions. The day of the week is also added as it contains information about weekly patterns and about the instance being on the weekend or not. Since the cyclicality and continuousness of the values are irrelevant in this case, the variable is transformed into a categorical feature.

All described input variables are preprocessed and combined into one feature vector with a corresponding label vector containing the classes of 'no action', 'window opening', and 'window closing'.

SELECTION OF MACHINE LEARNING ALGORITHMS

To build a model for the prediction of the target output, different machine learning algorithms from the scikit-learn library in python are tested and compared to each other in the process called 'spot-checking' (Pedregosa et al. 2011).

Based on the knowledge about the data and the problem at hand, the machine learning algorithms to be tested for suitability must meet the following requirements:

- classification algorithms capable of handling multiclass problems
- more computationally intensive algorithms (e.g. deep learning) are excluded
- able to handle large data sets of a million instances
- consider class imbalances automatically or have a built-in option for setting class weights to balanced mode

When choosing the learning algorithms to spot-check, it is good practice to pick a mixture of parametric and non-parametric algorithms, model-based, tree-based, and instance-based algorithms, linear and nonlinear functions (in case of model-based algorithms) and different learning algorithms for the same type of representation. Under consideration of all described aspects, the following machine learning algorithms are selected for the spot-checking process (Burkov 2019; Perner 2013):

- Decision Tree
- Random Forest
- Gradient Tree Boosting
- k Nearest Neighbors
- Gaussian Naïve Bayes
- Support Vector Machine with Stochastic Gradient Descent
- Logistic Regression

PERFORMANCE EVALUATION METRICS

To be able to test the performance on unseen data, the data set must be split into a training set and a holdout set (test set) with comparable properties. The usual approach to split a dataset is to shuffle the entire dataset and then split it into subsets. For the case of time series data, this approach is unsuitable. The model is only useful if it is trained exclusively on data from the past. The data set is first split while still in its chronological order. In the next step, the examples of each subset are randomly shuffled. The performance on the unseen examples of the test set is decisive for the selection of the best model.

For the quantitative performance evaluation and comparison of different models, one single decision criterion must be defined (Ng 2018). The simplest and most commonly used evaluation metric for

 DOI: 10.3217/978-3-85125-786-1-31

classification problems is the overall accuracy. It represents the ratio of correct predictions to the total number of examples (Burkov 2019). In the present classification problem, the classes are strongly imbalanced. On the entire data set 971,430 minutes are followed by the event ‘no action’ while opening and closing events happen only 190 times respectively. Hypothetically, a model that predicts the class ‘no action’ for each example (independently of the feature values) achieves an accuracy of 99.96% and all predictions for the examples of the underrepresented classes of window openings and closings are wrong. This shows that the accuracy as a single metric is insufficient for the evaluation of the models. It is thus necessary to additionally focus on the underrepresented classes. The recall value of a class is defined as the percentage of all examples of that class which are classified correctly by the trained model (Raschka et al. 2017). In the described case of only predicting the class ‘no action’, the recall value of both underrepresented classes would be 0%. The recall values of the underrepresented classes are a suitable metric for the evaluation of how well the opening and closing events are detected. A different hypothetical model that always predicts the shift towards an open window when the window is closed and towards a closed window when the window is open achieves a recall value of 100% for both underrepresented classes. As a single evaluation metric, the recall value of the underrepresented classes is thus also insufficient. Therefore, an evaluation of the models in two steps is carried out. In the first step, the recall value of the underrepresented classes is inspected, to see if the model recognizes the opening and closing events. Only those models that reach a determined threshold recall value of both underrepresented classes are evaluated further. Since there are currently no predictive models for the problem at hand, there are no comparative values for the achievable success and the threshold recall value must be chosen intuitively. For the purpose of this study, it is decided that a minimum recall value of 80% must be achieved for both underrepresented classes. Optimizing the models towards a high recall value of the underrepresented classes may come at the cost of misclassification of the overrepresented class of ‘no action’. The performance of the models achieving the threshold is thus compared at the overall accuracy in order to determine the best model.

In summary, the two steps for the comparison of the performance of the different classification models are

1. the selection of models that achieve recall values of at least 80% for both underrepresented classes and
2. the comparison of the overall accuracy of only the selected models.

