

# Machine Learning-Based Approach for Household Water Consumption Prediction and Efficient Water Usage Recommendations

**Assistant Professor Smita Mangesh Junnarkar<sup>1</sup>**

Department of DataScience, SIES COLLEGE OF ARTS,  
SCIENCE AND COMMERCE (EMPOWERED  
AUTONOMOUS), Sion(W), Mumbai<sup>1</sup>

*Smita Mangesh Junnarkar, [smitaj@sies.edu.in](mailto:smitaj@sies.edu.in)*

**JAYAGURU NADAR<sup>2</sup>**

Department of DataScience, SIES COLLEGE OF ARTS,  
SCIENCE AND COMMERCE (EMPOWERED  
AUTONOMOUS), Sion(W), Mumbai<sup>2</sup>

*JAYAGURU NADAR, [gurunadar911@gmail.com](mailto:gurunadar911@gmail.com)*

---

## Abstract

Water use at home is closely connected with everyday activities, but the amount consumed can change considerably from one day to another. A family may use more water on a weekend, during hot weather, or on a day when laundry and cleaning are done. This makes household demand more difficult to estimate using a simple average. The study proposed here uses machine learning to learn these patterns from historical information and then turn the prediction into practical suggestions for saving water. The approach combines previous consumption with time-related, household, behavioural, and weather variables. Linear Regression, Random Forest, Support Vector Regression, and Gradient Boosting are considered so that a simple baseline can be compared with models that can learn more complicated relationships. Model performance is assessed using MAE, RMSE, and  $R^2$ . The recommendation part of the system is intended to give understandable suggestions, such as checking for possible leaks or reviewing unusually high-use activities. Earlier research has shown that previous consumption can be an important predictor of short-term demand, while household and temporal information can provide additional context [4, 6, 5]. Since an actual household dataset has not been supplied for this paper, experimental performance values are not invented; they should be added after the proposed framework is tested on real data.

**Keywords:** household water consumption; machine learning; demand prediction; smart meters; water conservation; recommendation system; sustainable water use.

## Title of the Study

**Machine Learning-Based Approach for Household Water Consumption Prediction and Efficient Water Usage Recommendations**

## Purpose of the Study

The purpose of this study is to design a machine-learning-based system that can estimate near-future household water consumption and use the prediction to provide practical, personalized recommendations for reducing avoidable water use. The central idea is to move from passive measurement—simply recording how much water a household has consumed—towards a decision-support approach in which consumption data are interpreted and converted into understandable actions.

## 1. Background and Context

Water is essential for drinking, food preparation, personal hygiene, cleaning, and many other domestic activities. The quantity of water required by a household is not fixed: it changes with access, reliability of supply, household circumstances, climate, and behaviour. The World Health Organization notes that household water quantity and accessibility are closely connected with health and well-being, and that water use can also change when supply is unreliable or affordability is limited [1]. This makes household consumption an important topic not only for environmental sustainability but also for practical resource planning.

Residential demand is not easy to predict because it comes from many small activities carried out throughout the day. A household may use more water because more people are at home, because laundry is done on a particular day, or because the weather is hot. A sudden increase can also be a sign of an unusual event, such as a leaking toilet or tap. The U.S. Environmental Protection Agency has noted the amount of water that can be lost through household leaks and the value of monitoring unusual flows [2, 3]. Smart meters make it possible to observe consumption at much shorter intervals than traditional monthly bills. Research using such data has found that previous demand, seasonality, weather, and household characteristics can help with forecasting [5]. Other studies have also tested machine-learning methods for daily residential demand and reported that nonlinear models can be useful when several factors interact [4, 6].

### 0.1. Why Machine Learning?

Traditional forecasting methods can be useful when demand follows a fairly regular pattern. Household behaviour, however, is rarely completely regular. Machine learning is useful for this type of problem because it can learn relationships between several variables at the same time. For example, the effect of temperature may be different for a large household than for a small household, and weekend behaviour may differ from weekday behaviour.

In this study, water consumption is treated as a supervised regression problem. The target is the amount of water expected to be used in a future period, while the input variables can include recent consumption, time information, household characteristics, and contextual factors such as weather or supply conditions.

