

METRO FLOW-CITY LEVEL WATER CONSUMPTION FORECASTING USING ML AND CLIMATE INDICATOR

1st V.M.Manju Mithraa

Department Of Advanced Computing And Analytics
Vels Institute Of Science, Technology & Advanced Studies
Chennai, Tamil Nadu

2nd Dr.Vidhya Sathish

Department Of Advanced Computing And Analytics
Vels Institute Of Science, Technology & Advanced Studies
Chennai, Tamil Nadu

Abstract - The rapid rise of the world's urban populations and the change in climate are creating barriers to properly managing water resources in many cities. Therefore, if we want to plan cities in a sustainable manner, distribute water efficiently, and prevent future water shortages, we must precisely forecast future water demand in our cities. This paper describes a machine-learning-based water usage forecasting system called MetroFlow, which will forecast the amount of water consumed in a city by using a combination of historical water usage and climate factors. The model can improve its calculation in prediction by adding the environmental factors like rainfall, temperature, humidity, evaporation, and wind speed. The advanced feature engineering methods are used to analyze the short-term and long-term patterns of water consumption. The machine learning models involved in the project are Random Forest, XGBoost, Long Short Term Memory, these models are evaluated using regression metrics RMSE, MAE and MAPE. The machine learning model that provides the best predictions will be used to generate forecasts for the next 24 hours, 7 days, and 30 days. Additionally, an interactive visualization dashboard has been created using Streamlit to enable users to visually explore water consumption trends and view water usage forecasts. This proposed project demonstrates how machine learning techniques that consider climate will enhance forecasting of urban water demand and assist with smart city resources management.

Key Words: Urban Water Demand Forecasting, Machine Learning Models, Interactive Visualization Dashboard, Smart City Water Management

I. INTRODUCTION

Urban areas depend on water as a critical life-sustaining resource for domestic use, manufacturing, and agricultural production. Given the rapid population growth in urban cankers due to global urbanization rates, cities are facing increasing pressure to provide a secure and reliable supply of water. This has made it even more essential to develop accurate forecasts of future water demands as part of the effective management of our water resources. Many factors affect water use decision-making in urban cities, including population growth, seasonal fluctuations in demand patterns, and the various physical environmental elements. The main climates factors like rainfall and temperature have a very big impact on the water usage consumption in different seasons of the year. While demand for water is expected to be higher in the summer and it is reduced during winter seasons. These techniques may identify general trends but they lack the capacity to model the complex and non-linear relationships between environmental factors and water demand accurately. The advancement of machine learning methods continues to create opportunities for the development of models that can be used to analyze large datasets and to highlight previously undetected patterns in time-series data. Machine learning algorithms are capable of learning associations between historical water use and weather, so they should be able to have more accurate predictions based on these associations.

This study also outlines the new MetroFlow system, which uses machine learning to predict water demand at the city level from weather data. Multiple different machine learning models are incorporated into the system, which is displayed in an interactive format so that users can analyze predicted water usage and demand in different ways.

II LITERATURE SURVEY

Various studies have used different machine learning techniques for water demand prediction.

Study 1: Several researchers have also used traditional statistical models like ARIMA for predicting water consumption. These models are typically good predictors of patterns that are linear, but, they perform poorly at capturing nonlinear relationships between the weather and water use.

Study 2: To predict the water consumption and demand machine learning models like Random Forest and Support Vector Machines are used. These models capture the nonlinear relationships between predicted water consumption and historical data, but due to the sequential nature of time series forecasting they often fail to utilize sequential information.

Study 3: Deep Learning models have effective forecasting tools for time series data and identify the model longitudinally based on temporal relationships and long-term patterns between time based datasets.

Research Gap: Most studies have two different types of data: one is historical water consumption, and the other is climate variables. Very few studies have combined both data types to develop an interactive city-level water consumption forecasting system.

Solution: MetroFlow incorporates climate data and historical water consumption data into a hybrid machine learning type of framework and provides an interactive dashboard for future analysis.

