ECWS

MDPI

Implications of Experiment Set-Ups for Residential Water End-Use Classification

Nora Gourmelon [1,2,*], Siming Bayer [1,3,*], Michael Mayle [2], Guy Bach [4], Christian Bebber [4], Christophe Munck [4], Christoph Sosna [3] and Andreas Maier [1]

[1] Pattern Recognition Lab, Friedrich-Alexander University, Martenstraße 3, 91058 Erlangen, Germany
[2] Diehl Metering GmbH, Industriestraße 13, 91522 Ansbach, Germany
[3] Diehl Metering GmbH, Donaustraße 120, 90451 Nuremberg, Germany
[4] Diehl Metering SAS, 67 Rue du Rhone, 68300 Saint-Louis, France
* Correspondence: nora.gourmelon@fau.de (N.G.), siming.bayer@fau.de (S.B.); Tel.: +49-157-377-69375 (N.G.), +49-176-225-20695 (S.B.)

Abstract: With the increased population in urban areas worldwide, the security of water supply is gaining in importance. Water scarcity accelerated by climate change poses additional stress to water supply infrastructures. Water consumption data transmitted by smart water meters form the foundation of advanced data analysis, such as water end-use classification, with which the resilience of water supply can be improved. Especially with large amounts of high-resolution data, the accurate categorization of data from smart water meters into different end-uses such as toilets, showers or dishwashers is challenging and cannot be performed by humans. To this end, machine-learning (ML) approaches provide several benefits, such as real-time capability, scalability and generalizability. State-of-the-art methods to identify residential water end-uses include both unsupervised methods and supervised approaches. However, a comprehensive comparison of unsupervised and supervised techniques is still missing. In this study, we are aiming at a quantitative evaluation of various ML techniques for water end-use classification. Furthermore, we focus on deriving general implications on the setting and conduction of ML-based experiments for water end-use classification. For these purposes, a stochastic water consumption simulation tool with high capability to model the real-world water consumption pattern is applied to generate residential data. Subsequently, unsupervised clustering methods, such as dynamic time warping, k-means, DBSCAN, OPTICS and Hough transform, are compared to supervised methods based on SVM. The quantitative results demonstrate that supervised approaches are capable to classify common residential end-uses (toilet, shower, faucet, dishwasher, washing machine, bathtub and mixed water-uses) with accuracies up to 0.99, whereas unsupervised methods fail to detect those consumption categories. The major implications drawn from the quantitative results are two-fold: clustering techniques alone are not suitable to separate end-use categories fully automatically. Hence, accurate labels are essential for the end-use classification of water events, where crowdsourcing and citizen science approaches pose feasible solutions for this purpose.

Keywords: end-use classification; smart water meter; machine learning

1. Introduction

In recent years, climate change and population growth have exacerbated water scarcity and accelerated the vulnerability of water supply systems. A comprehensive and detailed understanding of water demand as well as water end-uses will improve the resilience of these systems. With the water consumption data generated and transmitted by smart water meters, advanced analysis of water demand using data-driven approaches is facilitated. In particular, the categorization of water use events into residential end-uses like washing machine, faucet or toilet permits a detailed investigation of the impact of different end-uses as well as the inhabitants on the overall demand. With the recent advances in artificial intelligence, categorization of residential water end-use can be

resolved as classification problem using Machine Learning (ML) techniques, thus having inherent real-time capability, scalability and generalizability. On a basic level, ML techniques can be subdivided into supervised and unsupervised approaches. The former directly perform a categorization/classification of data with given categories. Consequently, training data with labels annotating the samples' categories are necessary. The latter, search for groups of similar data points in the dataset in the absence of labels.

Both unsupervised and supervised approaches are widely applied in studies on end-use classifications of water use events. For example, Pastor-Jabaloyes et al. [1] proposed an unsupervised approach formulating a Partition Around Medoids clustering method using Gower distance as a similarity measure in the feature space, which is spanned by the volume and the average flow rate of the water use events. A clustering algorithm based on Dynamic Time Warping (DTW) as the similarity measure between events is described by Nguyen et al. [2]. This algorithm was refined to a hybrid method comprising k-medoids clustering, DTW and an swarm-intelligence-based global optimization, i.e. Artificial Bee Colony algorithm. Later, this approach is extended with a hybrid method consisting of Self-Organizing Maps and a k-means algorithm in the subsequent publications [3, 4]. Most supervised approaches are based on Hidden Markov Models and a subsequent optimization [5-16], but Artificial Neural Networks [9-13], Decision Trees [14] and Multi-category Robust Linear Programming [17] are used as well. Support Vector Machines (SVMs) were employed by Vitter et al. [18] to categorize water events using water consumption data and coincident electricity data. Furthermore, Carranza et al. [19] identify residential water end-uses in data measured by precision water meters equipped with pulse emitters using SVMs.

Both supervised and unsupervised approaches proposed in the literature demonstrate high performance in terms of accuracy on a particular dataset described in the corresponding publication, however, a comprehensive comparison of both approaches on a common database has not been performed yet.

In this paper, we focus on a comparative evaluation of supervised and unsupervised ML techniques for the application of residential water end-use classification, and aim at deriving a decision support for the selection of the appropriate approach, revealing possible pitfalls of water end-use classification. For this purpose, we implemented a large variety of ML-based approaches proposed in the state-of-the-art literature, i.e. the clustering algorithm based on DTW established by Nguyen et al. [2], k-means, Density-Based Spatial Clustering of Applications with Noise (DBSCAN), Ordering Points To Identify the Clustering Structure (OPTICS), Clustering in Arbitrary Subspaces based on the Hough transform (CASH), as well as two classifiers based on Support Vector Machines (SVM). Subsequently, the methods are evaluated on a common database generated using a stochastic water consumption simulation framework. Finally, evaluation results are analyzed and discussed in detail, with the intention to draw implications and recommendations of experiement set-ups for data-driven water end-use classification in general.