Table 1: Proposed predictor groups

| Group      | Example variables                                    | Reason for inclusion                                    |
|------------|------------------------------------------------------|---------------------------------------------------------|
| Historical | previous day/week use, rolling average, lag values   | Captures persistence and recent household behaviour.    |
| Temporal   | hour, weekday, weekend, month, season                | Represents recurring routines and seasonal effects.     |
| Household  | occupants, dwelling type, rooms, appliance ownership | Represents the scale and characteristics of demand.     |
| Contextual | temperature, rainfall, supply reliability, holidays  | Captures external conditions that can change water use. |

### 0.2. Research Gap

Recent reviews show that machine-learning research on household water consumption is growing, but datasets, predictor sets, modelling choices, and evaluation practices remain heterogeneous [10]. There is also a practical gap between a prediction model and a useful household intervention. A model can predict demand accurately while still being unhelpful to a resident if its output is difficult to interpret. This study addresses that gap by combining prediction with a simple recommendation layer.

## 2. Purpose, Objectives and Research Questions

The study is built around a practical question: *Can household water consumption be predicted with reasonable accuracy using information that can realistically be collected, and can those predictions be turned into useful suggestions for saving water?*

### 0.3. 2.1 Objectives

The specific objectives are:

1. To examine the major variables associated with household water consumption.
2. To prepare household water-consumption data for machine-learning analysis.
3. To develop and compare multiple regression models for short-term consumption prediction.
4. To identify which input variables contribute most strongly to prediction.
5. To convert prediction outputs into simple, understandable water-efficiency recommendations.
6. To design an evaluation framework that balances predictive accuracy, interpretability, and practical usefulness.

### 0.4. 2.2 Research Questions

1. Which household, temporal, behavioural, and contextual variables are most useful for predicting household water consumption?
2. Which of the selected machine-learning models provides the best balance between accuracy and interpretability?
3. How much does adding historical consumption improve prediction compared with using household characteristics alone?
4. Can model outputs be translated into recommendations that are specific enough to be useful without making unsupported assumptions about household behaviour?

### 0.5. 2.3 Working Hypotheses

**H1:** Historical consumption features will provide substantial predictive value because household water use tends to show temporal persistence.

**H2:** Ensemble tree-based models will capture nonlinear relationships more effectively than a simple linear regression baseline when several household and contextual variables are combined.

**H3:** A recommendation layer based on predicted demand and interpretable contributing factors will be more actionable than reporting a predicted volume alone.

These hypotheses are intended to guide empirical testing rather than to predetermine the result.

### 0.6. 2.4 Significance of the Study

The proposed system has value at three levels. At the household level, it can provide a simple warning when predicted use is unusually high. At the community or utility level, aggregated forecasts could help identify periods of increased demand. At the academic level, the project demonstrates how a data-science pipeline can connect data preparation, exploratory analysis, supervised learning, model interpretation, and decision support.

The approach is also relevant to the Indian context, where water supply may be intermittent in some locations and households may rely on storage tanks or other adaptations. Recent work from India has explored lower-cost sensing approaches for water monitoring, including non-intrusive methods suitable for intermittently operated systems [12]. This suggests that a useful future direction is not only better algorithms but also affordable data-collection mechanisms.

Table 2: Research objectives and measurable outputs

| Objective              | Expected output                                             |
|------------------------|-------------------------------------------------------------|
| Understand consumption | EDA plots, descriptive statistics, correlation analysis.    |
| Prepare data           | Clean dataset and engineered features.                      |
| Compare models         | MAE, RMSE, $R^2$ , and validation results.                  |
| Explain predictions    | Feature importance or SHAP-style interpretation.            |
| Recommend actions      | Rule-based recommendation layer linked to predicted demand. |

### 3. Literature Review

Research on household water demand has increasingly moved from broad, aggregate estimates towards more detailed, household-level forecasting. This change has been encouraged by smart meters, which provide much more information about when and how water is being used. A review of household water-consumption determinants also shows that consumption is affected by a mixture of physical, social, behavioural, and environmental factors [7, ?].

Lee and Derrible compared twelve statistical and machine-learning approaches for predicting daily residential water use [4]. Their work is useful for this study because it shows that model selection should depend on the data and the prediction task rather than simply choosing the most complicated algorithm.