III PROPOSED METHODOLOGY

Proposed Methodology: The proposed methodology for MetroFlow uses an intelligent predictive framework using historical consumption patterns and climate data to predict a city's consumption of water. In order to do this, the methodology goes

through several stages which include data collection, preprocessing data, feature engineering, training models, evaluating models, forecasting, and visualizing information.

Each of these stages contributes to improving the forecasting accuracy and creating a valid forecast.

A. Data Collection

The first stage in our methodology will be to assemble a set of requisite datasets needed to train the forecasting machine-learning models.

First Dataset: Historical Water Consumption Data - This dataset consists of historical water usage data for each city. Depending on the data source, this can consist of daily or hourly water used records and is expressed in either liters or cubic meters. The historical data provides the basis for the model to identify long-term trends in water demand.

Second Dataset: Climate Indicator Data - Climate conditions are generally major determinants of how much water will be used, and thus will be combined with the water consumption data. The main climate indicators that will be used for the project include:

- Rainfall
- Temperature
- Humidity
- Evaporation
- Wind Speed

These indicators establish a link between environmental conditions and the amount of water the consumer uses. In order to adequately analyze the climate conditions for the water consumption, both datasets will be merged using a common time or date. This means that in order to analyze the climate condition during a water consumption period, the time/date that are associated with the consumption will be used for that analysis.

B. Data Preprocessing

Collected datasets often have (data entries are) missing, incorrect, or unuseable (ineffective). Data cleaning is a necessary! Cleaning datasets will help individual datasets remove duplicates from the dataset, fix any errors in the records of collected data, and to remove null values from the datasets.

Dealing with Missing Records (Value(s)) - The filling of missing records (by) using a number of methods:

- Mean substitution
- Forward filling
- Interpolation

C. Feature Engineering

Feature engineering provides a way to derive meaningful and essential information from the initial dataset and is the second most crucial step in feature extraction.

Lag (previous water consumption values) - Lag features are created by taking the previous water consumption data and using it as a feature for predicting the current day's water consumption. Lag features will allow the capturing of the temporal dependency between the previous and current day's water consumption.

Moving average - A moving average will smooth the fluctuation in the water consumption patterns while also displaying long term trends in the water consumption data.

Climate anomaly - Climate anomaly features allow the identification of climate anomalies, such as excessive heat or extreme rainfall, which will greatly affect water demand.

E.g. - Day of the week, month, season, etc.

Temporal features allow a model to identify the day of the week and seasonal patterns of water consumption. The prediction performance of the model is improved by the original feature and newly created feature .

D. Model Training

After feature engineering is completed, the dataset will be split into 2 different datasets: a training dataset and a testing dataset. The training dataset will be used to train multiple machine learning models.

There were three different machine learning models implemented that are LSTM model, Random Forest model, and XGBoost.

LSTM Model:

This model is a deep learning it is designed for time series data which allows to record long term dependencies and sequential patters of water consumption. The LSTM model will also be beneficial in recognizing seasonality as well as temporal patterns.

Random Forest Model:

Random Forest is an ensemble machine learning algorithm that predicts outcomes using numerous decision trees. Additionally, Random Forests are effective at accurately capturing nonlinear relationships between climate factors and water consumption.

XGBoost Model:

XGBoost is a gradient boosting algorithm that has a reputation for high accuracy in predicting outcomes. It builds the decision tress sequentially by this each tree improves the error make by the previous tree and enhance the model performance.

Utilizing multiple models will enable the recognition of different types and patterns of data within the dataset which will improve the performance of MetroFlow.

E. Model Evaluation

After the step of machine learning models training using the historical water consumption data and climate factors, it is important to evaluate their performance to determine how accurately they can predict future water demand. The model evaluation helps to identify which model gives predictions that are very close to the actual observed values. In this project, three different models that are implemented (LSTM, Random Forest, and XGBoost). Their performance is compared using widely accepted evaluation metrics that is regression evaluation metrics.

Regression metrics measures the different between actual and predicted values , if the difference is small then it is a better model.

Root Mean Square Error (RMSE)

RMSE measures the average value of prediction errors. A Lower RMSE value means better model accuracy.

