Anomalous Water Use Detection Using Machine Learning

Lukas Kulikovas [1] and Šarūnas Packevičius [1]

[1]Faculty of Informatics, Kaunas University of Technology, Kaunas, Lithuania

Abstract

Water is an essential resource that is necessary for human life, agriculture, and industry. Numerous countries confront water shortages and inefficient water usage. Anomalous water usage detection is an important task in the efficient management of water resources and the prevention of water leaks. In this publication, we present a comparison between various machine learning models to detect unusual patterns in water usage data. All the machine learning models were tested on a real-world water usage dataset. The performance of each model was evaluated by accuracy, precision, recall, F1-score, ROC AUC, and MAE scores. The results indicate that PCA outlier detector can accurately detect uncommon patterns in water usage data. Our results outlined in this paper might be utilized by either individual homeowners or water utility corporations to detect water leaks more quickly and hence minimize water wastage.

Keywords

Anomalous water use detection, unsupervised learning, semi-supervised learning.

1. Introduction

Water is a limited and invaluable resource, which plays a crucial role in supporting every life on the Earth. As the population grows and urbanization develops, so will the need for water. With increased concerns about water shortage and its wasteful use, proper water management has emerged as a significant worldwide challenge.

According to UNESCO, global water consumption has increased by about 1% per year since the 1980s, driven by growing populations and changing habits of water consumption [1]. In accordance with Burek et al. [2], worldwide water use will likely continue to grow at 1% yearly rate, culminating in an increase of 20 to 30% above current levels by 2050. Approximately 2.2 billion people do not have access to safe drinking water, roughly 4.2 billion people face acute water scarcity for at least one month each year, and around three billion individuals do not have access to basic handwashing facilities [1].

In the USA a typical household uses about 138 gallons (~ 522 liters) of water every day, where the toilet flush accounts for the majority of this use (24%), followed by faucets (20%), showers (20%), clothes washers (16%), leaks (13%), baths (3%), others (3%) and dishwashers (2%) [3]. Even while leaks account for just 13% of total home water usage, approximately 1 trillion gallons (3.785 trillion liters) of water can be wasted by residential leaks in the United States, with the average household's leaks accounting for nearly 10,000 gallons (~37,854 liters) of water every year [4].

The "Leaving No One Behind" report highlights the importance of improving water resources management and how it is crucial to address various problems, such as poverty, health, food security and environmental sustainability [5]. Water leaks are a sign of larger concerns, which come from outdated infrastructure, poor maintenance procedures and inefficient water use management [5]. As recommended by World Water Assessment Programme (WWAP), optimization of water resource management could aid in the prevention of water leaks, which include frequent pipe monitoring, maintenance, enhanced metering, and leak identifications systems [5].

IVUS 2023: Information Society and University Studies, May 12, 2023, Kaunas, Lithuania
EMAIL: lukas.kulikovas@ktu.edu (L. Kulikovas); sarunas.packevicius@ktu.lt (Š. Packevičius)
ORCID: 0000-0001-7653-8526 (L. Kulikovas)

Because water resources are limited, we must manage them effectively and responsibly to assure that it will be available for our future generations. To enhance water metering and existing water management systems, we compare a number of machine learning models to see how they perform under different water usage conditions. We begin by presenting relevant research on the topic of water leak anomaly detection, then we analyze various machine learning models and select the best one for detecting water usage abnormalities.

2. Related Work

This interesting yet difficult field, known as anomaly or outlier detection, has been acknowledged and thoroughly investigated by a plethora of research over the years. Han et al. [6] scrutinized how, in the last two decades, developed anomaly detection algorithms perform with regard to varying levels of supervision, different types of abnormalities and noisy and polluted data. Boukerche et al. [7] present a taxonomy of newly created outlier identification algorithms and approaches for high-dimensional data, data streams, big data, and little labeled data, followed by an overview of benefits and limitations for each algorithm. Wang et al. [8] presented a comprehensive and organized review of the progress of outlier detection methods. Chandola et al. [9] presented basic anomaly detection techniques, followed by an overview of the advantages and disadvantages of each technique. Campos et al. [10] conducted an extensive experimental study on the performance of a representative set of standard k-nearest neighborhood-based methods for unsupervised outlier detection.

Various techniques have been developed, which have been applied to a variety of real-life scenarios, that include intrusion detection systems, fraud detection, medical anomaly diagnosis, anomaly detection in wireless sensor networks and urban traffic flow [7]. One of the real-life scenarios is an anomalous water use detection system, which has the main objective of detecting unusual water use.

It is critical to distinguish between the numerous types of anomaly detection algorithms used to identify water leaks. Where physics and predefined expert rules seem appropriate to use, a traditional anomaly detection approach might be applied. This strategy, however, may not be appropriate in various situations, and labeled data may be difficult to get in this context. Furthermore, because anomalies are usually infrequent and unexpected, unsupervised learning algorithms have grown in prominence for their ability to detect abnormalities in unlabeled data. Where there is a possibility to obtain some labeled data points, semi-supervised techniques are used to train the algorithm, whereas fully supervised methods use solely labeled data. Active learning models are another way that uses expert or user input to categorize data and enhance algorithm accuracy. To further understand what research has been done and how different approaches function, the next paragraph will discuss traditional, unsupervised, semi-supervised, active learning, and fully supervised anomaly detection algorithms.