Pesantez, Berglund, and Kaza used smart-meter readings together with previous demand, seasonal information, weather, and household characteristics to forecast water demand [5]. Their comparison of Random Forest, Artificial Neural Network, and Support Vector Regression models demonstrates why several types of predictors can be considered together.

Xenochristou and colleagues developed an interpretable machine-learning approach for short-term household water-demand forecasting [6]. Their work is particularly relevant here because it considers not only predictive performance but also how the model can be understood. This supports the idea of adding an explanation and recommendation stage instead of treating the prediction as a black-box answer.

Other researchers have used high-resolution meter readings and unsupervised learning to study individual water-use events [8]. This shows that the same data can be useful for more than forecasting. They can also help identify unusual patterns that may point towards possible conservation opportunities.

Recent reviews show that machine learning is now being used with a wide range of water-consumption datasets. Tree-based methods are common for structured tabular data, while neural and recurrent approaches are often considered for sequential data. MAE and RMSE are also frequently used to evaluate forecasting models [10]. A separate review describes the growing use of artificial intelligence for water-consumption monitoring and assessment [9].

#### 0.7. Research Gap Identified from the Literature

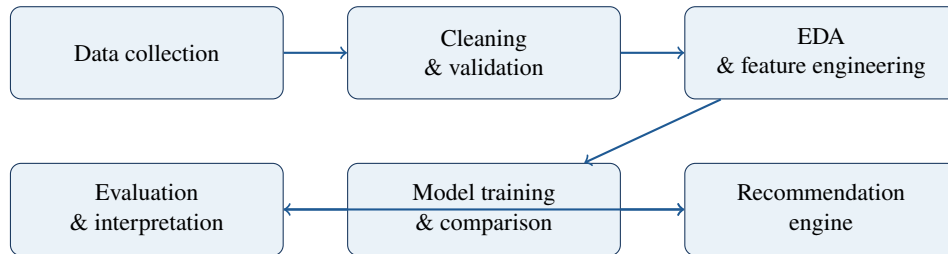
The literature points to three practical issues. First, the quality and amount of available data can vary considerably between studies. Some researchers have access to high-frequency smart-meter readings, while others have only billing or survey information. Second, a model can be accurate without giving a household any clear idea of what action to take. Third, household-level data raise questions about privacy and responsible use.

This study therefore focuses on a framework that can begin with a manageable set of variables and be expanded later. The aim is not to claim that one algorithm is best for every household. Instead, the study aims to show a clear connection between the data, the prediction, the explanation of that prediction, and a practical water-saving recommendation.

## 4. Methodology

### 0.8. 4.1 Research Design

The research uses a quantitative approach in which different machine-learning models are trained and compared. The work starts with collecting and cleaning the data, followed by exploratory analysis and feature engineering. The models are then trained and evaluated, and the final stage connects their output with practical water-use recommendations.



### 0.9. 4.2 Data Collection

For this project, the most useful dataset would contain household water readings collected daily or hourly over a reasonable period. If detailed smart-meter data are available, they can provide information about changes in usage during the day. A student project can begin with a smaller number of households and expand later. If original data cannot be collected, a suitable public residential water-demand dataset can be used after checking its licence and documentation.

Each observation should have a timestamp and a target consumption value. If household-level information is available, additional variables such as number of residents, dwelling type, water-storage capacity, and appliance ownership can be included. Weather variables such as temperature and rainfall can be joined using the appropriate date and location.

### 0.10. 4.3 Data Cleaning

The cleaning stage should include:

1. checking for missing timestamps and missing consumption values;
2. removing or investigating impossible negative readings;
3. identifying duplicated observations;
4. checking for sensor resets or meter rollover;
5. standardising units to litres or cubic metres;
6. detecting extreme observations using domain rules and robust statistical methods;
7. ensuring that training features do not accidentally use future information.

The final point is particularly important. If the goal is to predict tomorrow's consumption, the model must not receive tomorrow's actual consumption as an input. This type of information leakage can produce impressive but misleading performance.

### 0.11. 4.4 Exploratory Data Analysis

Exploratory data analysis will be carried out before modelling so that the general behaviour of the dataset is understood first. The analysis will look at daily and hourly consumption, weekday and weekend differences, monthly or seasonal changes, rolling averages, unusual observations, and possible relationships with temperature and rainfall.