APPROACH FOR TRAINING

In an iterative process, machine learning models are trained for each algorithm, the predictions are evaluated, and optimization techniques are implemented.

The models built by the algorithms Decision Tree, Random Forest, and k Nearest Neighbors are strongly overfitted. The complexity of those models is therefore reduced by adjusting a corresponding built-in hyper parameter. For the adaptation of the hyper parameter, an additional validation set is created. Instead of using 70% of the data, only 60% are used for training and the remaining 10% serve as the validation set.

As an attempt to improve the predictive power of all models, a dimensionality reduction is implemented. This approach leads to a deterioration of the performance of all models and is discarded.

Another applied potential optimization is creating additional informative features from the chronologically previous values. At first, the variable of the duration of the current window state is added to the feature vector. If, for example, the occupant has a routine of briefly ventilating every morning and usually closes the window after about five minutes, this information is contained in the new feature.

The duration of the current window state is not the only information about the short-term past that may be relevant for the occupant behavior prediction. The development of the climatic conditions in the room can also have an influence on the sensation and accordingly on the occupant’s behavior. Since the air temperature and the operative temperature correlate strongly, only the previous values of the operative temperature and the relative humidity inside the room are considered for the new features. Characteristic values are calculated for the period before each example. Regarding the length of this period, the four versions of ten, twenty, thirty, and forty minutes are examined.

First, only the differences between the first and last values of this period are added as a feature. The value difference represents the trend of the parameter during the given period.

As a final attempt of optimization, further variables are created from the periods before each example. Beyond the value difference between the first and last minute of the period, the mean value, the maximum value, and the minimum value are calculated and added as new features.

 DOI: 10.3217/978-3-85125-786-1-31

RESULTS

The best accuracy on unseen data under the condition of achieving a recall of 80% for the classes of window state transitions is 81.16% and is achieved by the Random Forest model after adding the feature of the duration of the current window state. The Gaussian Naïve Bayes model achieves an accuracy of 32.36% on the test set after the optimization steps of adding the duration of the current window state and the characteristic values of the period of ten minutes before each example. The two models built by Logistic Regression with the solvers 'newton cg' and 'lbfgs' both achieve an accuracy of 77.04%. This performance is reached after adding the features of the duration of the current window state and the characteristic values of the previous period of 40 minutes. The Logistic Regression model with the solver 'saga' achieves an accuracy of 71.48% on the test set after adding the duration of the current window state and the characteristic values of the period of ten minutes before each time step as features. The learning algorithms Decision Tree, Gradient Boosting, k Nearest Neighbors, Logistic Regression with the solvers 'sag' and 'liblinear', as well as Support Vector Machine with Stochastic Gradient Descent (with and without kernel approximation), are incapable of reaching the minimum requirement of a recall value of 80% for the underrepresented classes. Table 1 shows the comparison of the results of the successful algorithms.

Table 1:
Comparison of achieved accuracy scores on unseen data

Random Forest	**Gaussian Naïve Bayes**	**Logistic Regression ('newton cg' and 'lbfgs')**	**Logistic Regression ('saga')**
81.16%	32.36%	77.04%	71.48%

ALTERNATIVE PROBLEM DEFINITIONS

The previously described approaches attempt to optimize the creation of a predictive model for the window control behavior in a specific office room after each minute. To find out whether better results can be obtained with the existing data, alternative problem definitions are tested. The results cannot directly be compared to the original version but they indicate whether the alternative approaches are more promising. The feature vector used for these tests is the one having achieved the best result in the original problem definition.

One new problem definition is training machine learning models on the data of an alternative office room with more than one occupant. The goal is to determine whether several occupants together show an averaged behavior with clearer patterns that can more easily be detected by learning algorithms. The data used for this approach is collected in an office room on the eastern side of the ground floor of the same building. It is shared by two occupants and has two windows. No distinction is made whether one or two windows are open. The same evaluation metrics are used as for the original problem definition. Compared to the performance of the original version, two more models achieve the minimum requirement. Furthermore, the performance increases for three of the models that already achieve the minimum requirement in the original problem definition and decreases for two of them. The predictions are generally better for most models for the alternative office room with two occupants. The best achieved accuracy of 80.49%, however, is below the best performance for the original problem definition (81.16%).