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

Where:

- y_i = actual value
- \hat{y}_i = predicted value
- n = number of observations

Mean Absolute Error (MAE)

MAE calculates the average absolute difference between actual and predicted values.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

Mean Absolute Percentage Error (MAPE)

MAPE measures the prediction error in percentage, which helps simplify the model performance easily.

$$MAPE = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right|$$

Where:

- n - number of data points,
- y_i - actual value
- \hat{y}_i - forecast value

All this metrics are used to compare the performance of the models and the model with less prediction error is selected for forecasting.

F. Forecast Generation

After the best performing model is selected, the system automatically gives water demand predictions for different time range.

The *Short-Term Forecast* predicts the water consumption for the next 24 hours. The *Medium-Term Forecast* predicts the water consumption for the next 7 days. The *Long-Term Forecast* predicts the water consumption for the next 30 days.

These forecasts help urban planners predict the future water demand and prepare for possible shortages of the water.

G. Visualization and User Interface

In this stage of the methodology, the final element is the forecasting results presented on a visually appealing and interactive dashboard, developed using frameworks such as Streamlit and Flask. This allows for visualisation of water consumption historical data trend and forecast results.

The users can:

- Visualize historical water consumption trends
- Investigate how climate conditions affect consumption
- ML models performance comparison.
- Visualize forecasted water demand
- Download forecast reports

This interactive visualization of forecasted results provides a useful tool for city water management authorities and policy-makers.

IV ARCHITECTURE DIAGRAM

The architecture of the MetroFlow project is a data driven model as it is designed to use machine learning and climate factors to predict the amount of water used in cities. First, data is acquired at the data acquisition stage. The acquisition of historical water usage records from multiple sources and weather data for cities will take place in this stage. The accuracy and consistency of the data are archived by passing the data to preprocessing stage that includes cleaning, normalizing, and integrating.

Next, will be feature engineering stage, to enhance accuracy of the model accuracy the relevant data are identified and used to create new features. Then the data is used to train the model, it uses the machine learning technique to find out the connection between climate and water demand. After training, the model will be put to use to predict water demand through its predictive stage.

The predicted water demand will then be shown visually to officials using visualization applications. Authorities can utilize the predicted water demand to maximize their water management and resource planning efforts.

The model will utilize regression techniques to assess the accuracy of predicted water demand. Also the model accuracy of predictions remains constant over time. This architecture creates a scalable, adaptable system that is capable of addressing changing climate conditions.

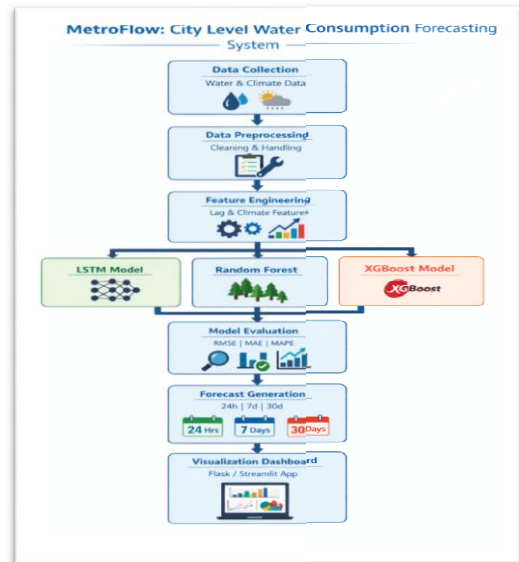


Fig 1: The Architecture diagram

V INPUT

MetroFlow's water consumption forecasting will stem from an input of two relevant historical datasets comprising of historical water consumption data and climate variable data. Both datasets must be entered into the MetroFlow system in CSV format or may be uploaded in a single CSV file.

The first historic water consumption dataset contains water consumption data for past uses of water. The first column of the dataset lists dates and times of days that the water was used and the second column list the amount of water consumed on specific dates. This first column of dates enables time-series analysis to be applied to the water demand history to identify water demand trends over time. The second dataset within the MetroFlow system consists of climatic variables that directly impact the demand for water. Weather factors significantly affect how individuals use water; hot weather leads to increased water consumption while rainy weather has the opposite impact.

