Peer-reviewed version available at Appl. Sci. 2019, 9, 3719; doi:10.3390/app9183719

Article

Using different ML algorithms and hyperparameter optimisation to predict heat meters' failures

Przemysław Pałasz [1] and Radosław Przysowa [2,*]

[1] Rapp Enserv, Hochstrasse 100, 4018 Basel, Switzerland; patryk.palasz@rapp.ch
[2] Instytut Techniczny Wojsk Lotniczych, ul. Księcia Bolesława 6, 01-494 Warszawa, Poland; radoslaw.przysowa@itwl.pl
* Correspondence: patryk.palasz@rapp.ch

Featured Application: Design of advanced heating systems for smart buildings and optimisation of stocks and maintenance processes in existing heat meter networks.

Abstract: The need to increase the energy efficiency of buildings as well as the use of local renewable heat sources has caused that heat meters are used not only to calculate the consumed energy but also for the active management of central heating systems. Increasing the reading frequency and the use of measurement data to control the heating system expands the requirements for the reliability of heat meters. The aim of the research is to analyse a large set of meters in the real network and predict their faults to avoid inaccurate readings, incorrect billing, heating system disruption and unnecessary maintenance. The reliability analysis of heat metres, based on historical data collected over several years, shows some regularities which cannot be easily described by physics-based models. The failure rate is almost constant and does depend on the past but is a non-linear combination of state variables. To predict meters' failures in the next settlement period, three independent machine learning models are implemented and compared with selected metrics because even the high performance of a single model (87% True Positive for Neural Network) may be insufficient to make a maintenance decision. Additionally, performing hyperparameters optimisation boosts models' performance by a few percent. Finally, three improved models are used to build an ensemble classifier which outperforms the individual models. The proposed procedure ensures the high efficiency of fault detection (>95%), while maintaining overfitting at the minimum level. The methodology is universal and can be utilised to study the reliability and predict faults of other types of meters and different objects with the constant failure rate.

Keywords: heat meter, district heating, fault detection, predictive maintenance, Machine Learning (ML); Artificial Neural Network (ANN); Bagging Decision Tree (BDT), Support Vector Machines (SVM), hyperparameter optimisation, ensemble model

1. Introduction

Sustainable development and controlling climate change can only be achieved with a safe and low-emission energy system. Its transformation involves, i.a. decarbonization of buildings responsible for approximately 36% of all CO_2 emissions in the European Union. Almost 50% of the consumption of final energy is spent on heating and cooling, from which 80% is used in buildings. This sector is treated as pivotal in accelerating the reduction of emissions of the energy system. It is also a strategic industry in the context of energy safety, since according to the forecasts, till 2030, heating and cooling will account for about 40 % of the consumption of renewable energy sources [1].

Peer-reviewed version available at Appl. Sci. 2019, 9, 3719; doi:10.3390/app9183719

Heat metres are used in central heating systems to calculate the consumed energy both in multi-family houses and single apartments or offices. Smart heat metres ensure that costs are settled based on the actual consumption and contributes to saving heat by inhabitants and to reducing emissions, particularly in buildings with many tenants.

Currently, approximately half of the installed heat meters, which are under research in this article, are read manually. However, year by year increases the number of meters read remotely, which belong to smart meters. It is a fact that more and more electricity meters are connected to the wide-area network, the so-called Smart Grid. The application of smart meters enables regular monitoring of power consumption and access to flexible rates [2,3]. Smart Grids and Smart Buildings feature optimised asset management, increase operational efficiency, ensure stable power supply, allow monitoring of system operation in real time or introduce the network's ability to reconfigure and self-heal.

It is necessary to develop a uniform communication system to introduce the IoT technology for the billing of media and enable the delivery of telemetric data to one node. Such an attempt was made in 2015, creating the OMS (*Open Metering System*) group. Initially, it was intended to be a non-profit organisation gathering leading meter manufacturers. Its objective is to create a protocol of communication as well as to develop a common standard of smart meters (not only electricity meters). To date, the group operates mainly in Europe.

Smart meters and advanced heating systems contribute to reduced emissions by limiting energy grid losses and the possibility to adjust heat supplies to weather conditions [4]. Recording heat consumption with the use of a smart meter can take place in one-hour intervals or even shorter. The information on the meter state can be directly sent to the plant, thanks to which it is possible to monitor and manage it regularly.

In smart cities and buildings, it is of utmost importance to create a prediction system, thanks to which the operation of the heat distribution network will be better adjusted to the needs of recipients. In consequence, it all enables to reduce the transfer losses, costs of heat purchase and pumping costs. Facilitating the measuring system is associated with the improvement of the efficiency of communication with distributed elements, remote control and monitoring of technological processes.

Building smart homes and cities aims to improve the quality of life of its inhabitants and the protection of natural environment. Interdisciplinary cooperation in the area of collecting, storing and analysing a significant amount of data as well as using, servicing, exchanging and maintaining IoT devices is needed to achieve this goal. The issue of reliability of smart meters has a fundamental significance for heat distributors and consumers, since the reliability of collecting data, the correctness of billing of heat and taking proper decisions in managing heat distribution network depends on it.

Topics similar to the subject of this article are discussed in [5] and [6]. The studies on the optimisation of the heating system in residential buildings are carried out in the paper [7]. The Authors utilised an optimised ANN model to determine the optimal start time for a heating system in a building. Also an ANN has been applied in [8] to evaluate the energy input, losses, output, efficiency, and economic optimisation of a geothermal district heating system. This has been used to determine if the existing system is operating at its optimal level, and will provide information about the optimal design and profitable operation of the system.

More recent studies on optimising heat usage can be found in [9]. To minimise the energy consumption and maintain a good comfort level anticipating the thermal behaviour of the building and external disturbances, a neural predictive controller for single-speed ground source heat pumps systems was developed. Still, according to our knowledge, studies dealing with reliability of heat meters are scarce. Furthermore, using a set of different ML models in parallel is unique. We show that these problems are crucial to implement smart buildings.

This article attempts to broaden our knowledge on the operation and reliability of the currently applied heat meters and create an effective and universal model to predict their failures.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 July 2019 doi:10.20944/preprints201907.0319.v1

Peer-reviewed version available at Appl. Sci. 2019, 9, 3719; doi:10.3390/app9183719

2. Materials and Methods

2.1. Heat meters

A heat meter is a microprocessor measuring device which calculates the consumption of heat in kWh by measuring the flow rate of the heat transfer medium[1] and the difference between supply and return temperature (Fig. 1). The meter has an in-built flow meter, which measures the volume of the flowing medium, and two temperature sensors for the inflow and outflow fluid.