The originally defined task of predicting the window control behavior to the minute is very complex. Even if the thermal conditions inside a room are known to be outside the comfort range, it does not necessarily mean that the occupants will directly act on it. Their reaction time can also depend on the type of activity they are involved in at that moment. One more attempt to simplify the problem definition is down sampling the time steps of one minute to time steps of ten minutes. The information on whether the window will be opened at one point during the subsequent period or not is defined as the new label. The closing events are no longer considered. The predictions of these models would be useful for the field of building automation but cannot be applied for modeling the occupant behavior in thermal building simulations. In this version as well, a recall value of 80% for the class of 'window will be opened in the following ten minutes' is set as a requirement. The best overall accuracy for this approach of 84.96% is achieved with the Logistic Regression model with the solver 'sag'.

The exploration of the ground-truth input data shows that window state transitions only happen on weekdays between minute 370 and 1,126 of the day, which is thus the time where window control actions can generally happen. One way to change the problem definition is therefore to eliminate the examples from periods where the window state always stays the same before using the data set for training. Intuitively, what changes is that the model doesn't need to focus on learning the times when events can potentially happen and can instead focus on detecting the patterns during the times of possible actions. The resulting model can only be used to predict the window control behavior from Monday to Friday between 6:10 a.m. and 6:46 p.m. The

 DOI: 10.3217/978-3-85125-786-1-31

number of opening and closing events (examples of the underrepresented classes) remains the same in the new data set and it is, again, required that at least 80% of them are classified correctly. Among all models that achieve this requirement, the best overall accuracy on unseen data of 54.71% is achieved by the model trained by the algorithm Logistic Regression with the solver 'sag'.

INTERPRETATION

The predictions of the model with the best achieved accuracy score for the original problem definition of 81.16% (an optimized Random Forest model) are inspected and interpreted.

The model classifies all window closing events of the test set and 34 out of the 42 window opening events correctly. The focus on the correct classification of the underrepresented classes leads to the prediction of both window state transitions more than 600 times as often as they occur.

The overall accuracy of over 80% may appear satisfying but it is mainly achieved because the model recognizes that the window state is never changed during certain periods. When only considering the predictions for the time during the nights and the weekends, the achieved accuracy is 94.96% and only 52.06% for the period of possible window control actions (workdays between 6:10 a.m. and 6:46 p.m.).

For each day of the 202 days of the test set, a plot is generated showing the window state, the ground-truth label, and the predicted label for each minute for further interpretation of the results.

The first pattern detected from the plots is the prediction of a window opening action every morning around minute 400 of the day (6:40 a.m.) for 15 to 120 continuous minutes. The exploration of the entire dataset shows that window opening events mostly happen in the morning around minute 400 of the day. By predicting window openings every morning, the model successfully detects all actual window openings that correspond to this morning routine.

The predictions for the examples when the window is open are examined. Whenever the time is within the period of potential activity, a window closing is predicted. Only when the window remains open during the night, it is predicted that the window state remains the same.

On 40 of the 202 days of the test set, the model predicts an opening for almost every minute between approximately 6 a.m. and 7 p.m. whenever the window is closed. On 25 of these days, window openings do happen at some point. The predictions for those days are therefore wrong for almost all minutes, but the occurring opening events are classified correctly. On the remaining 15 days of the 40 days, the window remains closed the entire day. These 15 days do, however, show a pattern. They are all on a weekday that is either one day before or after a day where opening events occur. It can thus be concluded that the model does recognize a pattern in the feature values that indicates a higher probability of window opening events occurring on specific days.