The two datasets are combined based on date or date and time so that they can be used to develop the historical demand trends of the areas represented in each respective dataset. The input to the system consists of two main datasets.

Water Consumption Data

- Date
- Water usage
- Historical demand records

Climate Data

- Rainfall
- Temperature
- Humidity
- Wind speed
- Evaporation

These datasets are combined to create a complete dataset for training the models.

VI PSEUDOCODE

Run application server

BEGIN

Initializing the Flask and the database application
Set the upload configurations and allowed the file types
Creates the required folders

Initializing the service of data ingestion

Generates live water flow level
Determining the status of water flow anomaly
Displaying the homepage with water flow details

Get the forecast input from user
Convert input into number of days

Receive the uploaded CSV files
Filter valid CSV files
Save uploaded files

Process uploaded files
If invalid, load default dataset
Validate dataset
If invalid, return error

Generate forecast using dataset
Generate status of the trend data and water flow
Create visualization charts
Generate forecast CSV file

Selecting the best model and extract its accuracy
Displaying the results in charts and forecast details
Check if forecast file exists
Allow user to download CSV file

Get manual input values
Calculate base water usage
Apply climatic factors
Apply day and seasonal improvement
Compute final predicted usage
Displaying status of prediction result with water flow
Initialize database tables

END

VII OUTPUT

The MetroFlow system gives output as predicted water consumption values and visual analytical insights generated by the trained machine learning models. After processing the input datasets and selecting the best performing model the system produces forecasts for different time range that includes 24 hours, 7 days, and 30 days.

The predicted results is the estimated water demand for the future based on the historical data and climatic factors, these predictions helps city planners and water management authorities to understand the expected water usage trends and preparing appropriate distribution strategies.

In addition to numerical predictions, the system also gives the visual outputs through an interactive dashboard developed using visualization frameworks like Streamlit or Flask. The dashboard displays the graphical visualization that includes the historical water consumption data, climatic factors that influences and comparison charts between actual and predicted values. These visualizations helps users to simplify the forecasting results more effectively.

The system also helps the users to download prediction results in CSV format for further analysis or reporting purposes. Both the numerical forecasts and visualization of the MetroFlow system provides a complete output that supports the decision-making for efficient urban water resource management.

VIII RESULT

The MetroFlow system is evaluated to determine how effective the machine learning models are predicting the city level water usage using the historical water usage data and climate factors. After the preprocessing stage and feature engineering the dataset was divided into two sets one is training set and other is testing sets, and the training dataset is used to train multiple models like Long Short-Term Memory (LSTM), Random Forest, and XGBoost to the forecast water demand.

The regression metrics like RMSE, MAE, MAPE are used to measure the model performance by finding the difference between the predicted water consumption value and actual water usage values. From the implemented ML models XGboost and Random forest shows a better prediction as they records the nonlinear relationships between climate factors and water usage. Whereas the LSTM model has performed well only in identifying the temporal patterns and long term dependencies time series data.

The final results shows that the added climate factors like rainfall, temperature, humidity, evaporation, and wind speed had significantly improved the forecasting accuracy when it is compared to models which has to rely only on historical water consumption data. The system successfully generated the forecasts for different time range like 24 hours, 7 days, and 30 days helping both short-term operational planning and long-term resource management.

Visualization graphs like historical consumption trends, actual versus predicted demand curves, and forecast charts further demonstrate the accuracy and reliability of the system. Overall, the results show the MetroFlow platform can effectively support data driven water management and help urban planners in optimizing water distribution and improving sustainability.

