Journal Publication of International Research for Engineering and Management (JOIREM)
Volume: 04 Issue: 07 | July-2026
ISSN (O) 3107-6696

AquaLens: A Machine-Learning Decision-Support System for Industrial Water Footprint Reduction

Manjula Devi K N

Department of Computer Science and Engineering

Abstract - Industrial water systems require operational decisions that balance production, cooling constraints, leak response, and reuse potential. This paper presents AquaLens, an end-to-end machine-learning decision-support prototype for reducing the water footprint of a simulated industrial facility. The system combines physics-informed synthetic telemetry, feature engineering for time-aware forecasting, histogram-based gradient boosting models, Isolation Forest-assisted anomaly detection, constrained cooling-setpoint optimization, reuse-quality prediction, and a React/FastAPI dashboard. Experimental results show one-hour demand forecasting with R2=0.898, MAE=2.26 m3/h, and 48.9% MAE improvement over persistence; leak detection achieves F1=0.910; cooling surrogates achieve water-flow R2=0.986; and sample recommendations estimate 308.9 m3/day (29.2%) savings for a recent facility window. The contribution is a transparent, reproducible, deployable architecture that translates multiple ML outputs into quantified operator actions.

***Keywords*--**industrial water management, machine learning, water demand forecasting, anomaly detection, cooling optimization, water reuse, FastAPI, React dashboard

1. Introduction

Freshwater security has become a strategic concern for governments, industries, and communities. The United Nations Sustainable Development Goal 6 calls for sustainable water and sanitation management, including improvements in water quality, recycling, reuse, and water-use efficiency. UN-Water also emphasizes that sustainable and equitable water management supports prosperity, stability, and resilience, while scarcity and pollution can undermine livelihoods and increase operational risk. Industrial facilities are important participants in this challenge because of processing water, cooling water, cleaning cycles, leaks, and effluent reuse decisions directly influence both environmental impact and production cost.

AquaLens addresses this problem as a machine-learning decision-support system. Rather than treating water conservation as a single prediction task, it frames industrial water reduction as a coordinated workflow: forecast demand, detect abnormal consumption, optimize cooling operation, estimate reuse quality, and convert model outputs into ranked recommendations.

2. Objectives of the Work

The primary objective is to design and implement a reproducible end-to-end ML prototype that identifies practical water-saving actions for an industrial facility. The specific objectives are:

- Generate realistic synthetic hourly telemetry for multiple industrial facilities without using confidential plant data.
- Develop a one-hour-ahead water-demand forecasting model and compare it against a persistence baseline.
- Detect leak or anomaly periods using forecast residuals and an unsupervised anomaly model.
- Recommend cooling-tower setpoints that reduce make-up water while respecting a process-temperature safety constraint.
- Predict recycled-stream quality and route water to the highest feasible reuse tier.
- Integrate model results into a FastAPI service and a React dashboard for operator-facing exploration.
- Evaluate the system using quantitative model metrics and actionable water-saving estimates.

3. Literature Review

Water-demand forecasting is a long-standing decision-support problem. Recent work shows that machine-learning and statistical-learning algorithms can be automated and applied at scale for urban water-demand forecasting, including probabilistic and quantile-regression approaches. Although AquaLens targets industrial rather than municipal demand, the same methodological idea applies: operational time series become more useful when prediction is paired with exogenous drivers such as production, calendar effects, and weather-like conditions.

Gradient boosting is widely used for structured tabular regression because it captures nonlinear feature interactions while remaining easier to deploy than many sequence-heavy deep-learning models. AquaLens uses scikit-learn's

Journal Publication of International Research for Engineering and Management (JOIREM)
Volume: 04 Issue: 07 | July-2026
ISSN (O) 3107-6696

HistGradientBoostingRegressor, which is documented as a histogram-based gradient boosting regression tree estimator and is faster than the exact GradientBoostingRegressor on large datasets; it also supports missing values natively.

Leak and abnormal-use detection aligns with anomaly detection research. Isolation Forest identifies unusual observations by random partitioning; anomalies tend to require shorter path lengths to isolate than normal points. The original Isolation Forest papers introduced the approach for efficient anomaly detection. AquaLens extends this idea pragmatically by combining Isolation Forest with positive forecast residuals.

Water reuse and operational optimization are central because conservation depends not only on prediction accuracy but also on translating predictions into safe actions. SDG 6 explicitly includes recycling, safe reuse, and increased water-use efficiency across sectors. AquaLens therefore uses reuse-quality tiering and constrained cooling optimization to convert model outputs into operator recommendations instead of presenting raw predictions alone.