Temperature sensors consist of a measuring element, that is a thermistor, which is usually platinum. It changes resistance depending on the temperature. In heat meters usually Pt100 or Pt500 sensors are used. The higher is the resistance of platinum in the sensor (respectively $100\Omega, 500\Omega$), the greater is the accuracy of temperature measurement.

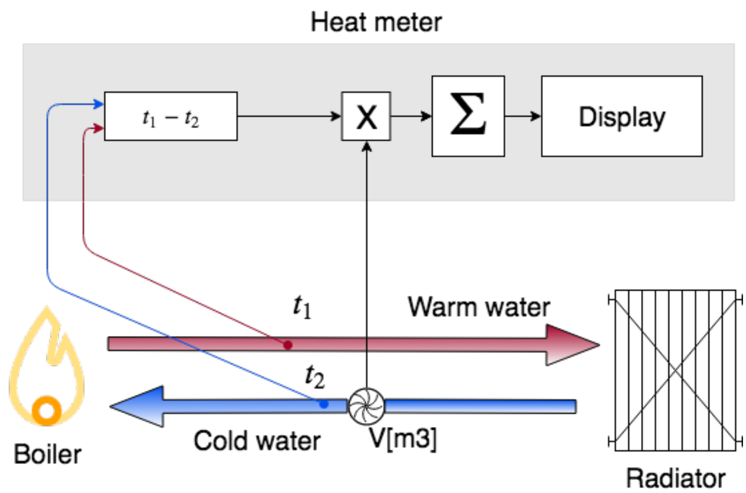

Figure 1. Principle of the heat meter

2.1.1. Maintenance of heat meters

Heat meters are generally considered as reliable. However, like all devices, they sometimes break down and require technical maintenance. Premature meter failures affect all parties, i.e. tenants, building managers and the billing firm, significantly disturbing the scheduled maintenance and causing losses. The consequences include incorrect readings, understated invoices and customer complaints.

The replacement of meters is a complex logistic process, which must be planned in detail. The old meters have to be removed from the network and delivered to a service centre or recycled. The installation of new meters is usually done in stages and it is convenient to carry out the service works in a limited area before proceeding to the next one. It is necessary to arrange meetings with building managers or tenants.

The whole procedure can generate high costs related to the work of qualified installers, planning and delivering of components, final inspection of the installed devices and updating the operational data regarding meters. It is often more efficient to replace a functioning device during a scheduled maintenance check if there are any indications of a possible failure in the next billing period. Developing a method for predicting the occurrence of the meter failure in the subsequent period is the subject of this article.

[1] Most often it is water, although in the case when the system is also used for cooling, it can be water with appropriate additives to prevent freezing.

Heat meters are as a rule the subject of planned maintenance. In the majority of the European countries, their verification should take place every five years due to metrological requirements. During this time, heat meters accumulate mineral deposit from water (especially warm and hot), which results in the wear of mechanical elements due to being moved by the stream flowing through the equipment. [2].

Reverification is time-consuming. Heat meters have to be dismounted from the network and delivered to the verification point. For the verification period, the older heat meters are replaced with the new ones, which have a valid heat meter verification marking. However, nowadays, the majority of the companies plan only one verification visit due to logistical costs. Once the verification is done, the dismounted heat meters are installed in other locations. Before the verification, the heat meters are cleaned, regulated and sealed so that they could work for another 5 years meeting metrological requirements. It is not uncommon that heat meters are verified two or even three times.

The users decide to replace them due to three reasons: new equipment has a longer warranty than the older verified one (5 years instead of 2 years), new heat meter is much more technologically advanced, or the old heat meter was broken, and the repair is impossible or useless.

2.1.2. Review of typical faults

The most common causes of meters' failures include:

- Failure of flow transducer (in case of a mechanical meter) caused by the accumulation of deposit on mechanical elements. The repair involves exchanging the rotor or the whole transducer.
- Failure of temperature meter – most frequently it occurs in the case of non-mechanical heat meters and is usually caused by rodents (rats) damaging the conductors. The repair is based on the replacement of temperature meter.
- Exhaustion of the battery. Despite the theoretical calculations that mounted batteries should work for 5-7 years (exceeding the required verification period), it frequently happens that the batteries run out already after 2-3 years of usage. The repair of such equipment requires only the replacement of the battery.

2.2. Data

Information on installation, operation and replacement of heat meters was accumulated over the last ten years in a relational database. Using a meter consists in cyclical readout of its current value necessary to calculate the energy consumption in a defined billing period. Potential failure should be detected at the time of meter readout at the latest. The billing period usually lasts 12 months (but it can also last 6, 18 or 24 months) and starts at the beginning of the chosen month (often it is January, June or September). Some modern smart meters also store the monthly values – although they do not impact the final billing.

The discussed database also includes many other items of information used for the billing of utilities (approximately 150 relational tables – some of them consist of 20 million records – in total, 250 GB of data) as well as data regarding other types of meters (water meters and heat distributors). However, the authors decided to focus on the heat meters addressed in the introduction. The available data may offer clues to many questions concerning the operation of heat meters. For economic reasons, the most significant problem is detecting and predicting a failure; thus, the data was prepared for this purpose.

After data preparation and initial statistical analysis, sixteen parameters are selected for further processing (table 1).

---

[2] To mitigate this problem, regular checks are done aimed at investigating the quality of water used for heating

Table 1. 16 selected features

Name	Description	Type
Age	Life of the meter in months from the moment of installation to the last readout or failure	ℕ Number
ZIP	Postal code of flat/office, in which the meter is installed	ℕ Number
Floor	Floor where the meter is installed	ℕ Number
FlatType	Type of usable area	String of letters
RoomType	Type of area	String of letters
AccConsumption	Consumption from the moment of registration in the system	ℝ Number
CurrentConsumption	Consumption in the last settlement period	ℝ Number
CommType	Type of communication with the meter	String of letters
Producer	Name of the manufacturer of the meter	String of letters
RatingFactor	Equalization factor	ℝ Number
BillingNo	Next number of the settlement period	ℕ Number
CurrentValue	Last recorded value of the meter	ℝ Number
AvgConsumption	Average consumption in all settlement periods	ℝ Number
MaxConsumption	Maximum consumption in the settlement period	ℝ Number
MinConsumption	Minimum consumption in the settlement period	ℝNumber
CalculatedAge	Calculated age of the meter in months: CurrentValue/AvgConsumption	ℝ Number

The type of the usable surface (e.g. flat, office or storeroom), as well as the type of room (e.g. room, kitchen, bathroom and corridor), were represented by enumerations. In the case of an unclear situation, the value *other* was used. The authors decided to distinguish sixteen types of rooms and five types of usable surface. Since some models accept only numbers, each type of room was assigned a successive natural number. The same was done in the case of meters' manufacturers (16 different values).