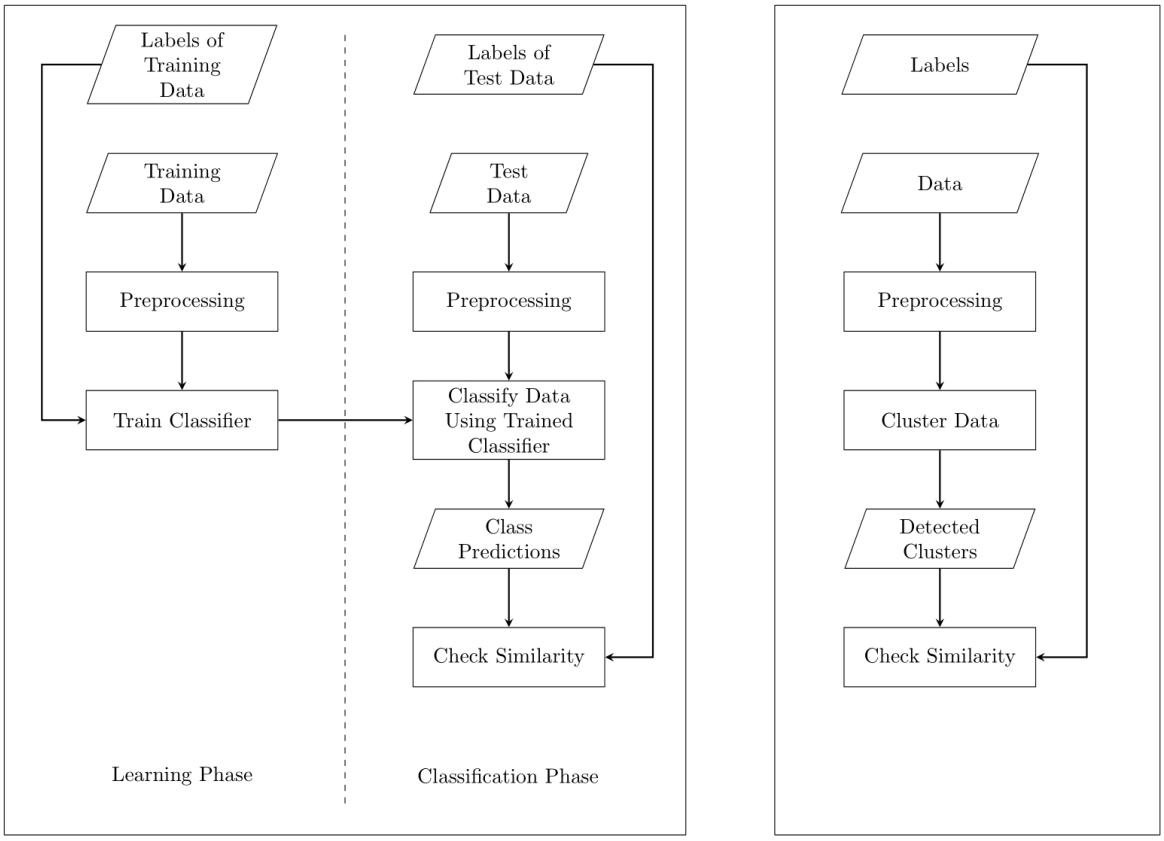
2. Methods

In this section, the simulated common water consumption database as well as the ML techniques employed and evaluated in this work are descriped in detail. An overview of supervised and unsupervised ML methods is presented in Figure 1. Both labels and water consumption data are generated with a stochastic simulation proposed in [16].

2.1. Simulation of Common Database and Data Preprocessing

The datasets employed in this paper are generated using the STochastic Residential water End-use Model (STREaM) developed by Cominola et al. [16]. It is a stochastic water consumption simulation tool with a high capability to model the real-world water consumption patterns of the end-uses, such as toilet, shower, faucet, dishwasher, washing machine and bathtub. STREaM is calibrated using observed and disaggregated water consumption data from 300 single-family houses in nine U.S. cities. With a stochastic model based on monte-carlo simulation, time series with different temporal resolutions (up to 10 seconds) and number of inhabitants can be emulated. To the best of

our knowledge, STREaM is the most advanced framework for realistic water comsumption simulation to date.



(a) Supervised Techniques (b) Unsupervised Techniques

Figure 1. Comparison of supervised and unsupervised learning: (**a**) Supervised techniques have a learning phase, in which the classifier trains on given labels. In the classification phase, the classifier is tested on a part of the dataset, which is not used in the training phase; (**b**) Unsupervised techniques search for grouping structures in the complete dataset.

In this work, the chosen temporal resolution is ten seconds, which corresponds to the highest possiple resolution of the simulation tool. Six datasets are generated with a varying number of inhabitants from one to six. In addition to the resolution and the number of inhabitants, the efficiency of the present end-uses and the time horizon of the datasets are fixed to "standard fixture"(i.e. fixtures which are not specified to improve the water use efficiency) and one year, respectively.

The output of the simulation includes six time series corresponding to the end-use toilet, shower, faucet, washing machine, dishwasher and bathtub. An additional time series aggregating the six above mentioned end-uses is also generated. The entire aggregated time series is segmented into water use events (or sequences) with the assumption, that a contiguous water consumption event does not have any interruption which is longer than 150 seconds. Since the timestamps of each event are known, labels for events extracted from the aggregated time series are computed by searching for water consumption at these timestamps in the time series representing the single end-uses. The result of the preprocessing is a set of water use event time series with corresponding end-use labels. Events with a consumption of less or equal to one liter as well as events with more than 1000 liter are removed from this set, as they indicate leaks and might not be connected to an end-use of interest.

2.2. Feature Description and Feature Extraction

Prior to the application of classification or clustering techniques, relevent features describing the water events are commonly selected from the simulated water consumption data, except for a DTW

based algorithm described in Sec. 2.4.1. Features selected in this work include volume, duration, maximum, minimum and most common flow rate, as well as the corresponding numbers of occurrences of these flow rates, the number of different flow rates and two orthogonal harmonic functions representing the time of the day. Since these features have different scales, we performed z-normalization in order to prevent the effect that features with a larger scale will dominate the other. Subsequently, we compute the pair-wise Pearson Correlation Coeeficient (PCC) of the features, and exclude one of the two features when the PCC is higher than 0.9. In order to deal with the curse of dimensionality, Principal Component Analysis is employed as dimensionality reduction technique prior to all algorithms except for CASH (refer to Sec. 2.4.5).