2.1. Traditional Anomaly Detection Methods

Sarangi [11] presented a technique for detecting water leaks and theft that is based on the concept of conservation of mass, which states that mass cannot be generated or destroyed. The author provided a solution to stop water theft by installing two sensors, where one of them detects the amount of water flow coming into the pipeline, the other one – water leaving the pipeline. If the difference in data collected by both sensors exceeds a specified limit defined to minimize false alarms, the microcontroller will send an alarm for further investigation of that location. Similarly, if there is a pipeline burst or water theft, the difference between the two measurements will be large.

Moni et al. [12] discussed how their approach could benefit farmers to detect water leaks. For this, they collected leak and no-leak vibration data from a pipeline using an accelerometer in three linear axes – x, y, and z. The results suggested that the root mean square (RMS) error, which in this case represents the difference between the measured vibration data and the expected vibration data for each condition, was always smaller when there was a leak. Final evaluation results indicated 87.9% accuracy when there was a leak and 96.3% accuracy when there was not.

Boudhaouia and Wira [13] proposed a general solution for collecting and managing water consumption by a non-intrusive approach which works at any measurement point from a water distribution system. The suggested approach is based on three parameters: maximum daily load curve, minimum night flow (MNF), and a non-null time period calculated from water flow rate (PWNC). During nighttime, minimum night flow parameter is used. The algorithm considers the fact that average day flow rate is different from zero and night water use must be close to zero, otherwise it is classified as a small leak. During daytime, maximum daily load curve and MFN are used. The maximum daily load curve reflects a maximum water usage threshold. The detection is performed by comparing the current consumption to the specified limit. If the current water usage exceeds or is at the given threshold, accordingly a big leakage or risk of leakage is reported. Finally, a period without null consumption parameter is defined, which gives an overview of how long the water may run before it is considered a leak. Authors claim that their proposed procedure detected all water leakages.

2.2. Unsupervised Water Anomaly Detection

Ji et al. [14] compared linear regression, ARIMA and additive regression models using the data of Osan City to find the best method for water leak detection. First, authors calculated Watson value for the data, which resulted in 2.83E-05. This indicated that the linear regression model was unfit for the provided dataset. For this reason, time series model ARIMA and additive regression Prophet models were proposed. The ARIMA model resulted in an accuracy rate of 64% with an average MAE for all houses of 39,230.88. Fbprophet provided a smaller MAE of 17,635.15 and smaller accuracy rate of 46% considering yearly trend, and accuracy rate was 65% with slightly higher MAE of 23,566.54 without considering yearly trend.

Fuentes and Mauricio [15] presented a smart water consumption measurement system, which involves house data collection, analyzation, and leakage alert functionality. Authors explain their 4-scenario algorithm, which involves negative trend evaluation, last 24-hour consumption evaluation, similar consumption evaluation and historical data process. Authors extract historical data into different features, apply the k-NN algorithm to obtain a list of the consumptions that are closest to (K = 4) and apply Tchebysheff theorem for confidence interval construction. Outstandingly, the algorithm demonstrated accuracy, recall, precision, and F1-score as 100%, surpassing the rest of the leak detection algorithms.

Patabendige et al. [16] developed a context aware anomaly detection algorithm that takes the relevant context for each day into account, applies the k-NN algorithm together with Gaussian error function to transform the outlier score into a probability value. The system also generates an anomaly score for each day together with a rationale that describes what could have caused an unusual water use and reports it to the user.

2.3. Semi-supervised Water Anomaly Detection

Lee et al. [17] built RNN-LSTM (Recurrent Neural Networks-Long Short-Term Memory) deep learning model. Authors applied a model on the actual leakage data and the leak was recognized at most points immediately after the accident. Also, this model resulted in good performance and showed more than 90% accuracy. Authors also mention that the model is highly scalable.

Pang et al. [17] proposed a deep reinforcement learning-based approach that enables an end-to-end optimization of the detection of both labeled and unlabeled data. This approach learns the known abnormalities by automatically interacting with an anomaly-biased simulation environment, while continuously extending the learned abnormality to novel classes of anomaly by actively exploring anomalies in the unlabeled data. The authors demonstrated how experiments on 48 real-world datasets proved that their model outperforms state-of-the-art competing methods.

Blázquez-García et al. [18] proposed a self-supervised water leak detection method based on a self-supervised classification of flow time series, called Self-Supervised Leak Detector (SSLD). This algorithm does not require external class labels and instead uses labels that have been assigned to artificially generated data. In the first step of their self-supervised framework, a self-labeled training set is generated. Later, the classifier is trained to learn the mapping between input and its corresponding label. Authors concluded that proposed SSLD method obtains the best trade-off between detecting the majority of the detectable leaks and providing a low FPR. Also, the provided model is purely data-driven and therefore does not require in-depth knowledge about the dynamics of the series.