Rating factor is a real number from the interval (0-1]. It is used in case of rooms which an increased consumption of heat, e.g. due to adjacency to external walls of the building. This ensures a fair distribution of the heating costs of the whole building between all tenants – irrespective of the fact whether they have an external flat or not.

An important parameter is also the method of communication with the equipment (*CommType*). There are four communication types:

- *bus* – meters regularly send their updates to the central panel installed in the same building which also collects data from other meters – cable connection
- *funk* – similar as bus, but the connection of the meter with the control panel does not require the additional wiring system
- *walk by* – on specific days (programmed) the meter sends data, which has to be collected by the technician sent to the neighbourhood and equipped with the receiving device
- *without a module* – the reading has to be done *manually* directly on the meter

The correction and normalisation of the data set are described in more detail in [10].

2.3. Data analysis

Preparation of data and selection of observation parameters theoretically enables to build prediction models. In practice, however, it is necessary to better understand data, on which the studies will be conducted [11]. By its very nature, a machine learning model is acutely sensitive to the quality of the data and it is of low-quality. Because of the huge volume of data required, even relatively small errors in the training data can lead to large scale errors in the system's output. Finding and analysing the relations between particular parameters facilities to draw correct conclusions and

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 July 2019 doi:10.20944/preprints201907.0319.v1

Peer-reviewed version available at Appl. Sci. 2019, 9, 3719; doi:10.3390/app9183719

enables the proper interpretation of the results [12]. Apart from this, such knowledge can be useful by selecting the appropriate machine learning model or its parameters [13].

In our work we used the Python language, which offers advanced tools for machine learning and data analysis. In order to train and evaluate selected models, authors used the following components and applications:

- Python 3.6.3
- Keras 2.2.0 — Open Source library for creating neural networks
- TensorFlow 1.8.0 — Open Source library written by Google Brain Team for linear algebra and neural networks
- Scikit-learn 0.19.1 — Open Source library implementing many different methods of machine learning

2.4. Machine Learning Algorithms

To ensure a high degree of independence between the models, we use three significantly different machine learning algorithms: SVM (Support Vector Machine) with the *rbf* kernel, ANN (Artifical Neural Network) and BDT (Bagging Decision Trees) in standard implementations. Underlying theory can be found in [14–18].

Next, we will show how to use the general method of hyperparameters optimisation to improve the obtained results. At the end, such improved models will be used to build an ensemble classifier. The purpose of this procedure is to obtain a model that will be able to best predict the failure rate of meters, while maintaining its generalisation, i.e. minimising overfitting.

For training selected models, the default parameters of the algorithms will be used as well as all the features of the observations presented in the table 1. Building and evaluation of the models were always based on the same dataset (51890 records), randomly divided into the training set (80% - 41512 records) and the testing set (20% - 10378 records).

3. Results

3.1. Statistical analysis

One of the first steps in exploratory research is usually constructing a correlation matrix [19]. Frequently, its analysis allows to eliminate the irrelevant parameters and may be a starting point for PCA analysis. Apart from that, the correlation matrix may reveal some dependencies between particular variables, their mutual relations and potential redundancies [20]. To visualise such information, we applied a thermal map. By analysing the correlations presented in the figure 2, a very weak relation of failure from all other parameters (the last row of matrix) can be observed.

The highest correlation coefficient at 0.26 has a parameter of a current meter value. Such correlation value is relatively small and most often omitted. On this basis, it can be concluded that failures do not directly depend on any single meter feature, but perhaps on some nonlinear combination of state variables.

A very strong correlation can be observed between the meter age and the number of billing (nearly 1). Presumably, one of these parameters is unnecessary (redundant) in further analysis. A little smaller, but also significant correlation can be seen between the maximum and the current energy consumption. Also, this relation is rather apparent and does not require any more explanations.

There are no rows below the current meter value, which demonstrate higher correlation with other parameters and can be treated as linearly independent. Attention is drawn only to the small correlation between the manufacturer and the type of communication, which is at 0.4. It indicates the situation, where manufacturers specialise themselves in making meters for particular communication types but also the maintenance company is supplied with particular types of meters not only by the selected manufacturers.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 July 2019 doi:10.20944/preprints201907.0319.v1

Peer-reviewed version available at Appl. Sci. 2019, 9, 3719; doi:10.3390/app9183719