IX SCREENSHOTS

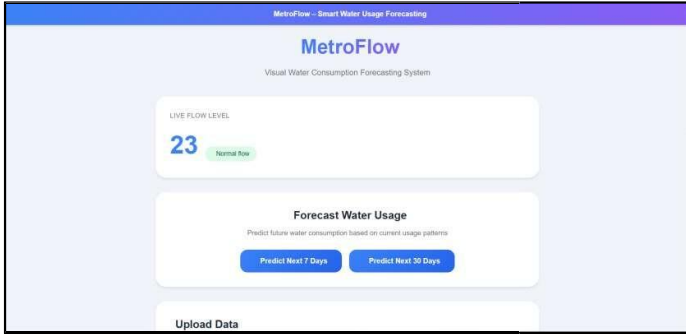


Fig 1: Home Page Displaying The Live MetroFlow Consumption

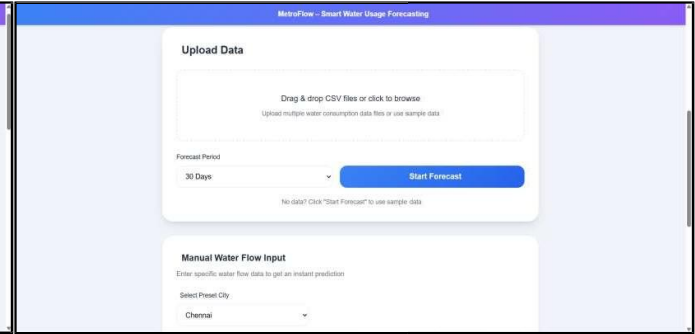


Fig 2: Input Interface To Upload Data Set

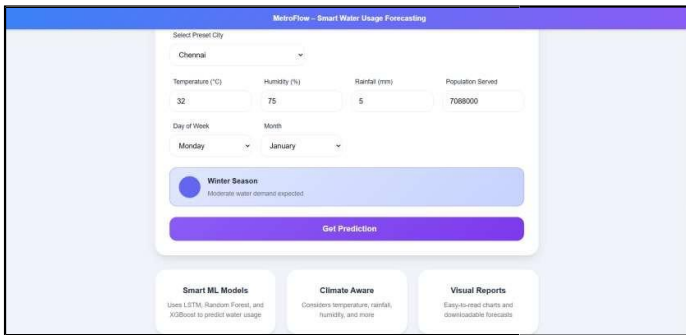


Fig 3: Input Interface Used For Entering User Data

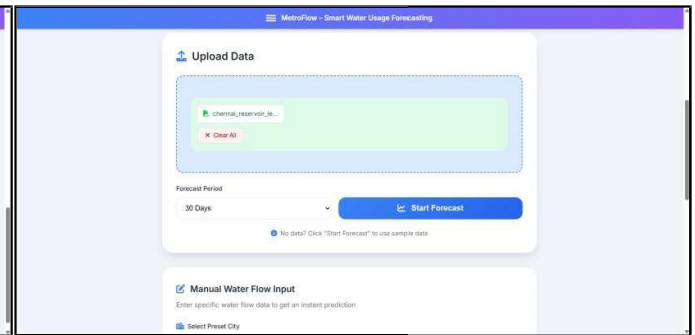


Fig 4: Web Page Of The Dataset Uploaded



Fig 5: Output Screen Displaying The Live Flow Level

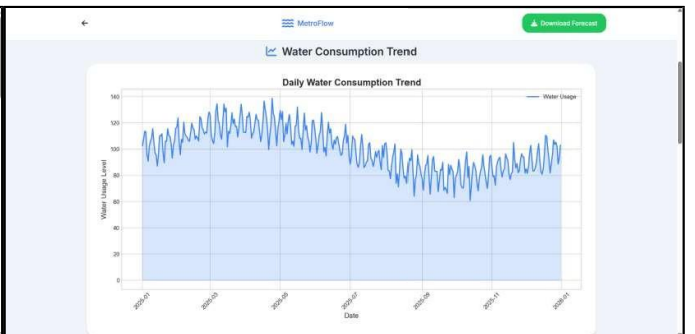


Fig 6: Output Screen Displaying The Daily Water Consumption Trend

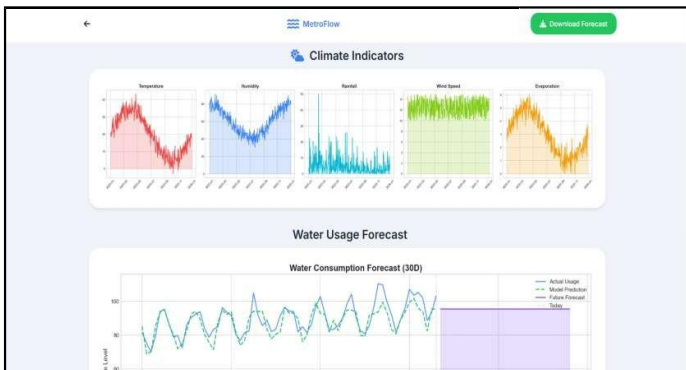


Fig 7: Output Screen Displaying The Climate Indicator

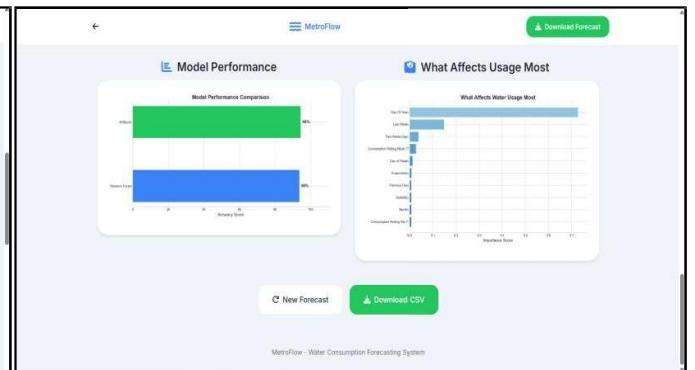


Fig 8: Output Screen Displaying The Model Performance

X CONCLUSION

Based on the results the MetroFlow system shows that it is an effective and reliable solution for forecasting city-level water consumption using machine learning techniques. The model's prediction accuracy is improved by using the historical water usage data together with the climate factors. Among the evaluated models the Random Forest and XGBoost model has showed the strong performance in capturing complex nonlinear relationships.

For the short term and long term resource planning the system has successfully generated accurate forecasts for multiple time range like 24 hours, 7 days, and 30 days. Evaluation metrics like RMSE, MAE, and MAPE ensures a good comparison of model performance and helps to select the most efficient model.