Table 3: Example EDA questions

| Analysis             | Question addressed                                                       |
|----------------------|--------------------------------------------------------------------------|
| Daily trend          | Does consumption rise or fall over time?                                 |
| Hourly profile       | Which hours have the highest demand?                                     |
| Weekday comparison   | Are weekends different from working days?                                |
| Weather comparison   | Does temperature or rainfall coincide with changes?                      |
| Outlier analysis     | Which observations may represent leaks, meter errors, or unusual events? |
| Household comparison | Which household characteristics correspond to higher demand?             |

## 5. Feature Engineering and Model Development

### 0.12. 5.1 Feature Engineering

Raw meter readings do not always give a model enough information on their own, so additional features will be created. For a one-day-ahead prediction, useful examples are the previous day's consumption, consumption from the same day in the previous week, and short rolling averages. These features help the model recognise whether recent household use has been relatively high or low.

Temporal features can include hour of day, day of week, weekend indicator, month, and season. Cyclical encoding may be used for variables such as hour and month because 23:00 and 00:00 are close in time even though their numerical values are far apart.

For example, hour can be represented using:

$$x_{\sin} = \sin\left(\frac{2\pi h}{24}\right), \quad x_{\cos} = \cos\left(\frac{2\pi h}{24}\right)$$

where  $h$  is the hour of the day.

### 0.13. 5.2 Candidate Algorithms

Four regression models are proposed for comparison:

**Linear Regression:** This provides an interpretable baseline and establishes how well a simple linear relationship can predict consumption.

**Random Forest Regression:** Random Forest can represent nonlinear relationships and interactions without requiring a large number of manual transformations. It has also been used in residential water-demand forecasting research [6].

**Support Vector Regression:** SVR can model nonlinear relationships through kernel functions and is useful as a comparison model, especially when the dataset is moderate in size.

**Gradient Boosting / XGBoost:** Boosting builds a sequence of trees that progressively correct previous errors. It is included because structured household datasets often contain nonlinear interactions.

The final model will not be selected simply because it is more advanced. The choice will depend on its performance on unseen data, its consistency, and how easily its predictions can be explained.

Table 4: Model comparison plan

| Model             | Role                       | Strength / limitation                                               |
|-------------------|----------------------------|---------------------------------------------------------------------|
| Linear Regression | Baseline                   | Easy to explain; limited for nonlinear patterns.                    |
| Random Forest     | Main tree model            | Handles nonlinear interactions and mixed predictors well.           |
| SVR               | Nonlinear comparison       | Powerful on suitable scaled data; sensitive to parameter selection. |
| Gradient Boosting | High-performance candidate | Strong predictive capability; requires careful tuning.              |

### 0.14. 5.3 Training and Validation

Because water consumption is time-dependent, a random train-test split should not be the only validation method. A chronological split is preferred. For example, the earliest 70% of observations can be used for training, the next 15% for validation, and the final 15% for testing.

For longer datasets, rolling-origin or walk-forward validation can provide a more realistic estimate of forecasting performance. This prevents information from the future from influencing model development.

### 0.15. 5.4 Evaluation Metrics

Three primary metrics are recommended:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$
$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$
$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

MAE is easy to interpret in the original unit of consumption. RMSE gives greater weight to large errors, which is useful when unusually high demand is operationally important.  $R^2$  indicates how much variation is explained by the model but should not be used alone.

## 6. Efficient Water Usage Recommendation Framework

A prediction by itself may not be very useful to a household. For that reason, the proposed system includes a separate recommendation stage. Its purpose is to take the model’s output and turn it into simple advice that a user can understand and act on.

The recommendation stage will consider the predicted amount, the household’s recent usage, and signs of unusual behaviour. For example, repeated water flow during a period when the household normally uses very little water may justify a suggestion to check for a leak. The system will use a small number of clear rules instead of giving complicated technical messages.