2.4. Active Learning Approach for Water Anomaly Detection

Numerous authors proposed various active-learning-based algorithms that can interactively query user's response to label data with the desired outputs. Wang et al. [19] proposed an active anomaly detection framework Active-MTSAD. Das et al. [20] proposed the Active Anomaly Discovery (AAD) algorithm. Zhu and Yang [21] proposed the tripartite active learning method. Vercruyssen et al. [22] proposed a novel constrained-clustering-based approach for anomaly detection that works in both unsupervised and semi-supervised setting. Active learning approach strategy starts with unsupervised learning where most important unlabeled instances are selected and then provided to the expert or user. Later, the model gets updated with new labels so it can achieve higher performance. This type of learning could adapt to the user's needs and provide better results. As authors mention, their experiments demonstrated how active learning models outperform most methods for domain-specific anomaly detection [19].

2.5. Supervised Water Anomaly Detection

Ismail et al. [23] proposed a comparison between four machine learning classification models. First, human annotators generated a ground truth dataset for water consumption. Secondly, the data was normalized using the z-score. Later, the data was applied to four machine learning models, including Decision Tree, k-NN, Naïve Bayes, and Random Forest. Authors concluded that Random Forest machine learning model gives the highest overall accuracy of 87%, precision of 75% and recall of 83 % compared to other three classification models.

Amora et al. [24] designed a Bidirectional LSTM (BiLSTM) machine learning model and compared it with Gated recurrent units (GRU) and Autoregressive models. The suggested method has two iterative loops, in which the outer loop optimizes the batch size, number of input time steps, and number of output units in each LSTM. Also, Hyperactive library and mean square error are used in this simulation. The inner loop of the suggested method updates each weighting element in the BiLSTM using the traditional Adam optimizer. According to the results gathered by authors, BiLSTM outperforms the GRU and Autoregression models when detecting water leak.

A supervised learning approach in water anomaly detection most of the time is a challenge due to lack of labels. Zese et al. [25] applied several supervised machine learning techniques for the automatic detection of leakages. Authors demonstrated how convolutional neural network models were the best in detecting both the presence and absence of water leaks. Overall results showed that the model was able to classify water leaks with accuracy, precision, recall, F-measure, and AUC ROC ranging from 92% to 99% [25]. Fan et al. [26] demonstrated how artificial neural networks (ANN) can accurately classify leaking versus non-leaking scenarios. However, it requires a balanced dataset under both leaking and non-leaking conditions. Nonetheless, authors claim that their model detected leaks in pipes with 100% accuracy [26].

3. The Data

In this study, we use SWM (smart water meter) time-series (Trial A) data from the DAIAD Trials, which is available on GitHub[2], and we also include one of our water use data, which was gathered from a smart water sensor in one person's home. Our water consumption data is also available on GitHub[3].

The DAIAD dataset contains hourly water consumption measurements for ninety-two households in Alicante. Each time-series starts at 1/3/2016 and ends at 28/2/2018. On average, there are 7108 measurements per household, which in total is 653,954 records. The outliers amount to 322 out of the total records in the dataset [27]. As they are a typical user behavior and not the water leak, thus any prediction as anomalous on the dataset's outlier would be considered false positive.

[2] https://github.com/DAIAD/data/blob/master/swm_trialA_clean.zip
[3] https://github.com/LukaLike/water-consumption-data

Our time-series dataset contains minutely measurements, starting on 8/13/2022 and ending at 3/2/2023. In total, this dataset contains 263,811 records. This dataset has no outlier and any prediction as anomalous on this dataset would also be considered false positive.

4. Methodology

In this section, we present a comparison of different outlier detection models to see which one is the most efficient in detecting anomalous water use. Different scenarios are used to achieve this, and important statistical parameters are evaluated. In these experiments we include CBLOF, COPOD, ECOD, HBOS, IForest, KNN, LOF, OCSVM, and PCA outlier detector models that are publicly available on GitHub platform[4]. The semi-supervised SSDO model that is used in these experiments is also publicly available on the GitHub platform[5].

To extensively analyze various machine learning models for detecting anomalous water use, we first prepared the data. This included reading timestamp and consumption values from the datasets. To ensure the quality of the data, preprocessing was applied, which included filling missing values and ignoring days with no consumption. Afterwards, different leak scenarios were added to the datasets. Throughout the data preparation process, we ensured that any modifications made to the dataset were performed only on the data that was temporarily stored in memory. This enabled us to maintain the original dataset intact for future runs. Subsequently, a feature and label matrix were built.

Since multiple tests were done, several feature matrices and their sizes were created, including 2, 3, or 4-dimensional feature matrices with mean, min, max, or longest non-zero water flowing duration, and varying window range sizes. The hourly datasets were separated into 1-, 2-, 3-, 4-, and 6-hour window sizes, while the minutely datasets were divided into 1-, 5-, and 10-minute window sizes. In this regard, labels matrices were created, which included values indicating whether or not a given window range had a water leak. Because none of the datasets had any unusual water use points, only the values generated by the water leak function generator were marked as anomalous. Finally, the data was separated into training and testing sets and submitted to all outlier detection algorithms.

In the following sections, we are going to further explain how we prepared the data and scenarios, and how we completed the evaluation of the models.