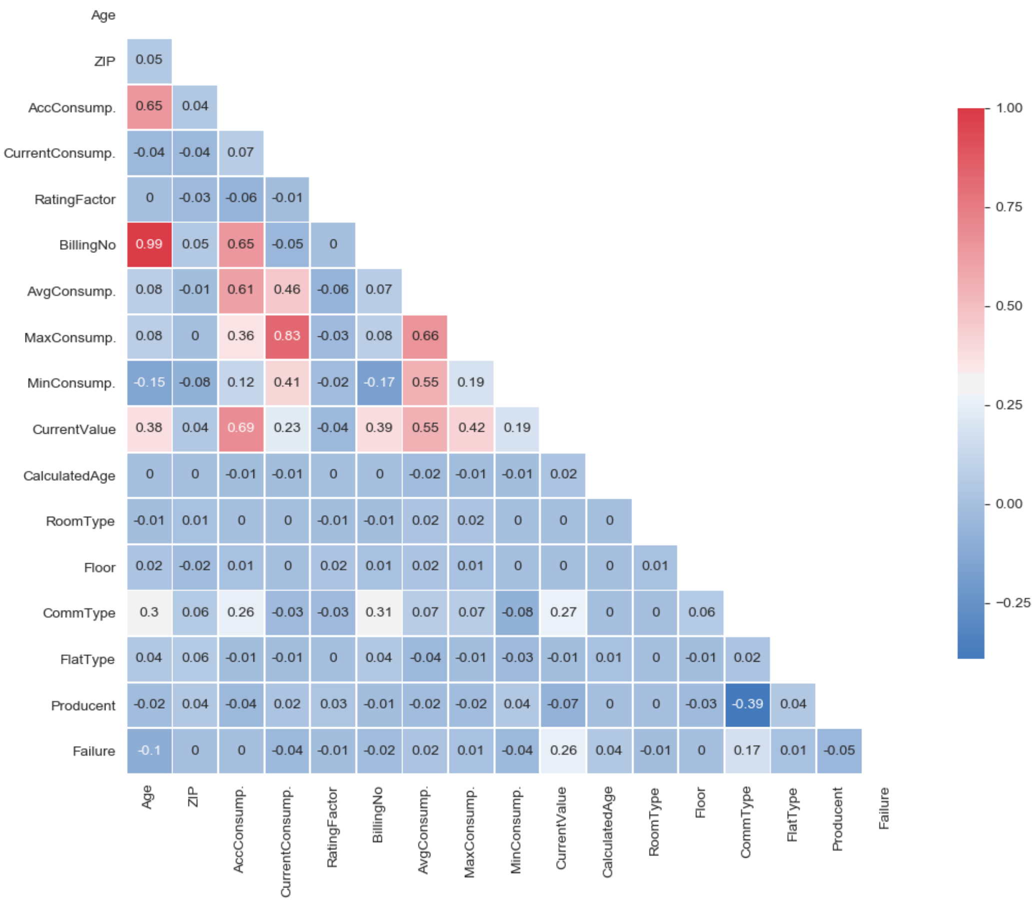

Figure 2. Correlation matrix – heat map

During the analysis of the correlation matrix, it should not be forgotten that a strong correlation does not necessarily imply a cause and effect – correlation is not causation.

3.1.1. Communication type

Because the minimisation and predictability of failures is crucial for cost optimisation, we will examine the dependence of the number of failures on other factors. The numerical dependence of failure on the meter communication type was presented in the table 2.

Table 2. Dependence of failure on the type of communication

Communication type	Meaters #	Failures #	% of Failures
Walk by	1 045	18	1.72 %
Funk	10 996	1 716	15.61 %
Bus	12 631	1 778	14.08 %
Without communication	27 577	8 273	30.00 %

The meters of "walk by" type have a very low percentage of failures, but the sample is small in relation to other types of communication, so the conclusion that they are the most reliable is a bit premature. You can see that meters without a communication module fail almost twice as often as those with communication. We can certainly conclude that the meters with remote communication – in other words the new generation of devices – are much less likely to fail.

Using the methods presented in [10], we can determine the distribution of the probability vector. On the figure 3 each state from 1 to 10 represents the age of meter in years. We can observe that,

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 July 2019 doi:10.20944/preprints201907.0319.v1

Peer-reviewed version available at Appl. Sci. 2019, 9, 3719; doi:10.3390/app9183719

also in long term, there is a significant difference in failure rate between meters with and without communication. The linearity visible on that figure confirms, that the failures' intensity is poorly dependent on the meter's operating time. Except that information about stationary distribution allows for better planning of inventory.

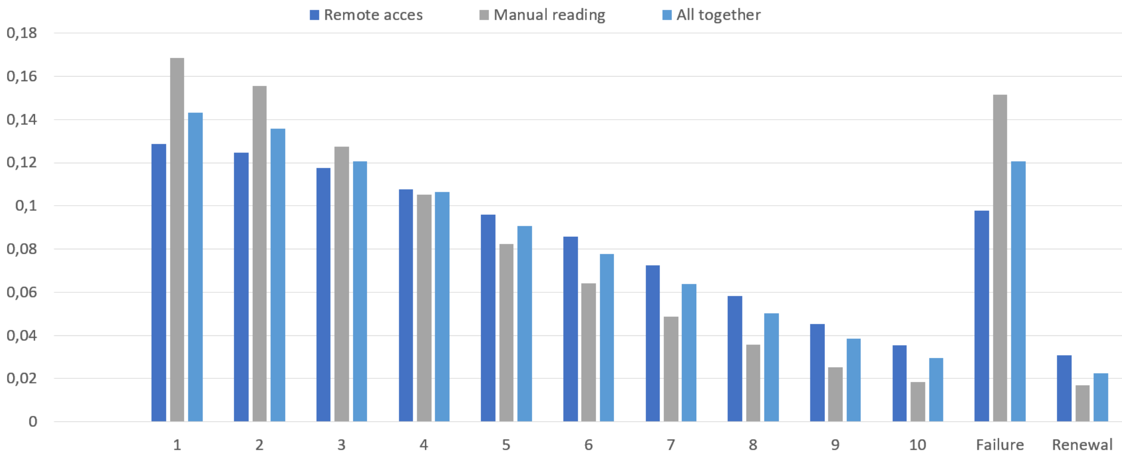

Figure 3. Stationary probability distribution

Therefore, in the context of smart meters and smart homes, devices with remote reading are definitely a better choice. Not only are more 'user friendly' (the presence of resident is not required during readings) but they are also more reliable.

3.1.2. Producer

Similarly to the grouping by the type of communication, we can present the violin charts with the division into the producer (fig. 4). The curves of Kernel Density Estimator (defined in [21]) for the number of devices from manufacturers No. 2 and 4 have their maxima above consumption of 20'000 kWh, which suggests that purchases have been abandoned and there are no new installations anymore. In the case of producer no. 10, the failures are slightly delayed in relation to the number of new meters, but it is difficult to draw any more specific conclusions here.