2.3. Supervised Techniques

In the category supervised learning techniques, we implemented two classifiers, namely a multi-class SVM as well as binary SVMs per end-use. Generally, SVMs separate classes with a maximal margin hyperplane in the feature space. The regularization parameters for the SVMs are set to 1.0 and Radial Basis Functions are utilized as the kernel. Basically, a binary SVM is trained to distinguish a single end-use class from all other end-uses. Therefore, seven binary SVMs are trained to each identify one end-use: Toilet, shower, faucet, dishwasher, washing machine, bathtub and overlapping events (i.e. events with more than one active end-use). In contrast a multi-class SVM is able to classify all seven end-uses at once. This is achieved by using a so-called one-versus-one approach. For every possible pair of classes, one SVM is trained to distinguish the two classes. The final multi-class decision is a majority vote of the 21 SVMs.

2.4. Unsupervised Techniques

The unsupervised methods evaluated in this paper include the clustering algorithm based on DTW established by Nguyen et al. [2], k-means, DBSCAN, OPTICS and CASH. Each of them is described in detail in the following sections.

2.4.1. Threshold-based Clustering using Dynamic Time Warping

The algorithm presented in [2] is based on the idea that the similariy of two events (or sequences) of a time series can be approximated by their distances to a reference event. This reference event is chosen arbitrarily from the dataset as described in [8] and will not be clustered. First, DTW distances of all water use events in the dataset to the reference event are calculated. Subsequently, relative distances are computed for each sequence. A threshold α is introduced to assign the event to a cluster. For algorithmic details we refer to the original work published in [2]. In this study, the threshold α is manually fine tuned to 0.6.

2.4.2. K-means

K-means is one of the most frequently used clustering algorithms, which use euclidean distances in the feature space (refer to 2.2) to determine cluster memberships. The optimal hyperparameter k, i.e. the number of clusters, is computed with an Elbow method using inertia. The output of this analysis, namely k=10 is fixed for all experiments conducted.

2.4.3. DBSCAN

In contrast to k-means, DBSCAN [20] is a density-based clustering algorithm. It uses the concept of core, border and noise points to form partitions in the feature space. Two hyperparameters, *eps* and *minPts*, are required to distinguish these three types of points. An neighborhood with a radius of *eps* around each feature vector is selected and the number of feature vectors in its neighborhood is counted. If a vector has more than *minPts* neighbors, then this vector is considered to be a core point. If a vector has less than *minPts* neighbors but one core point in its neighborhood, then this vector is a border point. All remaining vectors are noise points. The superposition of core and border points forms a clusters. Unlike k-means, DBSCAN determines the number of clusters intrinsically based on

the distribution of data points in the feature space. *minPts* and *eps* are set to 14 and 2, respectively, where the former is calculated with the rule of thumb minPts = 2 · number of dimensions after dimensionality reduction [21], and the latter is estimated with a 14-Nearest Neighbor Distance Graph.

2.4.4. OPTICS

OPTICS [22] is an extended version of DBSCAN, hence also applies the notion of core, border and noise points. However, instead of looking at a fixed neighborhood area, an infinite number of distance parameters eps_i is applied, which are smaller than a "generating distance" eps_{max} ($0 \leq eps_i \leq eps_{max}$). This allows OPTICS to find clusters of different densities in the feature space. eps_{max} can simply be set to infinity, as this will identify clusters across all scales [23]. In the original paper [22], experiments indicated that values between 10 and 20 for *minPts* will always lead to promising results. Hence, *minPts* is chosen to be 15 in this work.

2.4.5. CASH

CASH [24] is a clustering algorithm specifically designed for high-dimensional data. It uses a Hough transform to bypass the curse of dimensionalty. The feature vectors are transformed from the feature space to a parameter space, which corresponds to the space of all possible subspaces of the feature space. Vectors in the feature space are equivalent to sinusoidal curves in the parameter space. Instead of looking at a spatial proximity in the feature space to form clusters, intersections of curves in the parameter space are determined. An intersection in the parameter space means that the two vectors in the feature space lie on a common hyperplane, therefore are correlated. Since CASH does not rely on any distance-based measures in the feature space, it is not affected by curse of dimensionality. Consequently, dimensionality reduction techniques are dispendable. The three hyperparameters, namely minimum number of points in a cluster, the maximal allowed deviation and the maximum number of successive splits are set to 20, 0.1 and 3, respectively.

3. Results and Discussion

The datasets described in section 2.1 are addressed as *1P*, *2P*, *3P*, *4P*, *5P* and *6P* referring to the number of inhabitants in the simulated household. The following sections provide the evaluation of the supervised and unsupervised techniques using these six datasets.

3.1. Evaluation of the Supervised Techniques

For a fair comparison, we calculated accuracy and precision both for multi-class and binary SVMs. In detail, precision, recall and f1-score for all classified end-uses are first estimated for the multi-class SVMs. The accuracy is computed in the subsequent step. For binary SVMs, confusion matrices are used to compute accuracies and precisions of the classifiers, as the number of positive and negative samples is highly unbalanced. The split of training and test data is 0.8 to 0.2 for all experiments conducted.

3.1.1. Evaluation Results

Figure. 2 presents the quantitative comparison of the accuracies for multi-class and binary SVMs. The evaluation results for the multi-class SVMs are detailed in tables A1 to A6. The precisions, recalls and f1-scores for the *dishwasher* class have values close to zero. For the classes *washing machine*, *bathtub*, as well as *overlapping* events, the corresponding values are slightly higher than 0.5. Lastly, the precisions, recalls and f1-scores for the *toilet*, *shower* and *faucet* classes have a range between 0.73 and 0.96. The accuracies/micro averages of the classifiers vary between 0.78 and 0.82 as depicted in Figure 2.

