The Bioscan
AN INTERNATIONAL QUARTERLY JOURNAL OF LIFE SCIENCES
20(1): 445-454, 2025
www.thebioscan.com

From Dam to Tap in Mumbai: Real- Time Water Quality Monitoring using IOT and Machine Learning

[1] Pearl Dsouza, [2]Namrata Joshi,[3] Noelle Shaji, [4] Prof. Prachi Patil, [5] Prof. Monali Shetty

[1,2,3,] Student, Department of Computer Engineering, Fr. Conceicao Rodrigues College of Engineering. Mumbai [4,5,] Assistant Professor, Department of Computer Engineering, Fr. Conceicao Rodrigues College of Engineering. Mumbai

[1] crce.9538.ce@gmail.com, [2] crce.9545.ce@gmail.com, [3] crce.9577.ce@gmail.com, [4] prachip@fragnel.edu.in, [5] shettymonalin@gmail.com

DOI: https://doi.org/10.63001/tbs.2025.v20.i01.pp445-454

KEYWORDS
water quality monitoring,
public health,
and environmental sustainability,
machine learning algorithms,
IOT

Received on:

04-01-2025

Accepted on:

03-02-2025

Published on:

06-03-2025

ABSTRACT

This research explores water supply from dams to households at real-time monitoring of the quality of the water with IoT sensors and subsequent machine learning. It ensures that water supplied from dams to reservoirs and further to household service remains pure even after its supply. The methodology includes the collection of water at the Bhandup Complex for purification and its subsequent distribution to 26 service reservoirs throughout Mumbai. For training and testing different machine learning algorithms, namely SVM, Random Forest, and LightGBM, data from these reservoirs and purified water in the Bhandup Complex are used. LightGBM achieved high accuracy on the other algorithms and achieved an accuracy of 97.5% on purified water and 75.5% on the corresponding reservoir water. Real-time monitoring of water quality by IoT sensors in houses can give information on pH, turbidity, TDS, and many other parameters. It can help trace out specific areas of contamination in the water distribution network further. Major Findings for the present study comprised the reliability of AquaSage to predict water quality and the presence of potential points of pollution, thereby ensuring that the water supply from the treatment plant would be safe for consumption.

This technology brings to light the major life improvements for the citizens of public health through prompt intervention and the provision of timely water quality. Thanks to their features, one of which is the possibility of an early stage of knowing that contamination exists, they are also cutting the risk of waterborne diseases by limiting the growth of bacteria and such, so they are a great way to help sustainable water management also. On the other hand, the research shows some limitations. The data for dams and reservoirs, based on BMC's provided ranges, covers only five years and is static, not real-time. Efficiency of hardware sensors may decrease over time, and the newest scaling is possible only at households, but not in society or at building level. With this in mind, the major benefits of the study come in terms of creating an IoT infused machine learning algorithm for monitoring the projects in real-time and therefore using a user-friendly system that empowers citizens to ensure water safety. Practical implications involve eliminating the threats of waterborne illnesses in the city through giving the city dwellers an honest mechanism to avail clean water supply.

The real-time monitoring features of AquaSage also prove useful for precautionary measures against contamination and sustainable water management. Later studies should focus on refining precision in forecasting, additional sensor integration, extended performance testing, user interface optimization, and use cases in aquaculture and industrial on-site water quality monitoring. More scope for the improvement of water treatment processes include full system automation and integration into water purification efforts.

INTRODUCTION

Water is necessary to maintain ecosystems and, hence, human life. Worldwide, advanced systems are put in place to collect and filter water in order to meet the demands of industrial regions, rural villages, and urban municipalities. All these depend on lakes, rivers, and reservoirs as natural water sources for distribution through large pipelines and infrastructure for serving increasing population.

Yet, despite the sophistication at which they are built, water distribution systems have significant challenges that challenge their reliability and efficiency. Risks of poor water quality include old infrastructure, poor monitoring, and increased pollution. Recently, a report was published indicating that Mumbai has experienced serious cases of severe water contamination through pipelines that had not been adequately monitored (Report Cite). Such cases underscore the imperative to adopt effective water quality monitoring systems for the protection of public health and the environment's sustainability.