Furthermore, visualization of results using the graphs and charts has improved the interpretability of predictions and provided a meaningful insight of the water usage consumption patterns. Overall, the MetroFlow platform gives scalable and data driven approach to the urban water management that enables the authorities to optimize water distribution, reduce wastage, and support sustainable development.

XI REFERENCES

- [1] G. Liu, D. Savic, and G. Fu, "Short-Term Water Demand Forecasting Using Data-Centric Machine Learning Approaches," *Journal of Hydroinformatics*, vol. 25, no. 2, pp. 350–366, 2023.
- [2] M. Fathian, A. Minaei-Bidgoli, and H. Bagheri, "Developing an LSTM Model for Forecasting Monthly Water Consumption Based on Climatic Factors," *Environmental Research*, vol. 216, 2023.
- [3] J. Herrera, M. Torgo, J. Izquierdo, and R. Pérez-García, "Predictive Models for Forecasting Hourly Urban Water Demand," *Journal of Hydrology*, vol. 387, pp. 141–150, 2010.
- [4] Z. Zhang, Y. Li, and H. Wu, "Urban Water Demand Forecasting Using Deep Learning Models," *IEEE Access*, vol. 8, pp. 145939–145949, 2020.
- [5] S. Papacharalampous and A. Langousis, "Probabilistic Water Demand Forecasting Using Quantile Regression," *Water Resources Research*, vol. 57, no. 4, 2021.
- [6] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [7] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [8] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016.
- [9] Y. LeCun, Y. Bengio, and G. Hinton, "Deep Learning," *Nature*, vol. 521, pp. 436–444, 2015.
- [10] P. Herrera, R. Torgo, J. Izquierdo, and R. Pérez-García, "Urban Water Demand Forecasting: A Review of Methods and Models," *Water Resources Management*, vol. 24, pp. 1867–1884, 2010.