4.1. Data and Scenarios Preparation

To evaluate each model, correct and most realistic data must have been used. For this, the following process was followed:

1. In the first step, we prepared the data. This included filling the missing values with zeros and removing the days that had not used any water.
2. Secondly, water leak scenarios were generated and included into the dataset at the program's runtime. Each assessment was performed 5 times and each time new random water leak scenarios were included. The model was fit on the first 80% and tested on the last 20% of the data. In the training randomly [0; 3], and in the testing data – [1; 3] leak scenarios were included. Each scenario had a chance to overlap, contain the same scenarios, and last between 5 and 180 minutes. All of the possible scenarios are shown in **Table 1**.

Table 1
Water leak scenarios

Scenario	Minimum flowing speed (ml/min)	Maximum flowing speed (ml/min)
Dripping pipe	33.1	35.9
Dripping faucet	539.8	573.2
Broken mainline	3671.9	3899

[4] https://github.com/yzhao062/pyod
[5] https://github.com/Vincent-Vercruyssen/anomatools

Flowing faucet	7343.7	7798
Leaking toilet	15434	16389
Broken sprinkler	45898.1	48737.2

4.2. Evaluation

To evaluate each model's performance, we calculate accuracy, precision, recall, F1-score, ED-score, AUC ROC, and AUC PR values.

Precision

It is a metric that assesses the proportion of a model's positive predictions that are actually correct. Its definition is the proportion of accurate positive predictions to all positive predictions.

$$Precision = \frac{TP}{TP + FP}, \quad (1)$$

where:
TP: True Positive, the number of correct positive predictions
FP: False Positive, the number of incorrect positive predictions

Recall

It is a metric that assesses how many actual positive instances a model can identify. It is determined by dividing the total number of positive occurrences by the proportion of actual positive predictions.

$$Recall = \frac{TP}{TP + FN}, \quad (2)$$

where:
FN: False Negative, the number of incorrect negative predictions

F1-score

It is an evaluation metric that is defined as the harmonic mean of the precision P and recall R.

$$\mathrm{F}_1 = \frac{2PR}{P + R}. \quad (3)$$

AUC-ROC: Area Under the Receiver Operating Characteristic Curve

It is a statistic used to assess how well binary classification models perform. The AUC-ROC ranges from 0 to 1, with a score of 1 denoting flawless performance and a score of 0.5 denoting no improvement over a random guessing.

AUC-PR: Area Under the Precision-Recall Curve

It is also a metric used to assess how well binary classification models perform. The AUC-PR concentrates on the model's precision and recall rather than the true positive and false positive rate, like AUC-ROC does. At various levels, it calculates the precision and recall trade-off. The AUC-PR scales from 0 to 1, with a score of 1 denoting perfect performance and a score of 0.5 denoting no better than a random guessing.

ED-score: Early Detection score

It is a score that rewards detections that are close to the fault start time t_f, with the reward decreasing as the detection moves further away. The detection time is defined as the earliest time step within the fault window duration at which the algorithm registered a detection. In our tests, as in Vercruyssen's, a successful detection is recorded if the algorithm generates detections that persist for at least 75% of the time. The early detection score is calculated by first determining the delay of the first detection in the defined fault window, given by $x = t_d - t_f$, and then applying the following sigmoid function to this detection [22]:

$$\sigma(x) = \frac{2}{1 + e^{\left(\frac{\alpha}{T_w} * x\right)}}, \tag{4}$$

where:
σ – value that is defined such that $\sigma(x) \approx 0$, while $\sigma(0) = 1$ for any values of $\alpha < \infty$,
T_w – fault window duration [22].

In our tests, α was selected as 6. A detection that occurs outside the fault window is not included in the score.

5. Analyzes and Results

In the following sections, we present machine learning models comparisons in different scenarios, which include time windows, contexts, and data dimensions. The default parameter of each evaluation strategy is fixed 3 hour, contextual, 1 dimension (mean), non-overlapping window interval. For each strategy, only one default parameter will be changed or if stated otherwise.

5.1. Models Evaluation on Different Time Windows

During this experiment, we evaluated the performance of the models at different sliding window sizes. First, the models were tested using DAIAD data (in hourly frames) with a window width of 3 hours and a sliding window size of 1, 2, 3, 4, and 6 hours, respectively. The models were then tested using data from the system's developer (in minute frames) with a window width of 3 hours and a sliding window size of 1, 5, and 10 minutes, respectively. In the graphs, the *x* axis indicates the size of the sliding window, and the *y* axis indicates the model's performance for the specified criterion.