Current systems often fail to monitor the quality of water during transit, leaving users vulnerable to contamination before reaching the homes. This lack of transparency jeopardizes the integrity of the supply and poses risks to end-users. To fill this gap, AquaSage is proposed as an innovative solution, empowering each individual with real-time information on the water quality metrics. By allowing the users to track and evaluate the safety of their water, AquaSage helps to enhance public health outcomes and supports environmental conservation in a proactive, initiative-based approach.

1. Literature Review

[1] This paper presents the role of 14 different machine learning models in order to assess as well as predict the water quality. The model demonstrates the poteourntial to automate water quality assessment and enhance the efficiency of evaluating water quality parameters. The paper highlights that LightGBM, XGBoost, CatBoost, and Random Forest were the best-performing algorithms. XGBoost outperformed the other algorithms with an accuracy of 96.31%

[2] The purpose of this paper is to provide a robust statistically sound methodology for assessment of WQI model uncertainties. Various WQI models and methods were tested in this research paper MLP achieved the highest accuracy among the different models. The future scope of this paper includes application of the methodology to various types of water bodies such as rivers, lakes as well as lies in the continued refinement, validation, and application of the proposed methodology to address the uncertainties associated with water quality assessment across various environmental contexts and stakeholder needs.

[3] This article employs a classification model to classify samples of drinking water as safe or unsafe and to predict the quality index of the water. The results are based on the training and testing of nine unique machine learning models; namely Extreme Gradient Boosting (XGBoost), Light Gradient Boosting (LightGB), Decision Tree, Extra Tree classifier, MLP classifier, GB classifier, SVM, ANN, and RF classifier. All the machines were trained using data generated from 7996 water samples and 19 different features. The best model in terms of performance was LightGBM, which scored an accuracy rate of 97%.

[4] This study aims at filling in missing values with the precision of predicting water quality to a high degree. High accuracy and handling of missing data was ensured by H2O stacked ensemble methods and KNN imputer. Accuracy scores attained both for RF and GBM were 79% and 76%, respectively. In case of RF, both precision, recall, and F1 were all 79%. In the case of GBM, precision, recall, and F1 scores achieved 76%. The H2O stacked ensemble performed better than all the models, with accuracy and recall being 87%, whereas precision and F1 scores were 85% and 86%, respectively. Accuracy was achieved with the proposed method using the KNN imputer at 97%.

[5] The authors in this study used the supervised learning technique to create the most accurate predictive models. This is using a labeled training dataset to find out whether water is consumable or not for some other purposes. Results show that application of the classification model after SMOTE with 10-fold cross-validation shows better results compared to other techniques, and this includes Accuracy of 98.1%, Precision of 100%, and Recall of 98.1% as well as AUC of 99.9%.

[6] This paper aims to discuss the areas of intersection between IoT and machine learning for a comprehensive review and prediction of water quality. The research used an IoT setup with four sensors: temperature, pH, turbidity, and TDS sensors to collect data from the Rohri Canal in SBA. The machine learning models applied in this study were Random Forest, XGBoost, SVM, and Decision Tree. The highest achieved model was Random Forest at a 0.93, followed by XGBoost, closely at 0.92, and Decision Tree at a maximum of 0.88. The lowest achievement had SVM with an accuracy score of 0.74.

[7] This paper mainly discusses the application of traditional water quality indicators combined with neural network techniques to predict water quality in watersheds, with the aim of timely preventing the deterioration of water quality and providing relevant agencies with theoretical frameworks and scientifically sound analyses for mitigation and management. Several machine learning algorithms, including SVM, Random Forest, GDBT, and LightGBM, are used in this research. The SVM model demonstrated the weakest predictive capability, which had accuracy at only 74.3% and an F1-score of 68.9%. The LightGBM model showed significant improvements for the predictive capability as this model reached accuracy of 97.5% with an F1-score at 97.8%.

[8] Different machine learning algorithms have been applied in the paper. The decision Tree, KNearest Neighbor, Support Vector Machine, Random Forest, and LightGBM have been applied. LightGBM model has gained the maximum accuracy among other classifiers. Accuracy 99.74% and a harmonic mean of 98%.