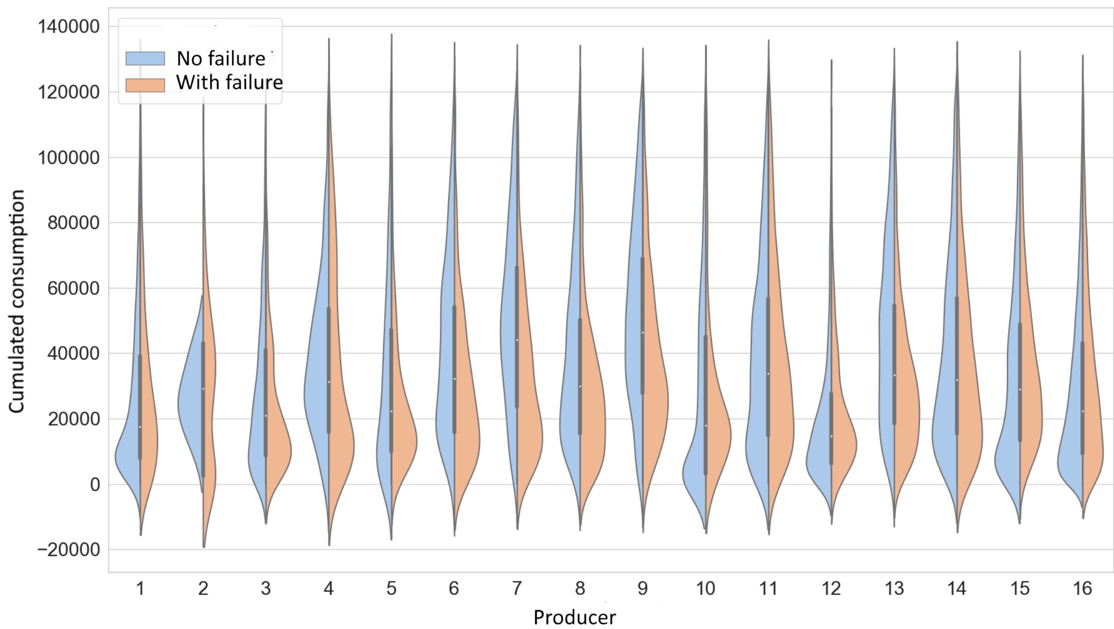

Figure 4. Consumption and failures by producer

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 July 2019 doi:10.20944/preprints201907.0319.v1

Peer-reviewed version available at Appl. Sci. 2019, 9, 3719; doi:10.3390/app9183719

The complement of the violin charts for producer is tab. 3. The failure results presented here are certainly influencing Rapp's purchasing policy. On the one hand, we have a strong leader (producer No. 16), which maintains the failure rate of its devices at a relatively low level of 13%. On the other hand, we see diversification and an attempt to become independent of this producer. The reason for the resignation from the services of the producer No. 2 was the high failure rate of over 90%, even in a few years.

Table 3. Failures by producer in numbers

Producer	Meters #	Failures #	% of failure
1	2 555	290	11.35 %
2	118	114	96.61 %
3	1 674	281	16.79 %
4	671	303	45.16 %
5	713	187	26.23 %
6	823	345	41.92 %
7	1 708	438	25.64 %
8	1 494	603	40.36 %
9	1 832	443	24.18 %
10	2 026	332	16.39 %
11	1 853	600	32.38 %
12	2 440	234	9.59 %
13	3 348	1 859	55.53 %
14	3 506	655	18.68 %
15	3 977	2 059	51.77 %
16	23 511	3 042	12.94 %

The above observation strongly supports the thesis, that statistical analysis performed on big data set can create significant savings and should be used in process of designing and maintaining smart buildings.

3.2. *Reliability analysis*

The cumulative heat consumption is a meter feature, which can be treated as an equivalent of operating time. In article [10] we showed that the distribution of probability of failures is exponential:

$$f(t) = \lambda e^{(-\lambda t)} \tag{1}$$

It is characterised by the constant intensity of failures, i.e. $\lambda = 3.08 * 10^{-5}$. It means that failures occur as external random events and does not depend on usage time-they appear randomly with the fixed intensity. Such feature is c a 'memoryless' of exponential distribution and implies that if we know that in moment x the element was fit for use, thus, counting from that moment, the fitness time has the same distribution as a distribution of a new element [22–24].

It can be stated that during these 10 years, the intensity of failures of heat meter is poorly dependent on the usage time. It is crucial since it enables the prediction of failures, which is independent of the history of operation. We do not have to possess information on how and when the equipment was used - if we know its parameters, it is entirely sufficient.

The above conclusion, as well as correlation matrix, suggest the selection of algorithms of machine learning to those, which perform well in classification problems with non-linear separable classes.

3.3. *Machine learning*

3.3.1. Metrics

There are a lot of different metrics which can be used to compare performance of trained ML models – each has its strong and weak sides (compare [25,26]). Due to the character of the data (high

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 July 2019 doi:10.20944/preprints201907.0319.v1

Peer-reviewed version available at Appl. Sci. 2019, 9, 3719; doi:10.3390/app9183719

predominance of records without failure – approximately 80%), as well as the goal of the model (equally important as predicting failure is predicting whether the meter will continue working), we focused on two metrics: *Area Under the ROC Curve* – AUC and *Matthews Correlation Coefficient* – MCC. The MCC takes values from $[-1, 1]$ (more means better) and AUC takes values from $[0, 1]$, whereby 0.5 means a random classifier. Detailed information about each of metric can be found in [27]. For the more comprehensive image we will provide also accuracy, precision, recall and f_1 for both classes.

The figure 5 shows the results of all 3 models in various metrics. The figure 6 presents a part of the confusion matrix regarding true predictions (True Positive and True Negative).