Comparison of the models in different hourly time frames

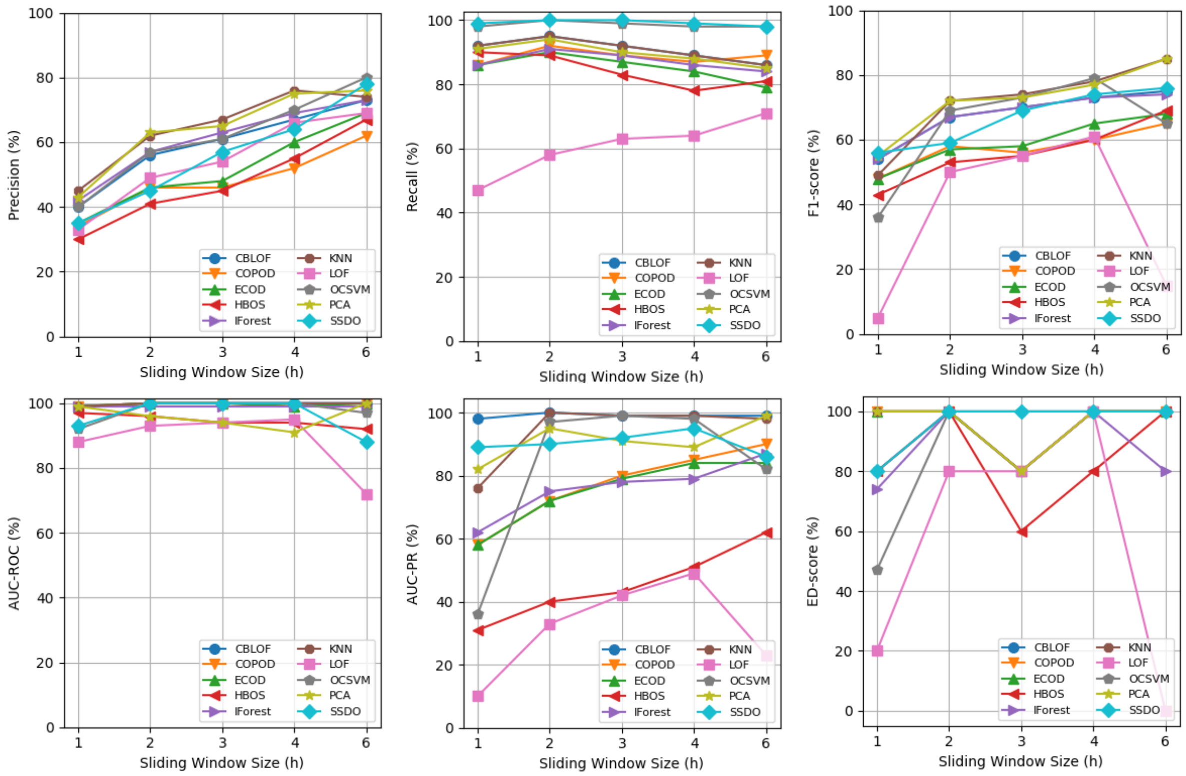

Figure 1: Comparison of models in different hourly frames

The experiment with different hourly frames showed that the precision decreased 2 times as the sliding window decreased from 6h to 1h, but the recall remained the same in most of the models. Also looking at the ED-score, most models performed poorly in terms of early detection of water leakage with a 1h rolling window. The semi-supervised SSDO model was able to perform better than most of the unsupervised machine learning models.

Comparison of the models in different minutely time frames

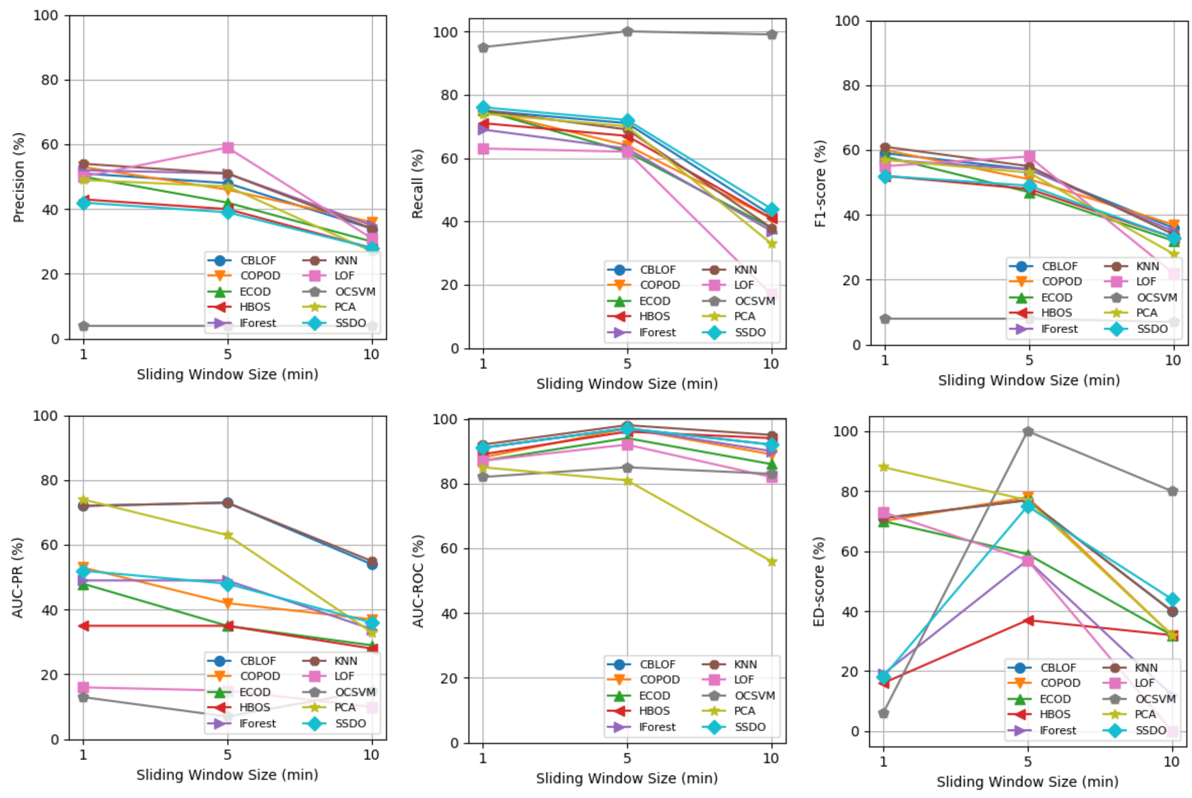

Figure 2: Comparison of models in different minutely frames

The experiment with different minute frames showed that decreasing the size of the rolling window from 10 minutes to 5 minutes improved accuracy and recall much more than changing the window size from 5 minutes to 1 minute. The partially trained SSDO model was not able to provide better accuracy compared to the unsupervised models but provided in the best recall values.