[9] Water samples were taken from wells in the study region of North Pakistan for creating WQI prediction models in this research. For this analysis, four independent algorithms namely, random trees (RT), random forest (RF), M5P, and reduced error pruning tree (REPT) were used. In addition, this study utilized twelve hybrid data-mining algorithms, which involve standalone methods and combination with BA, CVPS, and randomizable filtered classification. Models that resulted in the highest accuracy were RT-ANN, BA-RT, RF, BA-RF, BA-M5P, and M5P. The lowest mean absolute error and root mean square error values were seen in RT-ANN with 2.284 and 2.319, respectively.

[10] The research paper focused on the prediction of monthly BOD values for the Iraqi section of the Euphrates River. The objective was attained by constructing five different machine learning ensemble models, which include the quantile regression forest model, random forest model, radial support vector machine model, stochastic gradient boosting model, and GBM_H2O model. Among all the models, QRF showed a most similar distribution to that of the observed data. The interquartile values were the closest indicating that it was the best model. RF and SVM models performed the worst of all. The correlation coefficient between QRF and that of the observed data went above 0.95.

[11] In this paper different machine learning models were used which consisted of Long short-term memory (LSTM), Extreme Learning Machine (ELM),HammerStein Wiener Model (HW-RF), General Regression Neural Network (GRNNRF). The comparison of the predictive performance between the approaches demonstrates that HW-RF outperforms the others. The future scope of the paper includes investigation of the integration of ensemble techniques with emerging optimisation algorithms, deep learning models, and black box models.

A new hybrid prediction model called KIG-ELM, which is based on K-means, IGA, and ELM to address the issue of how to make an accurate prediction model, has been presented in this research. Other machine learning algorithms used in this study include SVM, XGBoost, Decision Tree, and Random Forest. The Random Forest model has obtained the highest accuracy at 0.93, and the XGBoost has reached an accuracy of 0.92.

Proposed System

1.1 Existing System

The water distribution system in Mumbai is a complex yet strong network designed to ensure the supply of clean and safe water to its residents. It begins with the collection of water from four important dams: Bhatsa, Upper Vaitarna, Middle Vaitarna, and Tansa. These dams serve as significant sources of water, providing the bulk water supply to the city. Once collected, the water undergoes a meticulous purification process at various complexes, namely the Bhandup Complex, Tulsi Filtration Plant, and Vehar Filtration Plant. At the Bhandup Complex, in particular, the water undergoes rigorous testing to ensure it meets the government's standards for various parameters. This involves conducting a series of experiments and tests in the laboratory to guarantee the water's quality and safety. After the purification, the water is distributed to 26 service reservoirs located across the city. These reservoirs act as crucial storage points where the water's contamination levels are continuously monitored. Additionally, measures such as chlorination are employed to control bacterial growth and maintain water quality.

From the service reservoirs, the water is then supplied to different societies and households throughout Mumbai. Each society typically has its own storage tanks where the water is stored before being distributed to individual households. This decentralized approach ensures efficient water distribution and accessibility to all parts of the city.

Overall, the water distribution system in Mumbai illustrates an exhaustive and well-coordinated effort to meet the water needs of a densely populated urban area, highlighting the significance of infrastructure, technology, and regulatory compliance in ensuring reliable access to clean water for all.

1.2 Drawback of Existing System

The existing system does not provide predictions based on historical data but lacks real-time updates such as finding out the exact location where the water has gotten contaminated or whether a specific attribute has crossed the threshold values. Implementing real-time monitoring capabilities and continuous data assimilation techniques can enable more dynamic and responsive water quality predictions.