Area	Paper / Source	How it supports AquaLens
Water demand forecasting	Papacharalampous & Langousis, *Probabilistic water demand forecasting using quantile regression algorithms*, 2021	Supports the forecasting module and use of ML/statistical learning for water-demand prediction.
ML in water systems	Vaquet et al., *Challenges, Methods, Data: A Survey of Machine Learning in Water Distribution Networks*, 2024	Good survey source for ML tasks in water networks, especially leakage detection and sensor-driven water management.
Leak/anomaly detection	Liu, Ting & Zhou, *Isolation Forest*, IEEE ICDM, 2008	Supports AquaLens' Isolation Forest-based anomaly/leak detection.
Anomaly detection overview	Chandola, Banerjee & Kumar, *Anomaly Detection: A Survey*, ACM Computing Surveys, 2009	General theoretical base for anomaly detection methods.
Water leakage ML	Vaquet, Hinder & Hammer, *Investigating the Suitability of Concept Drift Detection for Detecting Leakages in Water Distribution Networks*, 2024	Relevant to leak detection in water systems and temporal behavior changes.
Water resources ML	Willard et al., *Time Series Predictions in Unmonitored Sites: A Survey of Machine Learning Techniques in Water Resources*, 2023	Supports ML-based prediction where monitoring data is limited, matching AquaLens' synthetic-data motivation.
Water quality prediction	Grbcic et al., *Coastal water quality prediction based on machine learning with feature interpretation and spatio-temporal analysis*, 2021	Useful for reuse-quality prediction and feature interpretation in water-quality ML.
ML model basis	scikit-learn documentation, *HistGradientBoostingRegressor* and *IsolationForest*	Supports the implementation choices used in AquaLens.

4. Methodology

4.1 Overall Architecture

AquaLens follows three-tier architecture. The data/model layer generates synthetic telemetry and stores trained scikit-learn models. The application layer exposes model capabilities through FastAPI endpoints. The presentation layer is a React and Vite dashboard that visualizes metrics, time-series, anomalies, recommendations, and what-if controls. FastAPI is a modern Python API framework based on standard Python type hints, with automatic interactive documentation support. Vite provides a development server and production build workflow for modern frontend projects.

Table 1. AquaLens system layers

Layer	Technology	Role
Data and model	Python, pandas, scikit-learn, joblib	Simulation, feature engineering, model training, persistence
Application	FastAPI, Uvicorn	REST endpoints for forecasting, cooling, reuse, metrics, dashboard data
Presentation	React, Vite, Recharts	Interactive dashboard, charts, sliders, operator recommendations

4.2 Data Generation and Feature Engineering

The dataset is generated by a physics-lite simulator for three facilities over 120 days at hourly resolution. The simulator produces production rate, ambient temperature, humidity, cooling setpoint, process water, cooling water, cleaning water, total water use, leak labels, and recycled-water quality. Leak events are injected with known ground truth so that anomaly detection can be quantitatively evaluated. Features include hour, day-of-week, weekend flag, production, ambient conditions, cooling setpoint, cyclical hour encodings, lagged water use, and rolling means. Lag and rolling features are shifted to prevent target leakage.

4.3 Model Design

The system contains five coordinated ML capabilities: demand forecasting, anomaly detection, cooling optimization, reuse-quality prediction, and recommendation ranking. The regression tasks use histogram-based gradient boosting. The anomaly detector combines a residual baseline with Isolation Forest. Cooling optimization uses surrogate models and a constrained grid search across possible setpoints. The reuse model predicts quality and maps it to the highest qualifying reuse tier.

Table 2. Model capabilities and evaluation focus

Capability	Model / Logic	Evaluation
Demand forecasting	HistGradientBoostingRegressor with lag, rolling, calendar, production, ambient features	R2, MAE, MAPE, baseline improvement
Leak detection	Forecast residual threshold plus Isolation Forest	Precision, recall, F1
Cooling optimization	Water and temperature surrogate regressors plus constrained setpoint grid search	Surrogate R2, MAE, feasible saving
Reuse routing	Quality regressor mapped to reuse tiers	R2, MAE, tier-routing accuracy
Recommendations	Ranked aggregation of reuse, cooling, and leak saving estimates	m3/day and percentage saving

5. Implementation Details

The project is implemented as a Python package named aqualens with modules for data generation, feature engineering, forecasting, anomaly detection, cooling optimization, reuse prediction, recommendation building, pipeline execution, and API serving. The training pipeline generates telemetry, trains models, writes diagnostic plots, persists model artifacts under artifacts/models, and saves consolidated metrics in artifacts/metrics.json.

The FastAPI backend exposes health, metrics, forecast, cooling optimization, reuse-quality, facility-listing, and dashboard endpoints. The frontend is implemented in React and consumes backend routes through a Vite proxy from /api to http://127.0.0.1:8000. Tests cover data generation, model capability thresholds, and API smoke behavior.