SUMMARY AND CONCLUSION

Regarding the life cycle-based resource consumption as well as the environmental impacts, the building sector plays a major role in today's strive for a more sustainable future. A life cycle-based assessment and optimization of buildings is key in finding sustainable solutions. Within the life cycle of a building, the use stage is still responsible for a major share (de Larriva et al. 2014). During the planning phases of a building, the energy demand can be estimated using different methods, e.g. static calculations or dynamic thermal building simulations. There is, however, a discrepancy between the estimated energy demand and real energy consumption due to different aspects. One of the main aspects is the energy-relevant occupant behavior, especially regarding natural ventilation (Moeller et al. 2020). Therefore, a need for a deeper understanding and accurate prediction of the occupant behavior arises. This paper identifies the potentials and limitations of using machine learning classifiers to detect patterns and predict the occupants' window control behavior. To make use of the possibilities machine learning provides, actual data of parameters that potentially impact the occupants' behavior is needed. Within the framework of this study, an extensive energy and occupant behavior monitoring was and still is carried out over four years and data regarding the indoor climate, the outdoor climate, and the window handle state is collected.

In the first step, the data and the problem are analyzed and suitable machine learning algorithms are identified. As an attempt to improve the predictive power of all models, different optimization techniques such as dimensionality reduction, an adaptation of built-in hyper parameters, and the creation of additional features are implemented.

The accuracy and recall values are chosen as performance evaluation metrics. For the recall value, a minimum value of 80% is required for both underrepresented classes of the opening and closing events on unseen data of the test set.

Overall, only the models built by the Random Forest, Gaussian Naïve Bayes, and Logistic Regression algorithms are capable of meeting the minimum requirement. The achieved accuracy is 81.16% for the Random Forest model, 32.36% for the Gaussian Naïve Bayes model, 71.48% for the Logistic Regression model with the solver 'saga' and 77.04% for the two models built by the Logistic Regression algorithm with the solvers 'newton cg' and 'lbfgs'. The learning algorithms Decision Tree, Gradient

 DOI: 10.3217/978-3-85125-786-1-31

Boosting, k Nearest Neighbors, and Support Vector Machine with Stochastic Gradient Descent are unsuited for the task.

Some of the implemented optimization techniques contribute significantly to improved predictive performance. There is, however, no combination of adaptations that works best with all algorithms. Some of the optimization techniques improve the performance of certain models and deteriorate others.

At first sight, the achieved accuracy of over 80% appears to be a satisfactory result. Closer inspection reveals that almost half of all predictions during the potential occupancy period (workdays between 6:10 a.m. and 6:46 p.m.) are wrong. However, certain patterns are recognized successfully. For instance, the model detects the times when nothing ever happens, the morning ventilation routine, and the days when window openings are likely to occur. Attempting to improve the models with the same data but alternative problem definitions does not lead to significant improvements.

Overall, the approaches tested in this study cannot be used for practical applications in thermal building simulations and building automation. It can be assumed that the available data is insufficient for a solution to the described problem.

OUTLOOK AND POSSIBLE FURTHER APPROACHES

To further explore the field of predicting window opening behavior based on the described findings, several additional approaches can be carried out.

One approach could be the training of models with learning algorithms of the field of deep learning. This could make sense given the high dimensionality of the feature vector and the complexity of the problem. Statistical methods from the field of statistical modeling have been developed specifically for multivariate time series data. These methods do not belong to the field of machine learning, but they could potentially be appropriate solutions for the problem at hand.

Furthermore, the problem could be redefined accordingly to one specific application case. For thermal building simulations, the realistic representation of window opening durations is important. It is therefore not particularly important to detect the exact time of the window state transition. A new problem definition could be the prediction of the number of minutes with an open window within a defined time as a regression problem. For the field of building automation, the duration of window openings is less relevant. What is more important is to recognize in advance when the window will be opened in the future. This could also be defined as a regression problem, which predicts the number of minutes until the next window opening event.

Despite applying all described optimizations and simplifications, it may still not be possible to develop a successful model with the available data. As explained in the introduction, the parameters monitored in this study are not the only factors known to influence the window control behavior. For future projects, more information could be gathered as input variables. Sensors detecting the CO_2-concentration, the air velocity, and the presence of occupants inside the room could be installed. The degree of clothing, activity level, and mental state of the occupants also influence the behavior but are challenging to measure. Another aspect that might influence the decision to open the window is the weather forecast for the same and following day.

Based on the findings of this project, a self-sufficient mobile measurement box is currently being developed at the Institute of Energy Efficient and Sustainable Design and Building of TU Munich to address the need to measure more parameters (e.g. CO_2-concentration, air velocity, sound pressure level, indoor air quality) and to be able to measure the indoor climate closer to the occupants. The box will be used in two future research projects.