5.2. Models Evaluation on Different Context

During this experiment, we tested the performance of the models in different contexts. In the graph, not contextual data indicates that for training and testing, all data (from Monday to Sunday) was taken, while contextual data indicates that during different testing stages working days (from Monday to Friday) and weekend days (Saturday and Sunday) were taken, and the resulting estimates were summed and divided by two.

The graphs on the *x* axis indicate whether the result is contextual or not, and on the *y* axis is displayed model's performance for the specified criterion.

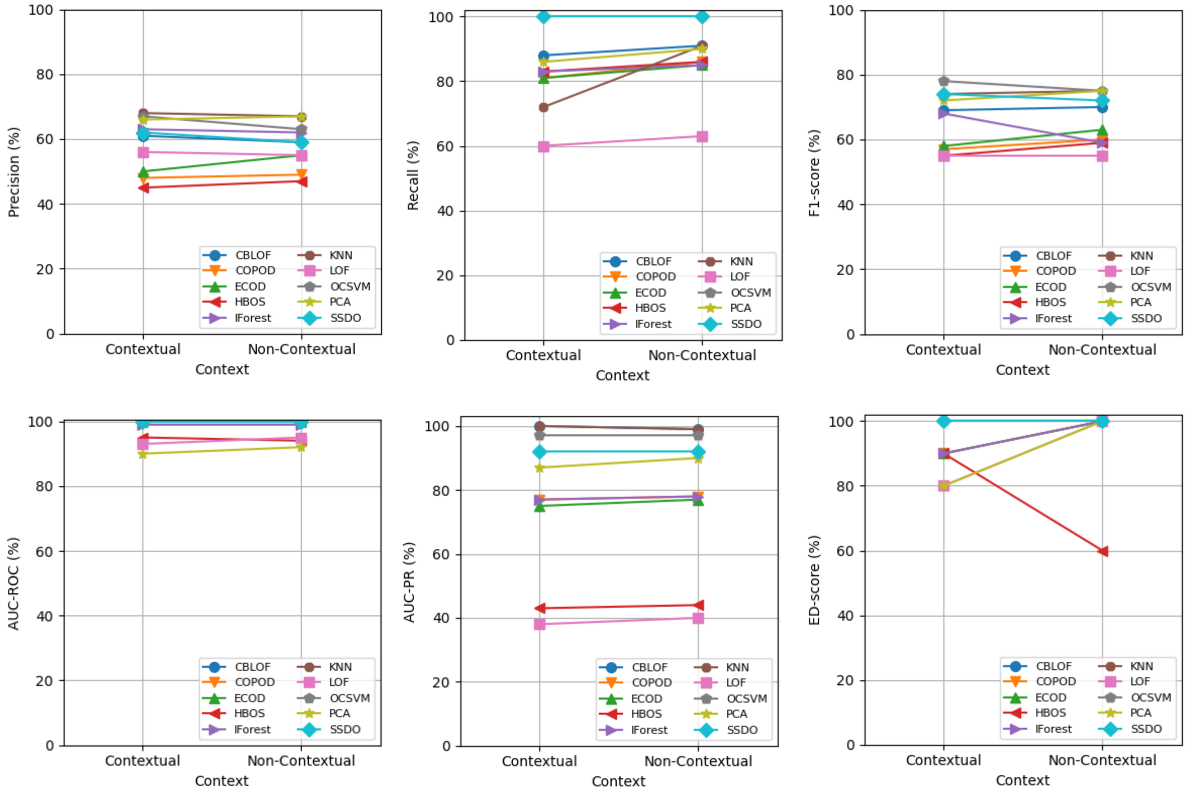

Figure 3: Comparison of models in different contexts

This experiment showed that the use of contextual data reduces the precision and improves the recall only marginally. However, when using non-contextual data, most models were able to detect water leakage much faster than when using contextual data.

5.3. Models Evaluation on Different Data Dimensions

During this experiment, we evaluated the performance of the models on different data dimensions. In the graphs, the dimensions marked on the *x* axis are expressed as follows:

- One dimension includes the mean of the data.
- Two dimensions include the mean of the data and the longest water running period (in minutes).
- Three dimensions include the average of the data, the minimum and the maximum water consumption over a three-hour period.
- Four dimensions include the third dimension plus the longest water running period (in minutes).

On the *y* axis the performance of the model is displayed for the specified criterion.