Table 3. Implementation modules

File / Module	Purpose
aqualens/data.py	Synthetic industrial water telemetry generator
aqualens/features.py	Shared feature engineering for training and serving
aqualens/forecasting.py	One-hour water-demand forecasting
aqualens/anomaly.py	Residual and Isolation Forest leak detection

Journal Publication of International Research for Engineering and Management (JOIREM)
Volume: 04 Issue: 07 | July-2026
ISSN (O) 3107-6696

aqualens/cooling.py	Cooling surrogate models and setpoint optimizer
aqualens/reuse.py	Reuse-quality prediction and tier mapping
aqualens/recommend.py	Ranked water-saving recommendation engine
aqualens/api.py	FastAPI service layer
frontend/src	React dashboard components and API wrapper

6. Results and Analysis

The final pipeline metrics show that AquaLens performs strongly on the main operational tasks. Forecasting achieves high explanatory power and a substantial improvement over a simple persistence baseline. Leak detection balances precision and recall, indicating that the hybrid residual-plus-Isolation-Forest design can identify most injected leak hours while limiting false alarms. Cooling surrogates are highly accurate, supporting their use in constrained setpoint search. Reuse-quality prediction is the weakest model by R2, but its tier-routing accuracy remains useful for recommendation-level decision support.

Table 4. Experimental results

Task	Metric	Value	Interpretation
Forecasting	R2	0.898	Strong fit for hourly demand
Forecasting	MAE	2.26 m3/h	Low average absolute error
Forecasting	MAE improvement	48.9%	Better than persistence baseline
Anomaly	Precision	0.922	Most flagged hours are true leak hours
Anomaly	Recall	0.899	Most true leak hours are detected
Anomaly	F1	0.910	Balanced leak-detection performance
Cooling	Water model R2	0.986	Reliable cooling-flow surrogate
Cooling	Temperature model R2	0.997	Reliable safety-constraint surrogate
Reuse	Tier accuracy	0.727	Useful but improvable routing model

For a sample recent 168-hour window at FAC-1, the system estimates current water use of 1057.7 m3/day and total potential saving of 308.9 m3/day, equivalent to 29.2% of daily use. The largest estimated saving comes from reuse routing, followed by cooling-setpoint optimization and leak inspection.

Table 5. Sample ranked recommendations

Recommendation	Saving (m3/day)	Share	Basis
Route recycled water to highest-qualifying reuse tier	213.60	20.2%	reuse capture 60% x predicted tier value
Optimize cooling-tower setpoint	88.42	8.4%	surrogate model over 168 h
Investigate flagged leaks / anomaly hours	6.85	0.6%	sum of positive forecast residuals on flagged hours

The top features also support the modeling assumptions: recent water use, cyclical time-of-day effects, production rate, 24-hour lag, and ambient temperature appear among the leading drivers. This is consistent with industrial operations where water use is shaped by recent demand, shift rhythm, production load, and cooling conditions.

7. Research Contribution

The main research contribution is not a single novel algorithm, but an integrated, transparent, and reproducible architecture for industrial water-footprint reduction. AquaLens contributes the following:

- **A multi-capability ML workflow that connects forecasting, anomaly detection, optimization, reuse routing, and recommendations.**
- **A reproducible physics-informed simulator that enables experimentation without exposing proprietary industrial telemetry.**
- **A hybrid leak-detection method that combines unexpected excess consumption with unsupervised anomaly scoring.**
- **A constrained cooling optimizer that prioritizes auditability through explicit surrogate predictions and grid search.**
- **A decision-support dashboard that translates model outputs into operator-readable m3/day and percentage savings.**

8. Conclusion and Future Work

AquaLens demonstrates that machine learning can support industrial water conservation when predictions are connected to operational actions. The prototype achieves strong forecasting, robust leak-detection performance, accurate cooling surrogates, and quantified recommendations. Dashboard-oriented deployment makes the work usable as a decision-support system rather than a collection of offline experiments.

Future work should focus on replacing synthetic telemetry with real plant data, adding uncertainty intervals to predictions, improving reuse-quality modeling with chemistry-specific features, validating savings estimates in live operations, adding model explainability for each recommendation, and expanding the optimizer to include energy-water trade-offs. A production system should also include authentication, data-quality monitoring, alert workflows, and periodic model retraining.

References

[1] United Nations Department of Economic and Social Affairs, Sustainable Development Goal 6: Ensure availability and sustainable management of water and sanitation for all, https://sdgs.un.org/goals/goal6.

[2] UN-Water, The United Nations World Water Development Report 2024: Water for Prosperity and Peace, 2024, https://www.unwater.org/publications/un-world-water-development-report-2024.

[3] G. Papacharalampous and A. Langousis, Probabilistic water demand forecasting using quantile regression algorithms, arXiv:2104.07985, 2021, https://arxiv.org/abs/2104.07985.

[4] scikit-learn developers, HistGradientBoostingRegressor documentation, https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.HistGradientBoostingRegressor.html.

[5] scikit-learn developers, IsolationForest documentation, https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.IsolationForest.html.

[6] F. T. Liu, K. M. Ting, and Z.-H. Zhou, Isolation Forest, 2008 Eighth IEEE International Conference on Data Mining, pp. 413-422, 2008.

[7] F. T. Liu, K. M. Ting, and Z.-H. Zhou, Isolation-Based Anomaly Detection, ACM Transactions on Knowledge Discovery from Data, vol. 6, no. 1, pp. 1-39, 2012.

[8] FastAPI project, FastAPI documentation, https://fastapi.tiangolo.com/.

[9] Vite project, Vite Guide: Getting Started, https://vite.dev/guide/.