1. Cost and complexity
   Online monitoring system implementation and maintenance can be costly, involving large investments in infrastructure, sensors, hardware, and employee training. Certain governments or water utilities may have difficulties due to the expenses involved in operation, installation, and maintenance. The municipalities may face difficulties in covering the expenses related to installation, operation, and maintenance. Because online monitoring systems are frequently intricate, proficiency in sensor technology, data processing, and system integration are prerequisites. Ensuring the reliability, accuracy, and compatibility of sensors and data collection systems can be challenging, especially in large water treatment facilities and distribution networks that serve urban regions such as Mumbai.
2. Sensor Reliability
   Such water monitoring systems may use a variety of sensors, some of which may be less accurate and more likely to wander or deteriorate with time. For sensors like pH sensors in particular, regular calibration, upkeep, and effective quality control procedures are crucial to guaranteeing the precision and dependability of sensor data. False alarms and inaccurate data might result from malfunctioning or failed sensors. As a result, periodic sensor maintenance is necessary.
3. Data Interpretation
   Interpreting and analyzing the large volumes of data generated by online monitoring systems can be daunting. Water quality managers and operators need the necessary skills and tools to process, analyze, and act upon the real-time data effectively. False alarms or ambiguous data interpretation may lead to unnecessary interventions or overlooked issues.
4. Infrastructure Vulnerability
   Piped water distribution systems rely on an extensive network of pipes, pumps, and treatment facilities. This infrastructure is vulnerable to various threats, including aging, corrosion, leaks, and physical damage from natural disasters or human activities.
5. Water Losses
   Leakage from aging pipes and infrastructure can result in significant water losses, reducing the efficiency of the distribution system and wasting valuable resources. In some cases, a large percentage of treated water may never reach the intended consumers due to leaks or unauthorized connections
6. Quality Degradation
   Water quality can deteriorate as it travels through the distribution network. Factors such as pipe material, age, and maintenance practices can contribute to the accumulation of sediments, biofilms, and contaminants, compromising water quality and posing health risks to consumers.
7. Operational Challenges
   Some of the operations that will be challenged to an extent in remote and underserved areas with limited resources and expertise will be managing and maintaining piped water distribution systems concerning the management of water quality, detection of leaks, optimizing the operation of pumps, and meeting regulatory requirements
8. Resilience to Disasters
   Distribution systems consisting of pipes are often severely compromised or destroyed during natural disasters and can lose pressure such that service interruptions, potential contamination, and the lack of access to potable water for the affected communities result.

2. Methodology

The goal of this research is to take a step forward to transform the monitoring of water quality, with a focus on important parameters like pH, turbidity, TDS, and total hardness. This work's boundaries are established by particular goals and techniques meant to produce precise predictive models for determining the potability of water. The scope of this research encapsulates the following objectives:

1) Data collection of Bhandup complex
   2) Data preparation and cleaning followed by the usage of machine learning techniques that include SVM, Random Forest, and LightGBM are used to evaluate water potability.
3) Taking data from the reservoir.
   4) Cleaning and pre-processing the dataset, followed by application of machine learning algorithms like SVM, Random Forest, and LightGBM for determining the potability of water.5) Collecting the data at the end-user's location with the help of hardware
6) Using python script in order to check the water potability
7) In the end, checking where exactly the water is getting contaminated by analysing the three checkpoints.