### 0.16. 6.1 Recommendation Logic

A simple rule structure can be used initially:

Table 5: Proposed recommendation rules

| Observed condition                                | Interpretation                  | Suggested action                                                       |
|---------------------------------------------------|---------------------------------|------------------------------------------------------------------------|
| Predicted use substantially above recent baseline | Possible high-use day           | Review laundry, cleaning, bathing, or irrigation activities.           |
| Continuous night-time flow                        | Possible leak or unattended use | Check toilets, taps, storage tanks, and visible plumbing.              |
| Repeated weekend increase                         | Behavioural routine             | Compare weekend activities and shift non-essential use where possible. |
| High use during hot weather                       | Weather-related demand          | Review outdoor water use and avoid unnecessary irrigation.             |
| Repeated high use despite normal conditions       | Persistent pattern              | Conduct a household water-use audit.                                   |

The wording deliberately avoids claiming that a particular appliance or activity caused the consumption unless the data actually support that conclusion. A machine-learning model can identify associations and unusual patterns, but a recommendation should not be presented as a proven causal diagnosis.

### 0.17. 6.2 Personalisation

Personalisation can be introduced through a baseline calculated for each household:

$$Baseline_{h,t} = \text{median} (Consumption_{h,t-k}, \dots, Consumption_{h,t-1})$$

A relative deviation score can then be defined as:

$$Deviation = \frac{Predicted\ Consumption - Baseline}{Baseline} \times 100$$

For example, a household whose predicted consumption is 30% above its recent baseline could receive a “review high-use activities” message. The threshold should be tested empirically rather than treated as a universal rule.

### 0.18. 6.3 Explainability

For tree-based models, feature importance can provide a first-level explanation. More detailed methods such as SHAP values can show how individual features push a prediction higher or lower. This is valuable because a household user may reasonably ask why the system is recommending conservation.

Interpretability also helps identify unrealistic model behaviour. If the model assigns unusually high importance to an accidental identifier or a variable that would not be available at prediction time, the feature pipeline should be reviewed.

### 0.19. 6.4 Proposed System Output

A simple user-facing output could be:

**Predicted household consumption:** 540 L/day  
**Recent baseline:** 410 L/day  
**Expected change:** approximately 32% higher  
**Recommendation:** Review unusually high-use activities today. If there is continuous water flow when the household is not using water, inspect toilets, taps, and storage-tank connections for possible leakage.

The numerical values in this example are illustrative and are not presented as experimental results.

## 7. Key Findings and Discussion

At this stage, the study does not include invented model scores because an actual household water-consumption dataset has not been provided. The findings discussed here are therefore based on earlier research and on what the proposed experiment is designed to examine. Once real data are analysed, these statements should be replaced or updated with the measured results.

### 0.20. 7.1 Historical Consumption Is Expected to Be Highly Informative

Prior household water-demand research repeatedly shows the usefulness of previous consumption observations. Pesantez et al. used lagged demand together with seasonality, weather, and household characteristics, while Xenochristou et al. reported that past consumption can be a particularly important explanatory factor when it is available [5, 6]. This supports the inclusion of lag and rolling-window variables in the proposed framework.

The practical implication is that a household does not need a very large number of unrelated variables if it has a reasonably long and reliable history of meter readings. However, historical data alone may fail when unusual conditions occur, which is why contextual variables remain useful.

### 0.21. 7.2 Household and Temporal Variables Add Context

Household size, property characteristics, day of week, season, and supply conditions can change the meaning of a consumption value. Two households can use the same quantity of water for very different reasons. Research on household water-consumption determinants confirms that consumption is influenced by a broad combination of physical, social, behavioural, and environmental factors [?].

The expected outcome is therefore that a combined feature set will provide a more useful model than a single-variable baseline.

### 0.22. 7.3 Ensemble Models Are Strong Candidates, but Accuracy Is Not the Only Criterion

Tree-based ensembles are attractive because household datasets often contain nonlinear interactions. The recent literature review by Canquil et al. found frequent use of tree-based ensembles in heterogeneous tabular settings and recurrent models in sequence-based forecasting [10]. This does not mean that Random Forest or gradient boosting will always be the best model for a particular dataset. Model choice should remain empirical.

For a college-level implementation, a strong practical result would be a model that achieves lower MAE and RMSE than the baseline while remaining explainable enough to support recommendations.