Confusion matrices depicted in Table B1 demonstrate the performance of the binary SVMs. For the end-uses *shower*, *washing machine*, *dishwasher* and *bathtub*, the binary SVMs have a high accuracy. The accuracy of binary SVMs for end-use events *toilet* and *faucet* are relatively low, but still outperforms the multi-class SVM. Compared to *bathtub* and *shower*, the precisions for *toilet*, *faucet* and

washing machine are significantly lower (refer to the confusion matrix in B1). In case of *dishwasher*, the precision is not computable, since there is no true positive event recognized.

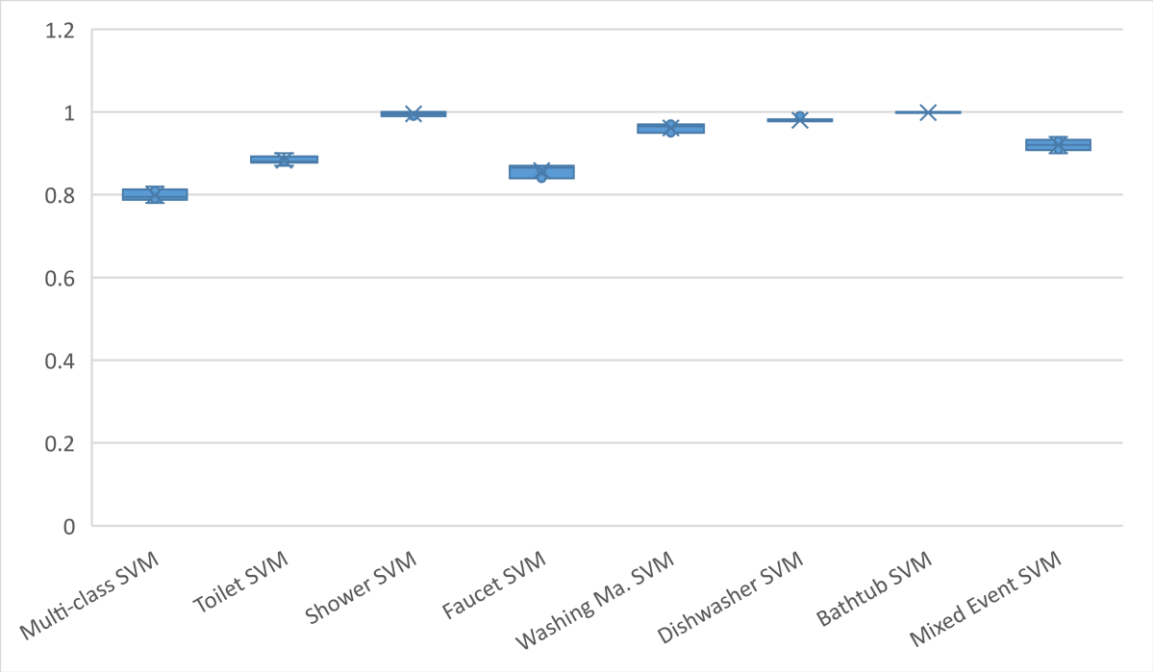


Figure 2. Accuracy ranges for multi-class and binary SVMs.

3.1.2. Discussion

At first sight, binary SVMs seem to outperform the multi-class SVM (refer to Figure 2). In case of *shower* and *bathtub*, binary SVMs demonstrate high performance in terms of accuracy and precision. This is the consequence of the characteristics of both events, i.e. both of them are characterized by a sharp ascent of flow rate, where the high flow rate is continuous for several minutes, mainly without any interruption.

However, the conclusion that binary SVMs outperform the multi-class SVM has to be made with care, as the inbalance in the sample distribution shifts the evaluation metrics for binary SVMs to higher values. For instance, the binary SVM is not able to recognize any *dishwasher* events. Although the corresponding accuracy is 100%, this classifier still fails to identify the *dishwasher* events correctly. Similar results can be observed for *washing machine*, where the binary SVM has a high accuracy but a relatively low precision. These results indicate that the classification framework described in this study comprising data preprocessing, feature extraction and binary SVMs are not ideal for water end-uses, where a single water consumption pattern is not representative for the end-use event. More concretely, an end-use event of *dishwasher* or *washing machine* usually continuous for more than an hour. However, the water consumptions assembling these end-uses are usually several separated consumptions patterns with large temporal intervals of interruption. Since the time series sequences are considered as coherent when the temporal interval of interruption is less than 150 seconds (refer to Section 2.1), the end-uses *dishwasher* or *washing machine* are not described by the selected features precisely.

For the end-uses *toilet* and *faucet*, binary SVMs tend to have high false positive and false negative rates. It indicates, that these two events are often confused with other end-uses. Both of them are abrupt and short events corresponds to a sharp curve of consumption volume and flow rate. Thus, additional features with more discriminative power should be utilized to increase the performance of the classification.

3.2. Evaluation of the Unsupervised Techniques

In contrast to the evaluation of supervised techniques, accuracy and precision cannot be computed for the evaluation of unsupervised techniques directly, as the clusters recognized by clustering methods are not necessarily equivalent to the end-use classes. For instance, we identified 10 clusters by using the k-means method, however there are only seven end-uses in the simulated database. Furthermore, the identified clusters are not assigned to any end-use classes inherently, hence the list of clusters will be a permutation of the list of end-uses. For these reasons, the Adjusted Rand Index (ARI) and the Adjusted Mutual Information (AMI), which are commonly used to assess whether detected clusters correspond to ground truth classes, are utilized as evaluation metrics.

Moreover, for CASH, DBSCAN, OPTICS and the clustering algorithm based on DTW, the number of detected clusters is presented. For k-means, the number of clusters is a hyperparameter fixed to ten (refer to Section 2.3.2) in this study.

3.2.1. Evaluation Results

The evaluation results for the unsupervised techniques described from Section 2.4.1 to 2.4.5 are summarized in Table 2 to 6, respectively. Obviously, the number of the estimated clusters differs from the actual number of end-uses greatly. With regard to the similarity between the estimated clusters and the real end-use classes (refer to ARI and AMI values), the performance of the clustering methods can be ordered as k-means, DTW-based method, DBSCAN, CASH and OPTICS, decendingly. All clustering methods demonstrate low ARI and AMI values for all experiments conducted.