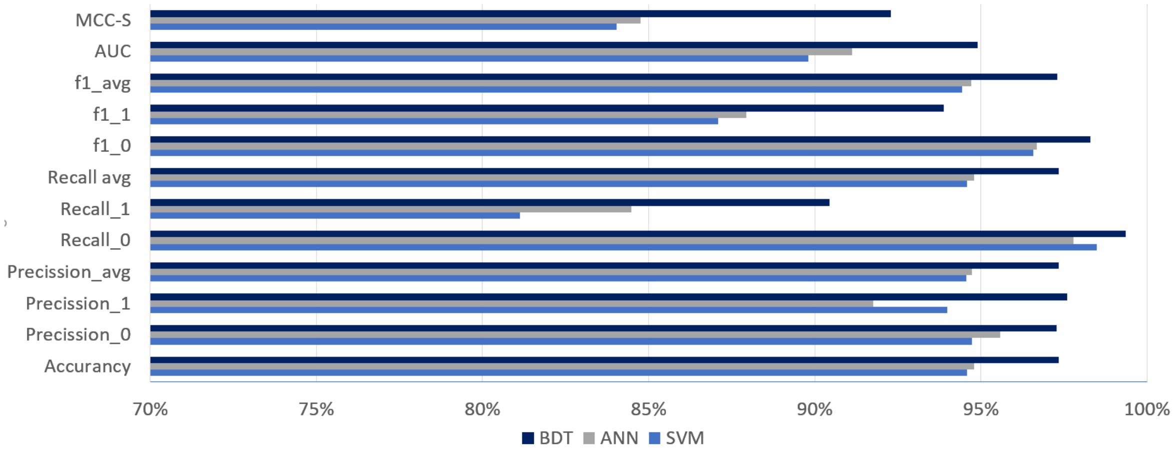

Figure 5. Average metrics for 10 tries

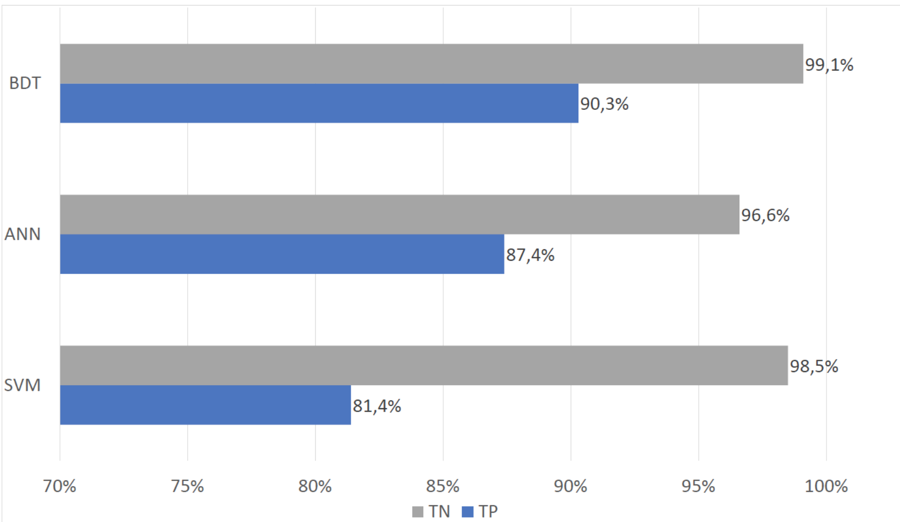

Figure 6. True Positive and True Negative

As you can see, BDT is the best and SVM is the worst. All models are much less successful in detecting meter failure than predicting survival for the next accounting period. Although we have not yet optimised the tested models, the results are very good. They mainly result from preprocessing and data normalisation, as well as proper selection of parameters.

3.3.2. Hyperparameter optimisation

Hyperparameter optimisation is a problem of finding a minimum of a certain objective function, the domain of which is the space of parameters of the examined model. The parameters can be continuous, discrete or categorical and additionally they can be dependent on each other [28]. It is

Peer-reviewed version available at Appl. Sci. 2019, 9, 3719; doi:10.3390/app9183719

worth highlighting that calculating the objective function is extremely expensive – it involves the full training and evaluation of the model.

There are different strategies of looking for optimum hyperparameters. The easiest way is 'manual' tuning. However, it requires expert knowledge of the model and the data, which does not foster generalisation. The other strategy is either full or random search of parameters' domain, the so-called 'grid search'. Checking all combinations is usually unrealistic due to the high costs. It has been confirmed that random search can work well in the case of a model with many parameters, out of which only some play a key role in its quality [29]. The next method of searching for optimum parameters of a classifier is a SMBO (Sequential Model-Based Optimisation) method. To put it simply, it consists in constructing a surrogate model approximating the objective function, the minimum of which we look for. Most frequently GP (Gaussian Process), RFR (Random Forest Regressions) or TPE (Tree-structured Parzen Estimator) are used as surrogate models. The selection of subsequent domain points (values of hyperparameters) is calculated in a way to optimise the selection function – here we most frequently use the EI function (Expected Improvement). Such a strategy usually provides the best results and eliminates the element of randomness [30].

To optimise the hyperparameters of models described in this paper the authors decided to use SMBO with TPE model. As the objective function, the AUC metrics was applied.

After optimising the hyperparameters, each of the models improved its performance – especially in failure detection. This was also the main goal of the optimisation that was achieved. It can be noticed that the larger the hyperparameters of the tested model, the easier it can be optimised. The neural network, noted the highest progress in each metric (compare 7). Significant progress is also visible in the SVM model.

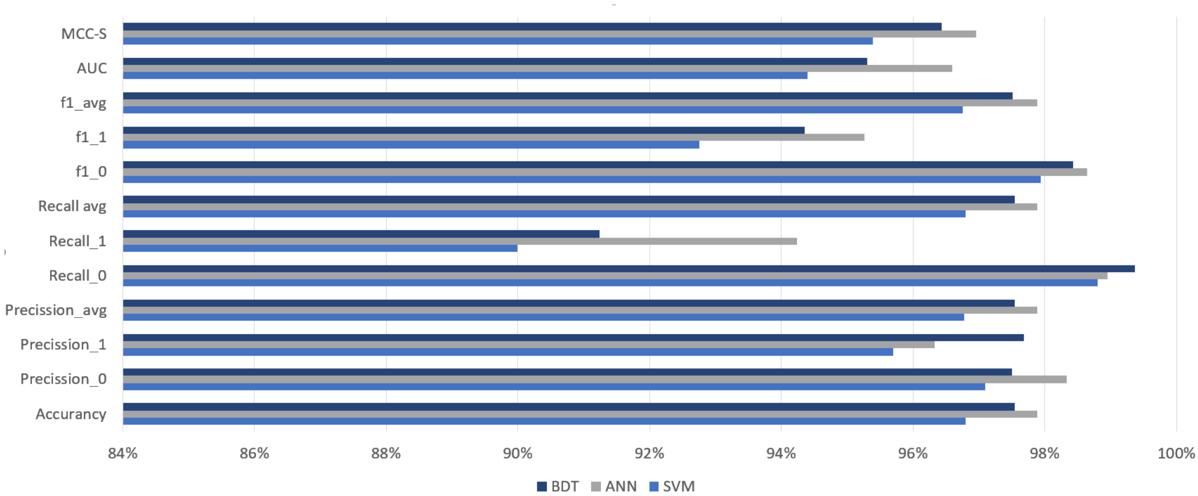

Figure 7. Models' metrics after optimisation

In general, most of the metrics show progress in the range of 3-5%, which is a very good result, especially considering that the quality of the examined models before the optimisation exceeded 92%.

3.4. Ensemble model

The ensemble classifier, as in the case of random forest (BDT) is a meta-classifier that internally uses several, preferably strongly independent models, respectively aggregating their predictions, generates the result [31]. The assumption about the differentiation of models is important because the similar classifiers make the same mistakes, so the final group model would only duplicate them. Sufficient differentiation can be achieved for the same algorithms by appropriate division of the training set into subsets (*bagging*). It also happens that subsets of data attributes are used instead of data subsets. In this paper, the authors decided on a different approach, namely the use of already

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 July 2019 doi:10.20944/preprints201907.0319.v1

Peer-reviewed version available at *Appl. Sci.* **2019**, *9*, 3719; doi:10.3390/app9183719

trained (and optimised) models and building of a collective classifier using the most popular voting algorithms with weights. The results of the obtained ensemble classifier can be seen in the table 4.