Table 6: Expected model-comparison interpretation

| Observation                      | Likely interpretation                                  | Action                                   |
|----------------------------------|--------------------------------------------------------|------------------------------------------|
| Low MAE and RMSE                 | Predictions are close to actual use                    | Consider deployment for daily guidance.  |
| High $R^2$ but high RMSE         | Model explains broad variation but misses large events | Investigate peak-demand cases.           |
| Strong training, weak test score | Possible overfitting                                   | Simplify model or tune regularisation.   |
| Historical features dominate     | Demand is strongly persistent                          | Maintain reliable meter history.         |
| Weather adds little value        | Household routines may dominate                        | Avoid unnecessary weather complexity.    |
| Weather adds substantial value   | External conditions influence demand                   | Include weather in production forecasts. |

### 0.23. 7.4 Prediction and Recommendations Should Be Evaluated Separately

A prediction model can be accurate without producing useful advice. Conversely, simple recommendations can be sensible even when prediction error is moderate. Therefore, the proposed study evaluates two layers: forecasting quality and recommendation usefulness.

A future user study could measure whether households understand the recommendation, whether they consider it relevant, and whether behaviour changes after receiving it. This would provide evidence beyond conventional ML metrics.

## 8. Limitations and Future Research

### 0.24. 8.1 Limitations

One important limitation is access to suitable data. Detailed household water records are not available everywhere, and smart-meter data can be difficult to obtain because of cost, availability, and privacy concerns. A model trained using households from one area may also perform differently in another area.

The second limitation is measurement quality. Smart-meter readings can contain missing values, sensor errors, communication gaps, or changes caused by meter replacement. These issues must be handled carefully during preprocessing.

The third limitation is that prediction does not establish causation. If high temperature and high consumption occur together, the model can learn their association, but it cannot by itself prove that temperature caused the increase. Recommendations must therefore be phrased cautiously.

The fourth limitation is that household behaviour changes over time. A model trained on historical routines may become less accurate when household occupancy, appliances, tariffs, or water-supply schedules change.

Finally, this paper does not claim experimental model-performance numbers because an actual household water dataset has not been attached to the study. Reporting invented accuracy, RMSE, or  $R^2$  values would make the research less credible. The methodology is intentionally structured so that genuine results can be added after the experiment is performed.

### 0.25. 8.2 Future Research

Future research can extend the framework in several directions.

**Larger and more diverse datasets.** A multi-household dataset from different neighbourhoods and seasons would allow stronger generalisation tests.

**Higher-frequency forecasting.** Instead of daily prediction, the system could predict hourly or even shorter intervals. Research has already demonstrated the value of smart-meter data for user-level hourly forecasting [5].

**Leak and anomaly detection.** Unsupervised algorithms such as DBSCAN, Isolation Forest, or clustering could identify unusual consumption without requiring labelled leak events. Such approaches have been explored for abnormal water-use detection [11].

**Affordable sensing.** In settings where conventional smart meters are difficult to deploy, low-cost sensing technologies could be integrated with the prediction framework. Recent Indian research has investigated non-intrusive water monitoring using tank-level information, which may be particularly relevant to intermittently supplied systems [12].

**Explainable AI.** SHAP, permutation importance, accumulated local effects, and individual conditional expectation plots could be compared to determine which explanation method is easiest for non-technical users to understand.

**Personalised interventions.** Rather than sending the same conservation message to everyone, recommendations could be selected according to a household's historical pattern. For example, one household might receive a leak-check suggestion while another receives advice about irrigation timing.

**Privacy-preserving learning.** Future systems could investigate federated learning or on-device prediction so that detailed household records do not have to be centrally collected. This would become increasingly important as residential sensing becomes more widespread.

### 0.26. 8.3 Expected Contribution

The main idea of this project is to bring prediction, explanation, and practical action together in one framework. Instead of stopping after comparing machine-learning algorithms, the study asks what can actually be done with the prediction. This makes the project more useful from a household water-management point of view while keeping the methodology manageable for a student research project.

## 9. Conclusion

Household water consumption is affected by several factors, including previous usage, household characteristics, daily routines, weather, and water-supply conditions. Because these factors can interact, machine learning is a suitable approach when enough historical data are available.