3.2.2. Discussion

The evaluation results for all clustering algorithms show that the identified clusters do not correspond to the ground truth end-use classes. One reason why the unsupervised techniques are not able to estimate the end-uses in the given datasets is, that clustering methods generally take the most significant differences in a dataset into account. This might not be the end-uses, but the variation of consumer behaviours in the household or other effects which are not related the end-use classification. Moreover, single end-uses might possess a variety of different consumption patterns. For example, a washing machine shows different patterns depending on its programm settings, the wash load or even the different sections of a wash cycle. These differences within individual end-uses make the task of separating end-use categories through clustering more challenging, especially since the most significant differences in a dataset determine the outcome of clustering.

Table 2. Evaluation metrics for the clustering algorithm based on DTW.

	1P	**2P**	**3P**	**4P**	**5P**	**6P**
Estimated number of clusters	21	17	19	23	27	19
ARI	0.06	0.06	0.05	0.05	0.05	0.05
AMI	0.15	0.11	0.10	0.14	0.12	0.12

Table 3. Evaluation metrics for k-means

	1P	**2P**	**3P**	**4P**	**5P**	**6P**
ARI	0.17	0.10	0.09	0.12	0.11	0.08
AMI	0.24	0.18	0.18	0.20	0.19	0.15

Table 4. Evaluation metrics for DBSCAN

	1P	**2P**	**3P**	**4P**	**5P**	**6P**
Estimated number of clusters	2	2	1	1	1	1
ARI	0.04	0.03	0.02	0.01	0.01	0.01
AMI	0.06	0.04	0.03	0.02	0.02	0.01

Table 5. Evaluation metrics for OPTICS.

	1P	2P	3P	4P	5P	6P
Estimated number of clusters	11	17	26	21	17	27
ARI	-0.03	-0.03	-0.03	-0.01	-0.02	-0.02
AMI	0.02	0.03	0.03	0.03	0.03	0.02

Table 6. Evaluation metrics for CASH.

	1P	2P	3P	4P	5P	6P
Estimated number of clusters	30	37	44	47	51	98
ARI	-0.04	-0.04	-0.04	-0.04	-0.03	-0.06
AMI	0.04	0.04	0.04	0.05	0.05	0.07

4. Conclusions

In this work, we perform a comprehensive quantitative comparison of several supervised and unsupervised ML techniques for residential water end-use classification. A common database is created with a stochastic simulation tool based on real consumption data. One of the most important findings of the quantitative results is, that the unsupervised methods alone, i.e. clustering techniques, are not sufficient to detect the correct end-uses of domestic water consumption fully automatically. This is somewhat consistant to the implications of the state-of-the-art literatures: in the context of end-use classifications, clustering techniques are commonly employed in combination with manual processing [1] or supervised techniques [2, 3, 4].

Another conclusion we can draw from this study is, that supervised ML techniques pose an efficient way to perform water end-use classification. The accuracies and precisions of such classifiers do not only depend on the classifier itself, but are also strongly influenced by the data preprocessing and feature extraction steps. Moreover, datasets solely including the water consumption data are not sufficient for the identification of complex end-uses (e.g. *washing machine* or *dishwasher*), since significant prior knowledge, i.e. program settings of the machine or accurate start and end time of the program, are not comprised in the datasets. Consequently, annotated datasets with labeled ground truth events are essential for the application of water end-use classification.

In order to establish a large representative dataset comprising annotations, a manual labeling process is inevitable. For this purpose, end consumers in different circumstances (e.g. housing and household situation, age, gender etc.) need to be encouraged to participate. Furthermore, manual labeling tends to produce inaccurate labels, since human-beings are not predestined for repetitive tasks. Thus, quality checks of the manual labels are necessary. Additionally, technical remedies could be utilized to generate accurate end-use labels. For washing machine or dishwasher, additional sensor data, such as machine internal time logger, or external electricity logger, can be used as labels. A fully automatic labeling of some other daily end-use events (e.g. *toilet* and *faucet*) can be achieved with an additional smart water meter installed directly on the water faucet. Considering the above mentioned aspects, crowdsourcing approaches which have been applied for medical applications [25] could be a possible solution. A suitable framework to implement such approaches successfully is the so called citizen science project, where the citizens are contributing to a scientific project actively with their resources and knowledge. Scientific results and other output of the project are accessable to the participants as an exemplary reward.

Supplementary Materials: The datasets described in section 2.1 are available online at https://github.com/Nora-Go/Water_Consumption_Datasets.

Acknowledgments: The methods and information presented in this work are based on research and are not commercially available. Public foundings are not required for this study.

Author Contributions: The experiments are conceived and co-created with the contribution of all authors; N.Gourmelon performed the experiments; N. Gourmelon and S. Bayer analyzed the data; all authors listed wrote the paper.

Conflicts of Interest: The authors declare that they have no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ML: Machine-Learning
SVM: Support Vector Machine
DTW: Dynamic Time Warping
DBSCAN: Density-Based Spatial Clustering of Applications with Noise
OPTICS: Ordering Points To Identify the Clustering Structure
CASH: Clustering in Arbitrary Subspaces based on the Hough transform
ARI: Adjusted Rand Index
AMI: Adjusted Mutual Information

Appendix A

Table A1. Classification report for the *1P* dataset.

	Precision	Recall	F1-Score	Support
0 [1]	0.66	0.63	0.65	150
1 [1]	0.80	0.84	0.82	466
2 [1]	0.80	0.96	0.87	49
3 [1]	0.83	0.89	0.86	775
4 [1]	0.84	0.56	0.67	137
5 [1]	0.25	0.02	0.04	47
6 [1]	0.56	0.56	0.56	9
Accuracy [2]			0.80	1633
Macro Average	0.68	0.64	0.64	1633
Weighted Average	0.79	0.80	0.79	1633

[1] The rows in the classification report correspond to the end-use labels: 0 = Overlapping, 1 = Toilet, 2 = Shower, 3 = Faucet, 4 = Washing Machine, 5 = Dishwasher, 6 = Bathtub. [2] The micro average corresponds to the accuracy. If only one subset of the classes is present in the predictions for the test data, the two metrics are not equivalent.

Table A2. Classification report for the *2P* dataset.