Metric	Result
$Accuracy$	98.15%
$Precision_0$	98.08%
$Precision_1$	98.42%
$Precision_{avg}$	98.15%
$Recall_0$	99.56%
$Recall_1$	93.34%
$Recall_{avg}$	98.32%
$f1_0$	98.82%
$f1_1$	95.81%
$f1_{avg}$	98.22%
AUC	96.45%
$MCC\text{-}S$	97.34%

Table 4. Ensemble classifier – average for 10 tries

For the most important metrics (MCC and AUC), the ensemble classifier is better than the average of the optimised models by more than one percent, which gives respectively 22% and 29% of the total possible improvement. The increase in detectability of failures by more than 1.5% for the $Precission_1$, $Recall_1$ and $f1_1$ metrics is also important.

4. Discussion

Access to large data sets containing information about the operation of already installed devices in residential buildings is nowadays relatively easy. The analysis of these data provides significantly new knowledge about their exploitation and can be treated as an intermediate step in the implementation of smart meters or smart homes. Using the reliability analysis tools, we have discovered that the failure intensity of the heat meters is random and weakly dependent on the time of their use. In addition to statistical analysis, the appropriate use of machine learning algorithms, including neural networks and ensemble models, allows a better understanding of the operation of such devices. Next to the failure prediction model presented in this work, you can also try to predict water usage, heat consumption or periods when such consumption will be minimal. Such knowledge can be used to optimise the costs of media transmission as well as to reduce costs for individual users (compare [32]). Similar approach has been shown in [33], but authors the used an adaptive neuro-fuzzy network.

The knowledge needed for data analysis and model building is available to researchers and fairly well-established. However, studies that use hyperparameter optimisation are not common, which encouraged us to present this issue a bit more widely. We also showed that this is an important step to improve the efficiency of machine learning models.

It should be emphasised that the preparation of data for analysis is a complex and laborious process, but extremely important for the quality of future models. As stated in [12], many ML methods are very sensitive to the type and quality of the data. To increase the quality of predictions, you can build an ensemble model with help of several (preferably independent) models. As a rule, its results are better than its individual components. However, the choice of algorithms in that case can not be accidental. In the presented example, we showed, on the basis of the correlation matrix, that selected parameters of heat meters are linearly independent. This means that, for example, the logistic regression method would be a weak choice.

5. Conclusions

It is not possible to replace all heat meters currently being in use at once. Therefore in order to realise the idea of smart homes, we need to get known the devices much better. We have shown that both statistical analysis and prediction models provide a significant new knowledge on the operation and failures of heat meters. The reliability of heat meters (especially those with remote communication) seems to be good enough for smart city applications and optimisation of heat consumption.

Currently, most applications of machine learning are primarily based on neural networks. It is understandable that this method is very flexible and probably the only one suitable for solving many types of problems such as regression, classification, clustering, reinforcement. Consequently, other types of algorithms are rarely considered. We showed that prior exploratory data analysis, the right choice of parameters, hyperparameters optimisation and the construction of the ensemble classifier can significantly enhance the quality of predictions and create a solution which outperforms the results of an individual neural network.

Presented data analysis and results of created models lead to following findings and implications:

- The intensity of failures of heat meters is almost independent on the operating time.
- Due to the high reliability of the meters and their considerable cost, condition-based or predictive maintenance of heat meters is justified and possible. It was shown, that required 5 years verification period could be extended.
- Collecting different parameters about devices in use, makes sense even if they are not required at the moment. It allows to build more general and better ML models in future.
- Heat meter data allow to build a ranking list of their producers and optimise deliveries.
- Heat meters with remote communication are twice as reliable as ones with manual reading.

This article is one of the few which deal with reliability and predictability of heat meters' failures. It is also, according to our knowledge, the first attempt to use more independent ML models based on a single database. Achieving the result above 95% for the AUC metric by the model, while maintaining overfitting at the minimum level, is a remarkable outcome.

It is not certain whether the developed models achieve equally good efficiency for meters and data derived from other sources. Due to the fact that training data supplied by only one meters' operator, the models can be biased. However, the presented approach and methodology of model construction shall perform well independent of data sources. The methods applied by us are so universal that they can be utilised to study the reliability and predict failures of other types of meters, e.g. water meters or heat cost allocators.

The developed Machine Learning model and the acquired knowledge will be used in the design of new heating systems as well as for optimisation of stocks and maintenance actions in existing heat meter networks, which will bring significant benefits for their operators, tenants and the environment. In particular, the optimisation of meters' maintenance in large buildings will allow companies to save both time and resources. Such optimisation is also crucial for the tenants, who are the end users of heat meters. It does not only shorten the time needed for their presence during replacement, but it also guarantees accurate meter readings and the fair distribution of heating costs.

Supplementary Materials: Due to the sensitivity and ownership of the data, datasets used will not be publicly available. Under special circumstances the access can be granted by Rapp Enserv AG. In that case please contact Rapp Enserv, Hochstrasse 100, 4018 Basel, +415859577744.

Author Contributions: P. Pałasz and R. Przysowa conceived and designed the research; P.P. processed and analysed the data; R.P. verified and evaluated the results. P.P. and R.P. drew conclusions and produced the paper.

Funding: This research received no external funding

Acknowledgments: We are very grateful to Rapp Enserv AG for allowing us to use and analyse the data. We would like to thank Prof. Józef Żurek from Instytut Techniczny Wojsk Lotniczych for sharing his expertise with us during the course of this research. We are also much obligated to Krzysztof Przysowa for his comments on an earlier version of the manuscript, although any errors are our own and should not tarnish the reputations of these esteemed persons. Special thanks should be given to Magda Tomaszewska for English language editing.

Conflicts of Interest: The authors declare no conflict of interest. Rapp Enserv AG did not sponsor this research and had no role in the design, execution, interpretation, or writing the study. The views, information, or opinions expressed herein are solely those of the authors and do not necessarily represent the position of any organisation.