First of all the box will be used in the research project 'NuOpt Office' which is a project based on an already finished research project called 'Eco+Office ASBau' where a life cycle-based plus-energy and CO_2-neutral office building was developed (Harter et al. 2019). Since the use stage plays a major role, the energy consumption, the indoor climate, and the energy-relevant occupant behavior will be monitored extensively. Special focus will be placed on the occupant-oriented determination of relevant indoor climatic parameters. As a case study, an office building in Hof, Germany will be used.

In regard to the goal of achieving climate-neutral buildings, the building stock also has a major share. The second research project 'NuData Campus' therefore aimes to optimize existing buildings and facilities. As a case study buildings of the University of Applied Sciences Munich will first be analyzed and classified in terms of energy consumption and the type of use. The Institute of Energy Efficient and Sustainable Design and Building of TU Munich is researching the partial aspects of socio-economic modeling of occupant influences and sustainable reference system concepts.

ACKNOWLEDGEMENT

The presented study is partly based on research conducted in the project 'Ferdinand Tausendpfund -- Lebenszyklusanalyse und Gebäudemonitoring', funded by the *Bavarian Building Industry Foundation (BBIV).*

 DOI: 10.3217/978-3-85125-786-1-31

REFERENCES

Abergel, T., et al. (2017). Towards a zero-emission, efficient, and resilient buildings and construction sector: Global Status Report 2017. UN Environment and International Energy Agency: Paris, France.

Beckhoff Information System. Retrieved March 2, 2020 from https://infosys.beckhoff.com/index_en.htm

Burkov, Andriy. 2019. The Hundred Page Machine Learning Book. 2019.

de Larriva, R. A., Rodríguez, G. C., López, J. M. C., Raugei, M., & i Palmer, P. F. (2014). A decisionmaking LCA for energy refurbishment of buildings: Conditions of comfort. Energy and Buildings, 70, 333-342.

DIN, EN 15251:2012-12. (2012). Indoor environmental input parameters for design and assessment of energy performance of buildings addressing indoor air quality, thermal environment, lighting and acoustics; German version EN 15251:2007

DIN, V. 18599-10: 2018-19. (2018). Energy efficiency of buildings - Calculation of the net, final and primary energy demand for heating, cooling, ventilation, domestic hot water and lighting - Part 10: Boundary conditions of use, climatic data

EnOcean GmbH. “868 Mhz EnOcean Für Europa.” EnOcean, Retrieved March 2 from https://www.enocean.com/de/produkte/enocean_ module/

Fabi, Valentina, et al. 2012. Occupants’ window opening behaviour: A literature review of factors influencing occupant behaviour and models. Building and Environment 58 (2012). 2012.

Harter, H.; Meier-Dotzler, C.; Vollmer, M.; Lang, P.; Pfoh, S.: AS-Bau Hof GmbH - Eco+Office -- Plusenergie und CO2-Neutralität. Innovation 2019

Markovic, Romana, et al. 2017. Comparison of Different Classification Algorithms for the Detection of User's Interaction with Windows in Office Buildings. Aachen : s.n., 2017.

Moeller, S., Weber, I., Schröder, F., Bauer, A., & Harter, H. (2020). Apartment related energy performance gap–how to address internal heat transfers in multi-apartment buildings. Energy and Buildings, 109887.

Ng, Andrew. 2018. Machine Learning Yearning. 2018.

Pedregosa, F., Varoquaux, Gaël, Gramfort, A., Michel, V., Thirion, B., Grisel, O., … others. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12(Oct), 2825–2830.

Perner, Petra. 2013. Machine Learning and Data Mining in Pattern Recognition. New York : s.n., 2013.

Raschka, Sebastian and Mirjalili, Vahid. 2017. Python Machine Learning. 2017.

Sammut, Claude and Webb, Geoffrey I. 2010. Encyclopedia of Machine Learning. 2010.

Vollmer, M.; Harter, H.; Schneider-Marin, P.; Lang, W.; Pfoh, S.: Ferdinand Tausendpfund -- Lebenszyklusanalyse und Gebäudemonitoring: Innovation 2019.

 DOI: 10.3217/978-3-85125-786-1-31