	Precision	Recall	F1-Score	Support
0 [1]	0.69	0.66	0.68	289
1 [1]	0.79	0.87	0.83	692
2 [1]	0.89	0.91	0.90	56
3 [1]	0.85	0.86	0.86	1098
4 [1]	0.77	0.58	0.66	110
5 [1]				0
6 [1]	1.00	0.50	0.67	2
Micro Average	0.81	0.82	0.82	2247
Macro Average	0.83	0.73	0.77	2247

Weighted Average	0.81	0.82	0.82	2247

[1] The rows in the classification report correspond to the end-use labels: 0 = Overlapping, 1 = Toilet, 2 = Shower, 3 = Faucet, 4 = Washing Machine, 5 = Dishwasher, 6 = Bathtub.

Table A3. Classification report for the *3P* dataset.

	Precision	**Recall**	**F1-Score**	**Support**
0 [1]	0.69	0.70	0.69	379
1 [1]	0.77	0.80	0.78	708
2 [1]	0.87	0.92	0.90	66
3 [1]	0.81	0.85	0.83	1253
4 [1]	0.82	0.61	0.70	137
5 [1]				0
6 [1]	0.80	0.57	0.67	7
Micro Average	0.78	0.80	0.79	2550
Macro Average	0.79	0.74	0.76	2550
Weighted Average	0.78	0.80	0.79	2550

[1] The rows in the classification report correspond to the end-use labels: 0 = Overlapping, 1 = Toilet, 2 = Shower, 3 = Faucet, 4 = Washing Machine, 5 = Dishwasher, 6 = Bathtub.

Table A4. Classification report for the *4P* dataset.

	Precision	**Recall**	**F1-Score**	**Support**
0 [1]	0.78	0.75	0.77	471
1 [1]	0.79	0.83	0.81	844
2 [1]	0.95	0.89	0.92	89
3 [1]	0.82	0.86	0.84	1275
4 [1]	0.76	0.71	0.73	166
5 [1]				0
6 [1]	1.00	0.33	0.50	3
Micro Average	0.81	0.82	0.81	2848
Macro Average	0.85	0.73	0.76	2848
Weighted Average	0.81	0.82	0.81	2848

[1] The rows in the classification report correspond to the end-use labels: 0 = Overlapping, 1 = Toilet, 2 = Shower, 3 = Faucet, 4 = Washing Machine, 5 = Dishwasher, 6 = Bathtub.

Table A5. Classification report for the *5P* dataset.

	Precision	**Recall**	**F1-Score**	**Support**
0 [1]	0.70	0.77	0.73	505
1 [1]	0.77	0.82	0.79	873
2 [1]	0.94	0.93	0.93	113
3 [1]	0.81	0.83	0.82	1315
4 [1]	0.81	0.53	0.64	207
5 [1]	0.60	0.05	0.08	66
6 [1]	0.80	0.60	0.69	20
Accuracy [2]			0.78	3099

Macro Average	0.77	0.65	0.67	3099
Weighted Average	0.78	0.78	0.77	3099

[1] The rows in the classification report correspond to the end-use labels: 0 = Overlapping, 1 = Toilet, 2 = Shower, 3 = Faucet, 4 = Washing Machine, 5 = Dishwasher, 6 = Bathtub. [2] The micro average corresponds to the accuracy. Solely when only a subset of the classes is present in the predictions for the test data, the two metrics differ.

Table A6. Classification report for the *6P* dataset.

	Precision	**Recall**	**F1-Score**	**Support**
0 [1]	0.73	0.78	0.76	747
1 [1]	0.73	0.82	0.77	956
2 [1]	0.87	0.82	0.84	112
3 [1]	0.85	0.83	0.84	1604
4 [1]	0.78	0.29	0.42	132
5 [1]	0.67	0.04	0.08	50
6 [1]	1.00	0.50	0.67	4
Accuracy [2]			0.79	3605
Macro Average	0.80	0.58	0.62	3605
Weighted Average	0.79	0.79	0.78	3605

[1] The rows in the classification report correspond to the end-use labels: 0 = Overlapping, 1 = Toilet, 2 = Shower, 3 = Faucet, 4 = Washing Machine, 5 = Dishwasher, 6 = Bathtub. [2] The micro average corresponds to the accuracy. Solely when only a subset of the classes is present in the predictions for the test data, the two metrics differ.

Appendix B

Table A7. Confusion matrices for the binary SVMs.