Abbreviations

Peer-reviewed version available at Appl. Sci. 2019, 9, 3719; doi:10.3390/app9183719

The following abbreviations are used in this manuscript:

ANN: Artificial Neural Network
AUC: Area Under the ROC Curve
BDT: Bagging Decision Tree
LDA: Linear Discriminant Analysis
MCC: Matthews Correlation Coefficient
MDPI: Multidisciplinary Digital Publishing Institute
ML: Machine Learning
PCA: Principal Component Analysis
rbf: Radial Basis Function
ROC: Receiver Operating Characteristic
SMBO: Sequential Model-Based Optimisation
SVM: Support Vector Machine
TPE: Tree-structured Parzen Estimator

1. European Commission. Energy Efficiency, 2019.
2. Makonin, S.; Wang, Z.J.; Tumpach, C. RAE: The rainforest automation energy dataset for smart grid meter data analysis. Data **2018**, 3, 8. doi:10.3390/data3010008.
3. Zhang, G.; Wang, G.G.; Farhangi, H.; Palizban, A. Data mining of smart meters for load category based disaggregation of residential power consumption. Sustainable Energy, Grids and Networks **2017**, 10, 92–103. doi:10.1016/j.segan.2017.03.006.
4. Khairalla, M.; Abdrabou, A.; Dahir, A.; Gaouda, A. Fault location based on smart meters time synchronized measurements. Smart Grid and Renewable Energy (SGRE), 2015 First Workshop on. IEEE, 2015, pp. 1–6. doi:10.1109/sgre.2015.7208720.
5. Lee, J.Y.; Yeo, M.S.; Kim, K.W. Predictive control of the radiant floor heating system in apartment buildings. Journal of Asian Architecture and Building Engineering **2002**, 1, 105–112.
6. Javed, A.; Larijani, H.; Ahmadinia, A.; Emmanuel, R. Comparison of the robustness of RNN, MPC and ANN controller for residential heating system. 2014 IEEE Fourth International Conference on Big Data and Cloud Computing. IEEE, 2014, pp. 604–611.
7. Yang, I.H.; Yeo, M.S.; Kim, K.W. Application of artificial neural network to predict the optimal start time for heating system in building. Energy Conversion and Management **2003**, 44, 2791–2809.
8. Keçebaş, A.; Alkan, M.A.; Yabanova, İ.; Yumurtacı, M. Energetic and economic evaluations of geothermal district heating systems by using ANN. Energy policy **2013**, 56, 558–567.
9. Yabanova, İ.; Keçebaş, A. Development of ANN model for geothermal district heating system and a novel PID-based control strategy. Applied Thermal Engineering **2013**, 51, 908–916.
10. Pałasz, P.; Żurek, J.; Przysowa, R. Markov reliability model for heat meters. Journal of KONBiN **2018**, 45, 83–96. doi:10.2478/jok-2018-0005.
11. Witten, I.H.; Frank, E.; Hall, M.A.; Pal, C.J. Data Mining: Practical machine learning tools and techniques; Morgan Kaufmann, 2016; chapter 4.
12. Gudivada, V.; Apon, A.; Ding, J. Data quality considerations for big data and machine learning: Going beyond data cleaning and transformations. International Journal on Advances in Software **2017**, 10, 1–20.
13. Hall, M.A. Correlation-based feature selection of discrete and numeric class machine learning. Proceedings of the Seventeenth International Conference on Machine Learning, 2000, pp. 359–366.
14. Goodfellow, I.; Bengio, Y.; Courville, A.; Bengio, Y. Deep Learning; Vol. 1, MIT press Cambridge, 2016.
15. Pérez, R.; Vásquez, C. Fault location in distribution systems with distributed generation using support vector machines and smart meters. Ecuador Technical Chapters Meeting (ETCM), IEEE. IEEE, 2016, pp. 1–6. doi:10.1109/etcm.2016.7750857.
16. Quinlan, J.R.; others. Bagging, boosting, and C4.5. AAAI/IAAI, Vol. 1, 1996, pp. 725–730.
17. Sapna, S.; Tamilarasi, A.; Kumar, M.P.; others. Backpropagation learning algorithm based on Levenberg Marquardt Algorithm. Comp Sci Inform Technol (CS and IT) **2012**, 2, 393–398.

18. Zhang, D.; Zhou, X.; Leung, S.C.; Zheng, J. Vertical bagging decision trees model for credit scoring. Expert Systems with Applications **2010**, 37, 7838–7843. doi:10.1016/j.eswa.2010.04.054.
19. Johnson, R.A.; Wichern, D.W. Applied Multivariate Statistical Analysis, 6 ed.; Prentice Hall, 2007.
20. Biały, J.; Rurak, A.; Tomaszewska, J.; Toruń, A.; Woch, M.; Zieja, M.; Żurek, J. Statistical Analysis of the Volume of Passenger Air Traffic in Europe in years 2004-2015. Transportation research procedia **2018**, 35, 72–79.
21. Martinez, W.L.; Martinez, A.R.; Solka, J. Exploratory data analysis with MATLAB; Chapman and Hall/CRC, 2017.
22. Bradley, E. Reliability Engineering A Life Cycle Approach; CRC Press Taylor & Francis Group, 2017.
23. He Ren, X.C.; Chen, Y. Reliability Based Aircraft Maintenance Optimization and Application; Aerospace Engineering, Academic Press Elsevier, 2017.
24. Rykov, V. Reliability of Engineering Systems and Technological Risks; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2016. doi:10.1002/9781119347194.
25. Murphy, K.P. Machine Learning: A Probabilistic Perspective; Number 12 in Adaptive computation and machine learning series, MIT Press: London, England, 2013.
26. Domingos, P. The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World; Basic Books, 2015; p. 354.
27. Davis, J.; Goadrich, M. The relationship between Precision-Recall and ROC curves. Proceedings of the 23rd international conference on Machine learning. ACM, 2006, pp. 233–240. doi:10.1145/1143844.1143874.
28. Bergstra, J.; Yamins, D.; Cox, D.D. Making a science of model search: Hyperparameter optimization in hundreds of dimensions for vision architectures. ICML **2013**.
29. Bergstra, J.; Bengio, Y. Random search for hyper-parameter optimization. Journal of Machine Learning Research **2012**, 13, 281–305.
30. Snoek, J.; Larochelle, H.; Adams, R.P. Practical bayesian optimization of machine learning algorithms. Advances in neural information processing systems, 2012, pp. 2951–2959.
31. Zhang, C.; Ma, Y. Ensemble machine learning: methods and applications; Springer, 2012.
32. Sajjadi, S.; Shamshirband, S.; Alizamir, M.; Yee, L.; Mansor, Z.; Manaf, A.A.; Altameem, T.A.; Mostafaeipour, A. Extreme learning machine for prediction of heat load in district heating systems. Energy and Buildings **2016**, 122, 222–227.
33. Shamshirband, S.; Petković, D.; Enayatifar, R.; Abdullah, A.H.; Marković, D.; Lee, M.; Ahmad, R. Heat load prediction in district heating systems with adaptive neuro-fuzzy method. Renewable and Sustainable Energy Reviews **2015**, 48, 760–767.