This paper proposes a framework for predicting household water consumption and then using the prediction to provide simple water-efficiency recommendations. The workflow includes data cleaning, exploratory analysis, feature engineering, chronological validation, comparison of regression models, and model interpretation. A separate recommendation layer is used to turn unusual or higher-than-expected demand into practical suggestions.

The literature reviewed in this paper supports the importance of previous consumption and also shows that household, temporal, and contextual variables can add useful information [4, 5, 6, 10]. At the same time, a prediction is not automatically a useful recommendation. The next part of the work is therefore to test the framework on a real dataset and use the measured MAE, RMSE, and  $R^2$  values to decide which model is most suitable.

The system should be viewed as a decision-support tool rather than as a replacement for a professional plumbing inspection or water-utility planning. Its main purpose is to help households notice patterns in their consumption and make better-informed choices. Once a real dataset is analysed, the paper can be extended with genuine experimental results, graphs, and a final model comparison.

## References

## References

- [1] World Health Organization. *Domestic Water Quantity, Service Level and Health*, 2nd ed., WHO, 2020. Available at: <https://www.who.int/publications/i/item/9789240015241>
- [2] U.S. Environmental Protection Agency. *WaterSense: Statistics and Facts*. EPA, accessed 2026. Available at: <https://www.epa.gov/watersense/statistics-and-facts>
- [3] U.S. Environmental Protection Agency. *WaterSense Labeled Homes Technical Reference Manual*. EPA-832-F-23-016, 2023. Available at: <https://www.epa.gov/system/files/documents/2023-08/ws-homes-TechnicalReferenceManual.pdf>
- [4] D. Lee and S. Derrible, “Predicting Residential Water Demand with Machine-Based Statistical Learning,” *Journal of Water Resources Planning and Management*, vol. 146, no. 1, 2020. doi:10.1061/(ASCE)WR.1943-5452.0001119.
- [5] J. E. Pesantez, E. Z. Berglund, and N. Kaza, “Smart meters data for modeling and forecasting water demand at the user-level,” *Environmental Modelling & Software*, vol. 125, 104633, 2020. doi:10.1016/j.envsoft.2020.104633.
- [6] M. Xenochristou, C. Hutton, J. Hofman, and Z. Kapelan, “Short-Term Forecasting of Household Water Demand in the UK Using an Interpretable Machine-Learning Approach,” *Journal of Water Resources Planning and Management*, vol. 147, no. 4, 2021. doi:10.1061/(ASCE)WR.1943-5452.0001325.
- [7] A. Cominola et al., “Data Mining to Uncover Heterogeneous Water Use Behaviors From Smart Meter Data,” *Water Resources Research*, vol. 55, no. 11, pp. 9315–9333, 2019. doi:10.1029/2019WR024897.
- [8] G. M. Bethke, A. R. Cohen, and A. S. Stillwell, “Emerging investigator series: disaggregating residential sector high-resolution smart water meter data into appliance end-uses with unsupervised machine learning,” *Environmental Science: Water Research & Technology*, vol. 7, pp. 487–500, 2021. doi:10.1039/D0EW00724B.
- [9] A. Morain, N. Ilangoan, C. Delhom, et al., “Artificial Intelligence for Water Consumption Assessment: State of the Art Review,” *Water Resources Management*, vol. 38, pp. 3113–3134, 2024. doi:10.1007/s11269-024-03823-x.
- [10] D. Canquil, D. San Martín, I. Inoquio-Renteria, and P. Leger, “A Systematic Literature Review on Machine Learning Techniques for Predicting Household Water Consumption,” *Water Resources Management*, vol. 40, no. 2, pp. 1–30, 2026. doi:10.1007/s11269-026-04496-4.
- [11] H. Ghamkhar et al., “An unsupervised method to exploit low-resolution water meter data for detecting end-users with abnormal consumption,” *Sustainable Cities and Society*, vol. 94, 104516, 2023. doi:10.1016/j.scs.2023.104516.
- [12] M. Jamadarkhani, R. Raphael, S. H. P. Ramprasad, H. Babu, and S. Narasimhan, “IoT enabled smart water metering using multi sensor data and machine learning techniques,” *Frontiers in Water*, vol. 7, 2025. doi:10.3389/frwa.2025.1586916.