	1P	**2P**	**3P**	**4P**	**5P**	**6P**
0 [1]	$\begin{bmatrix}1456 & 27\\ 73 & 77\end{bmatrix}$	$\begin{bmatrix}1941 & 53\\ 127 & 162\end{bmatrix}$	$\begin{bmatrix}2164 & 71\\ 154 & 225\end{bmatrix}$	$\begin{bmatrix}2377 & 62\\ 149 & 322\end{bmatrix}$	$\begin{bmatrix}2492 & 102\\ 159 & 346\end{bmatrix}$	$\begin{bmatrix}2719 & 139\\ 239 & 508\end{bmatrix}$
1 [1]	$\begin{bmatrix}1077 & 90\\ 79 & 387\end{bmatrix}$	$\begin{bmatrix}1470 & 121\\ 121 & 571\end{bmatrix}$	$\begin{bmatrix}1759 & 147\\ 168 & 540\end{bmatrix}$	$\begin{bmatrix}1898 & 168\\ 174 & 670\end{bmatrix}$	$\begin{bmatrix}2032 & 194\\ 188 & 685\end{bmatrix}$	$\begin{bmatrix}2424 & 225\\ 256 & 700\end{bmatrix}$
2 [1]	$\begin{bmatrix}1579 & 5\\ 4 & 45\end{bmatrix}$	$\begin{bmatrix}2224 & 3\\ 7 & 49\end{bmatrix}$	$\begin{bmatrix}2542 & 6\\ 6 & 60\end{bmatrix}$	$\begin{bmatrix}2817 & 4\\ 14 & 75\end{bmatrix}$	$\begin{bmatrix}2983 & 3\\ 9 & 104\end{bmatrix}$	$\begin{bmatrix}3486 & 7\\ 25 & 87\end{bmatrix}$
3 [1]	$\begin{bmatrix}741 & 117\\ 100 & 675\end{bmatrix}$	$\begin{bmatrix}1062 & 123\\ 180 & 918\end{bmatrix}$	$\begin{bmatrix}1167 & 194\\ 233 & 1020\end{bmatrix}$	$\begin{bmatrix}1448 & 187\\ 225 & 1050\end{bmatrix}$	$\begin{bmatrix}1588 & 196\\ 290 & 1025\end{bmatrix}$	$\begin{bmatrix}1835 & 166\\ 312 & 1292\end{bmatrix}$
4 [1]	$\begin{bmatrix}1484 & 12\\ 73 & 64\end{bmatrix}$	$\begin{bmatrix}2159 & 14\\ 57 & 53\end{bmatrix}$	$\begin{bmatrix}1077 & 90\\ 79 & 387\end{bmatrix}$	$\begin{bmatrix}2718 & 26\\ 82 & 84\end{bmatrix}$	$\begin{bmatrix}2871 & 21\\ 124 & 83\end{bmatrix}$	$\begin{bmatrix}3471 & 2\\ 119 & 13\end{bmatrix}$
5 [1]	$\begin{bmatrix}1586 & 0\\ 47 & 0\end{bmatrix}$	$\begin{bmatrix}2247 & 0\\ 36 & 0\end{bmatrix}$	$\begin{bmatrix}2550 & 0\\ 64 & 0\end{bmatrix}$	$\begin{bmatrix}2848 & 0\\ 62 & 0\end{bmatrix}$	$\begin{bmatrix}3033 & 0\\ 66 & 0\end{bmatrix}$	$\begin{bmatrix}3555 & 0\\ 50 & 0\end{bmatrix}$
6 [1]	$\begin{bmatrix}1620 & 4\\ 5 & 4\end{bmatrix}$	$\begin{bmatrix}2281 & 0\\ 1 & 1\end{bmatrix}$	$\begin{bmatrix}2606 & 1\\ 3 & 4\end{bmatrix}$	$\begin{bmatrix}2907 & 0\\ 2 & 1\end{bmatrix}$	$\begin{bmatrix}3076 & 3\\ 8 & 12\end{bmatrix}$	$\begin{bmatrix}3601 & 0\\ 3 & 1\end{bmatrix}$

[1] The rows correspond to the end-use labels: 0 = Overlapping, 1 = Toilet, 2 = Shower, 3 = Faucet, 4 = Washing Machine, 5 = Dishwasher, 6 = Bathtub.

References

1. Pastor-Jabaloyes, L.; Arregui, F.J.; Cobacho, R. Water End Use Disaggregation Based on Soft Computing Techniques. *Water* **2018**, *10*, 46, https://doi.org/10.3390/w10010046. Available online: https://www.mdpi.com/2073-4441/10/1/46 (accessed on 30/10/2020).
2. Nguyen, K.A.; Zhang, H.; Stewart, R.A. Application of Dynamic Time Warping Algorithm in Prototype Selection for the Disaggregation of Domestic Water Flow Data into End Use Events. Proceedings of the 34th World Congress of the International Association for Hydro- Environment Research and Engineering: 33rd Hydrology and Water Resources Symposium and 10th Conference on Hydraulics in Water Engineering, Brisbane, Australia, 26/06/2011 – 1/07/2011; Valentine, E.M.; Apelt, C.J.; Ball, J.; Chanson, H.; Cox, R.; Ettema, R.; Kuczera, G.; Lambert, M.; Melville, BW (Editor); Sargison, JE; Engineers Australia: Barton, A.C.T., Australia 2011.
3. Yang, A.; Zhang, H.; Stewart, R.A., Nguyen, K.A. Water End Use Clustering Using Hybrid Pattern Recognition Techniques - Artificial Bee Colony, Dynamic Time Warping and K-Medoids Clustering. *IJMLC* **2018**, 8, 483–487, 10.18178/ijmlc.2018.8.5.733. Available online: https://research-repository.griffith.edu.au/bitstream/handle/10072/380953/YangPUB5696.pdf?sequence=1 (accessed on 30/10/2020).
4. Yang, A. Artificial Intelligent Techniques in Residential Water End-use Studies for Optimized Urban Water Management Artificial. Master's Thesis, Griffith University, Brisbane, Australia, 26/09/2018.
5. Kalogridis, G.; Farnham, T.; Wilcox, J.; Faies, M. Privacy and Incongruence-Focused Disaggregation of Water Consumption Data in Real Time. *Procedia Engineering* **2015**, *119*, 854–863, 10.1016/j.proeng.2015.08.950. Available online: https://core.ac.uk/download/pdf/82459692.pdf (accessed on 13/04/2020).
6. Nguyen, K.A.; Stewart, R.A.; Zhang, H. An Intelligent Pattern Recognition Model to Automate the Categorisation of Residential Water End-Use Events. *Environmental Modelling and Software* **2013**, *47*, 108–127, https://doi.org/10.1016/j.envsoft.2013.05.002. Available online: https://www.sciencedirect.com/science/article/abs/pii/S1364815213001084 (accessed on 02/04/2020).
7. Nguyen, K.A.; Zhang, H.; Stewart, R.A. Development of an Intelligent Model to Categorise Residential Water End Use Events. *Journal of Hydro-Environment Research* **2013**, *7*, 182–201, https://doi.org/10.1016/j.jher.2013.02.004. Available online: https://www.sciencedirect.com/science/article/abs/pii/S1570644313000221 (accessed on 02/04/2020).
8. Nguyen, K.A.; Stewart, R.A.; Zhang, H. An Autonomous and Intelligent Expert System for Residential Water End-Use Classification. *Expert Systems with Applications* **2014**, *41*, 342–356, https://doi.org/10.1016/j.eswa.2013.07.049. Available online: https://www.sciencedirect.com/science/article/abs/pii/S0957417413005368 (accessed on 02/04/2020).
9. Nguyen, K.A.; Stewart, R.A.; Zhang, H. Development Of An Autonomous And Intelligent System For Residential Water End Use Classification. Proceedings of the 11th International Conference on Hydroinformatics, New York, USA, 17/08/2014 – 21/08/2014; Curran Associates, Inc.: New York, USA, 2015.
10. Nguyen, K.A.; Stewart, R.A.; Zhang, H.; Jones, C. Intelligent Autonomous System for Residential Water End Use Classification: Autoflow. *Applied Soft Computing Journal* **2015**, *31*, 118–131, https://doi.org/10.1016/j.asoc.2015.03.007. Available online: https://www.sciencedirect.com/science/article/abs/pii/S1568494615001519 (accessed on 02/04/2020).
11. Nguyen, K.A.; Sahin, O.; Stewart, R.A. AUTOFLOW© – A Novel Application for Water Resource Management and Climate Change Response Using Smart Technology. Proceedings of the 8th International Congress on Environmental Modelling and Software, Toulouse, France, 10/07/2016-14/07/2016.
12. Nguyen, K.A.; Sahin, O.; Stewart, R.A.; Zhang, H. Smart Technologies in Reducing Carbon Emission: Artificial Intelligence and Smart Water Meter. Proceedings of the 9th International Conference on Machine Learning and Computing, Singapore, Singapore, 24/02/2017 – 26/02/2017; Association for Computing Machinery: New York, NY, USA, 2017.
13. Nguyen, K.A.; Stewart, R.A.; Zhang, H.; Sahin, O. An Adaptive Model for the Autonomous Monitoring and Management of Water End Use. *Smart Water* **2018**, *3*, 1–21, https://doi.org/10.1186/s40713-018-0012-7. Available online: https://link.springer.com/article/10.1186/s40713-018-0012-7 (accessed on 30/10/2020).
14. Nguyen, K.A.; Stewart, R.A.; Zhang, H.; Sahin, O.; Siriwardene, N. Re-Engineering Traditional Urban Water Management Practices with Smart Metering and Informatics. *Environmental Modelling and Software* **2018**, *101*, 256–267, https://doi.org/10.1016/j.envsoft.2017.12.015. Available online:

https://www.sciencedirect.com/science/article/abs/pii/S1364815217305893 (accessed on 13/04/2020).
15. Cominola, A. Modelling Residential Water Consumers' Behavior. PhD thesis, Politecnico di Milano, Milano, Italy, 2016.
16. Cominola, A.; Giuliani, M.; Castelletti, A.; Rosenberg, D.E.; Abdallah, A.M. Implications of Data Sampling Resolution on Water Use Simulation, End-Use Disaggregation, and Demand Management. *Environmental Modelling and Software* **2018**, *102*, 199–212, https://doi.org/10.1016/j.envsoft.2017.11.022. Available online: https://www.sciencedirect.com/science/article/abs/pii/S1364815217311301 (accessed on 22/04/2020).
17. Vašak, M.; Banjac, G.; Novak, H. Water Use Disaggregation Based on Classification of Feature Vectors Extracted from Smart Meter Data. *Procedia Engineering* **2015**, *119*, 1381–1390, https://doi.org/10.1016/j.proeng.2015.08.992. Available online: https://www.sciencedirect.com/science/article/pii/S1877705815026624 (accessed on 30/10/2020).
18. Lloyd, S.P. Least Squares Quantization in PCM. *IEEE Transactions on Information Theory* **1982**, *28*, 129–137, doi: 10.1109/TIT.1982.1056489. Available online: https://ieeexplore.ieee.org/abstract/document/1056489 (accessed on 02/11/2020).
19. Jain, A.K. Data Clustering: 50 Years beyond K-Means. *Pattern Recognition Letters* **2010**, *31*, 651–666, https://doi.org/10.1016/j.patrec.2009.09.011. Available online: https://www.sciencedirect.com/science/article/abs/pii/S0167865509002323 (accessed on 02/11/2020).
Halkidi, M.; Batistakis, Y.; Vazirgiannis, M. On Clustering Validation Techniques. *Journal of Intelligent Information Systems* **2001**, *17*, 107–145, https://doi.org/10.1023/A:1012801612483. Available online: https://link.springer.com/article/10.1023/A:1012801612483 (accessed on 02/11/2020).
20. Ester, M.; Kriegel, H.; Sander, J.; Xu, X. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, Portland, Oregon, 02/08/1996 – 04/08/1996.
21. Sander, J.; Ester, M.; Kriegel, H.; Xu, X. Density-Based Clustering in Spatial Databases: The Algorithm GDBSCAN and Its Applications. *Data Mining and Knowledge Discovery* **1998,** *2*, 169–194, https://doi.org/10.1023/A:1009745219419. Available online: https://link.springer.com/article/10.1023/A:1009745219419 (accessed on 08/09/2020).
22. Ankerst, M.; Breunig, M.M.; Kriegel, H.P.; Sander, J. OPTICS: Ordering Points to Identify the Clustering Structure. *SIGMOD Record* **1999**, *28*, 49–60, 10.1145/304181.304187. Available online: https://dl.acm.org/doi/abs/10.1145/304181.304187 (accessed on 02/11/2020).
23. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, E. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* **2011**, *12*, 2825–2830, issn:1532-4435. Available online: https://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf (accessed on 02/11/2020).
24. Achtert, E.; Böhm, C.; David, J.; Kröger, P.; Zimek, A. Global Correlation Clustering Based on the Hough Transform. *Statistical Analysis and Data Mining: The ASA Data Science Journal* **2008**, *1*, 111–127, https://doi.org/10.1002/sam.10012. Available online: https://onlinelibrary.wiley.com/doi/abs/10.1002/sam.10012 (accessed on 02/11/2020).
25. Marzahl, C.; Aubreville, M.; Bertram, C.A.; Gerlach, S.; Maier, J.; Voigt, J.; Hill, J.; Klopfleisch, R.; Maier, A. Is Crowd-Algorithm Collaboration an Advanced Alternative to Crowd-Sourcing on Cytology Slides?. In *Bildverarbeitung für die Medizin 2020*; Tolxdorff T., Deserno T., Handels H., Maier A., Maier-Hein K., Palm C., Eds.; Publisher: Springer Vieweg, Wiesbaden, Germany, 2020; pp. 26-31.