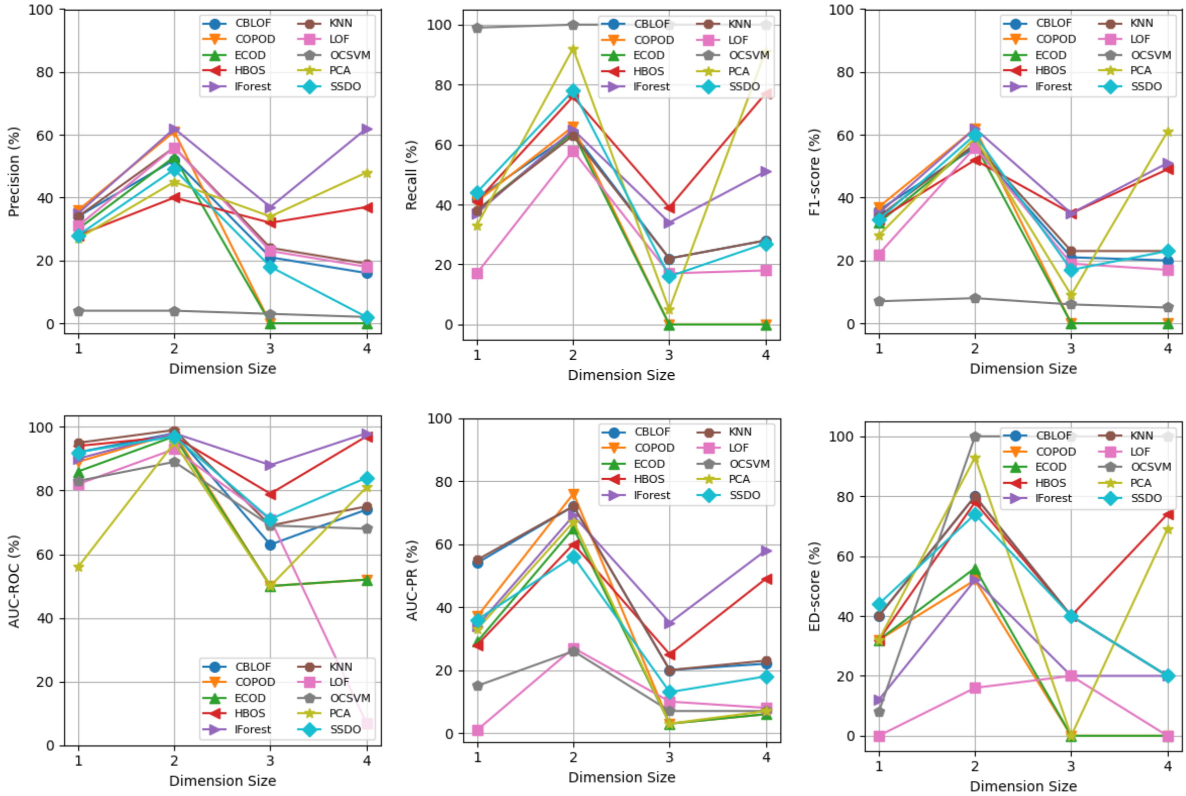

Figure 4: Comparison of models in different data dimensions

The experiment on different data dimensions showed that using two dimensions (average and longest water running period) gives the best overall results – the models were able to provide the best precision, recall and also provide the fastest water leakage detection rate.

6. Conclusion

In this work, we compared unsupervised CBLOF, COPOD, ECOD, HBOS, IForest, KNN, LOF, OCSVM, PCA and semi-supervised SSDO outlier detectors on water usage data and how they perform in detecting anomalous water usage. Experiments performed in different minute frames showed that when the system collects water consumption data every minute, all models perform much better with a smaller sliding window size, both in terms of precision and recall estimates. The results thus suggest that using the PCA outlier detector with a minute sliding window will be able to detect water leakage approximately 78% of the time, as well as to detect the onset of anomalous water use very quickly in order to stop unwanted consequences in time. Experiments performed at different levels of context showed that the difference between contextual and non-contextual data is not incredibly significant, but in most cases the models were able to detect leakage much faster using non-contextual data. Experiments performed using different dimension values indicated that using two dimensions (average water consumption and longest water running period) models were able to give the best results, and overall, the PCA model produced the best outcome, being able to detect water leakage 95% of the time and efficiently identify the start of the anomaly.

7. References

[1] UNESCO. The United Nations World Water Development Report 2021: Valuing Water, Mar 22, 2021.

[2] P. Burek, Y. Satoh, G. Fischer, M. T. Kahil, A. Scherzer, et al. Water Futures and Solutions: World Water Scenarios Report, 2016. URL: https://pure.iiasa.ac.at/id/eprint/13008/1/WP-16-006.pdf.