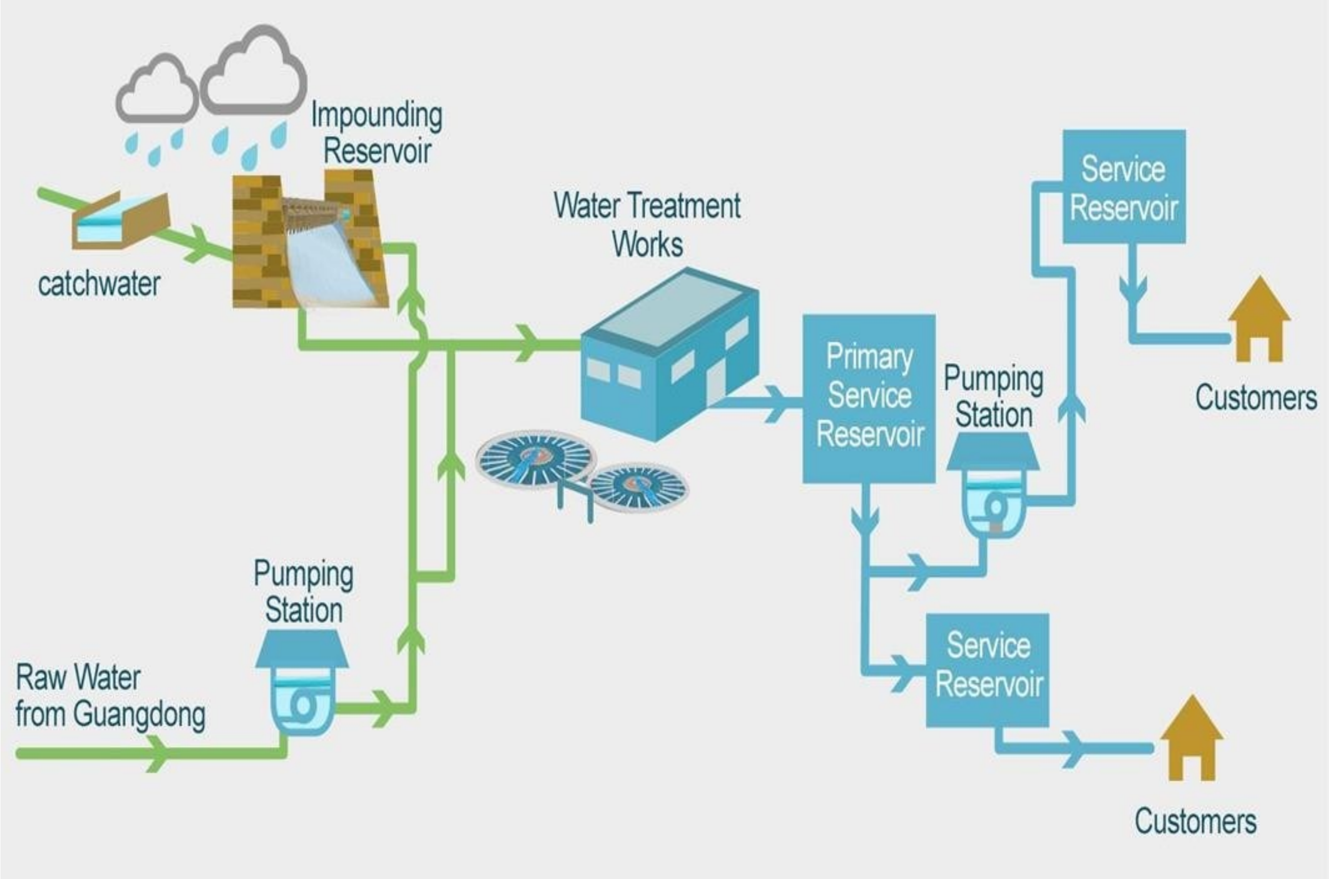


Fig. 4.1 Water Distribution System

4.1 Process of Distribution

In the above Fig 4.1, first the water from the dams is collected into the Bhandup Complex. After the water is collected at the Bhandup complex, the water is purified and cleaned by performing various tests in the laboratory. After the water is treated at the Bhandup Complex, it is supplied to 26 service reservoirs across Mumbai. The contamination levels of water are also checked and monitored continuously in these 26 service reservoirs. These 26 service reservoirs are responsible for supplying water to all the societies across Mumbai. The water supplied from the service reservoirs is stored in the tanks of the societies. This water is then supplied to different houses residing in the society.

Reservoir data

In the above diagram (Figure 4.1), at the intermediate source, the reservoir, data is gathered. Data was cleaned and pre-processed. It was divided into three sets: training set 80%, testing set 10%, and validation set 10%. To improve the robustness and reproducibility of the findings, 5-fold cross-validation was implemented during the model's training and evaluation phases. This method splits the dataset into five parts, using four for training and one for validation in each iteration. This approach allows the models to be assessed on various dataset segments, which reduces the likelihood of overfitting and provides a comprehensive evaluation of performance. For the final assessment, 10% of the dataset was reserved as a test set to evaluate the generalizability of the chosen model. The machine learning techniques utilized in this analysis include Random Forest, SVM, and LightGBM. These methods were selected due to their superior ability to manage structured data and for predictive modeling. One effective ensemble method that excels at handling non-linear relationships and reducing overfitting is Random Forest. With huge datasets, the gradient boosting technique LightGBM is very effective and provides faster training with good accuracy. Support Vector Machines (SVM) were chosen because of their capacity to manage intricate data distributions and provide exact decision bounds. Because these algorithms can represent non-linear correlations between water parameters including pH, turbidity, TDS, and hardness, they are well-suited for predicting water quality. Based on the input dataset, machine learning algorithms are trained. Finally, the algorithms will allow us to determine whether the water is fit for human consumption.