[3] WWAP (UNESCO World Water Assessment Programme). The United Nations World Water Development Report 2019: Leaving no One Behind, 2019.
[4] W. B. DeOreo, P. Mayer, B. Dziegielewski, J. Kiefer. Residential End Uses of Water, Version 2 Executive Report, 2016. URL: https://www.circleofblue.org/wp-content/uploads/2016/04/WRF_REU2016.pdf.
[5] Epa.Gov, Fix a Leak Week. URL: https://www.epa.gov/watersense/fix-leak-week.
[6] S. Han, X. Hu, H. Huang, M. Jiang, Y. Zhao. ADBench: Anomaly Detection Benchmark, 2022. doi:10.48550/arXiv.2206.09426.
[7] A. Boukerche, L. Zheng, O. Alfandi. Outlier Detection: Methods, Models, and Classification. ACM Computing Surveys, 2020, vol. 53, no. 3. doi:10.1145/3381028.
[8] H. Wang, M. J. Bah, M. Hammad. Progress in Outlier Detection Techniques: A Survey, 2019. doi:10.1109/ACCESS.2019.2932769.
[9] V. Chandola, A. Banerjee, V. Kumar. Anomaly Detection: A Survey. ACM Computing Survey, 2009, vol. 41, no. 3. doi:10.1145/1541880.1541882.
[10] G. O. Campos, A. Zimek, J. Sander, R. J. G. B. Campello, B. Micenková, et al. On the Evaluation of Unsupervised Outlier Detection: Measures, Datasets, and an Empirical Study. Data Mining and Knowledge Discovery, 2016, vol. 30, no. 4. pp. 891-927. doi:10.1007/s10618-015-0444-8.
[11] A. K. Sarangi. Smart Water Leakage and Theft Detection using IoT, 2020. doi:10.1109/I4Tech48345.2020.9102701.
[12] N. A. Moni, B. Sigweni, M. Mangwala, L. Kolobe. Water Leak Detection from Irrigation Pipelines in Botswana using Vibration Interpretation Technique, 2019. doi:10.1109/AFRICON46755.2019.9133829.
[13] A. Boudhaouia, P. Wira. Water Consumption Analysis for Real-Time Leakage Detection in the Context of a Smart Tertiary Building, 2018. doi:10.1109/ICASS.2018.8651976.
[14] D. J. A. Amora, M. A. V. Janapin, B. A. M. Calayag, C. L. P. P. Riofloriodo, N. M. Estur, et al. Design of a Household Consumption Based Water Leak Detection System Utilizing Machine Learning Algorithm, 2022. doi:10.1109/IET-ICETA56553.2022.9971564.
[15] H. Ismail, R. Elabyad, A. Dyab. Smart Residential Water Leak and Overuse Detection System using Machine Learning, 2022. doi: 10.1109/AICCSA56895.2022.10017508.
[16] M. Ji, G. Yi, J. Jung. Central Prediction System for Time Series Comparison and Analysis of Water Usage Data, 2020. doi:10.1109/ACCESS.2019.2963373.
[17] H. Fuentes, D. Mauricio. Smart Water Consumption Measurement System for Houses using IoT and Cloud Computing. Environmental Monitoring and Assessment, 2020, vol. 192, no. 9. pp. 602. doi:10.1007/s10661-020-08535-4.
[18] S. Patabendige, R. C. Oliver, R, Wang, W. Liu. Detection and Interpretation of Anomalous Water use for Non-Residential Customers. Environmental Modelling & Software, 2018, vol. 100. pp. 291-301. doi:10.1016/j.envsoft.2017.11.028.
[19] G. Pang, A. van den Hengel, C. Shen, L. Cao. Toward Deep Supervised Anomaly Detection: Reinforcement Learning from Partially Labeled Anomaly Data. Virtual Event, Singapore ed. New York, NY, USA: Association for Computing Machinery, 2021. doi:10.1145/3447548.3467417.
[20] A. B. García, A. Conde, U. Mori, J. A. Lozano. Water Leak Detection using Self-Supervised Time Series Classification. Information Sciences, 2021, vol. 574. pp. 528-541. doi:10.1016/j.ins.2021.06.015.
[21] W. Wang, P. Chen, Y. Xu, Z. He. Active-MTSAD: Multivariate Time Series Anomaly Detection with Active Learning, 2022. doi:10.1109/DSN53405.2022.00036.
[22] S. Das, W. K. Wong, T. Dietterich, A. Fern, A. Emmott. Discovering Anomalies by Incorporating Feedback from an Expert. ACM Transactions on Knowledge Discovery from Data, 2020, vol. 14, no. 4. doi:10.1145/3396608.
[23] Y. Zhu, K. Yang. Tripartite Active Learning for Interactive Anomaly Discovery, 2019. doi:10.1109/ACCESS.2019.2915388.
[24] V. Vercruyssen, W. Meert, G. Verbruggen, K. Maes, R. Bäumer. Semi-Supervised Anomaly Detection with an Application to Water Analytics. IEEE, 2018. doi:10.1109/ICDM.2018.00068.
[25] R. Zese, E. Bellodi, C. Luciani, S. Alvisi. Neural Network Techniques for Detecting Intra-Domestic Water Leaks of Different Magnitude, 2021. doi:10.1109/ACCESS.2021.3111113.

[26] X. Fan, X. Zhang, X. B. Yu. Machine Learning Model and Strategy for Fast and Accurate Detection of Leaks in Water Supply Network. Journal of Infrastructure Preservation and Resilience, 2021, vol. 2, no. 1. pp. 10. doi:10.1186/s43065-021-00021-6.
[27] S. Athanasiou, G. Giannopoulos, Y. Kouvaras, P. Chronis, G. Hatzigeorgakidis, et al. Trials Evaluation and Social Experiment Results, 2017. URL: http://www.daiad.eu/wp-content/uploads/2017/11/D7.3_Trials_Evaluation_v1.0.pdf.