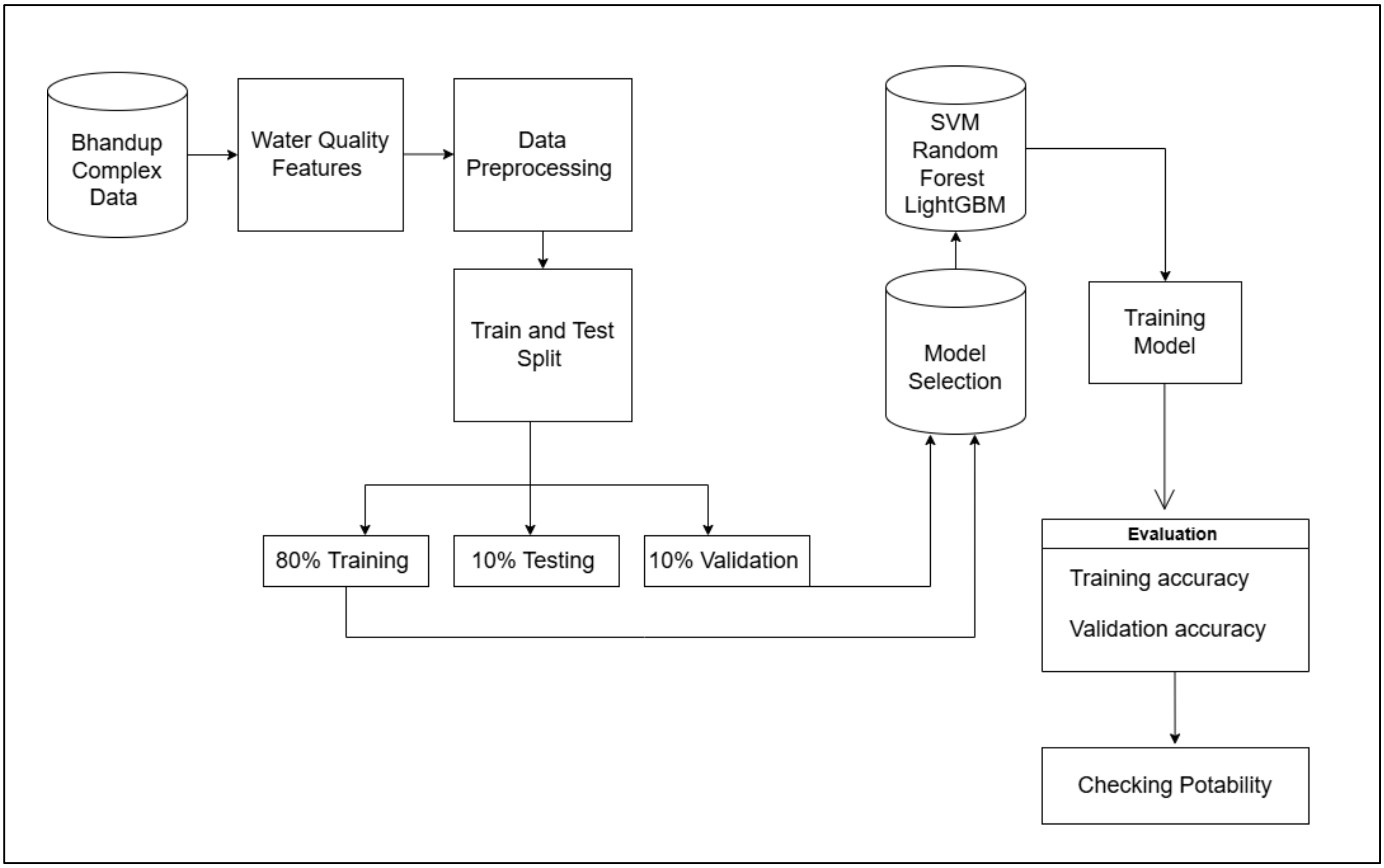


Fig. 4.2 Architecture Diagram

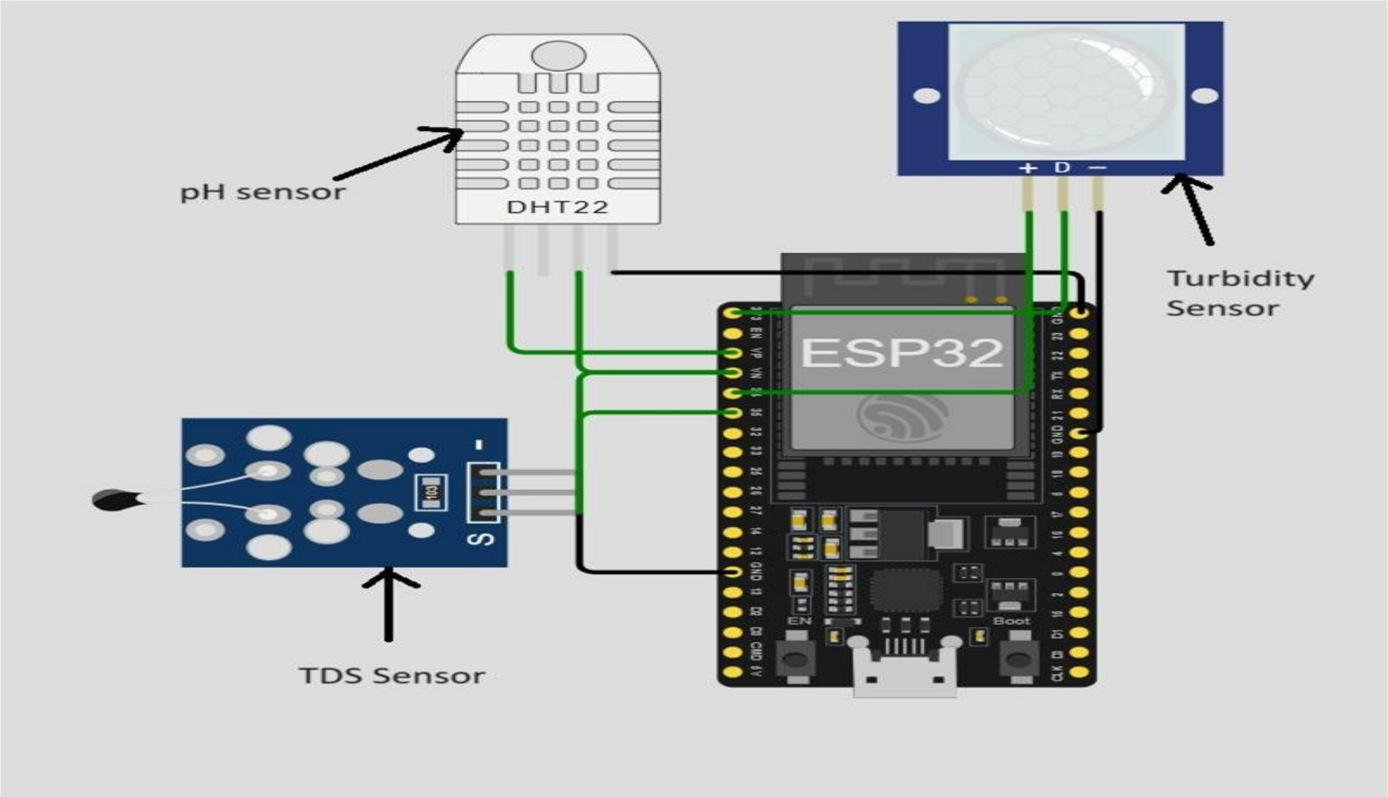


Fig. 4.3 Process of water quality check at end-user's

In the above Fig 4.3, the data at the end-user's location is collected with the help of hardware setup which includes pH sensor, Turbidity sensor, TDS sensor and then a python script containing if else statements is applied onto the data. To address sensor inaccuracies or missing data during analysis, error handling methods were implemented. Missing data points were imputed using mean values based on the distribution of the respective water quality parameters. The script gives the necessary output based on the input data and checks if the water at the end-user's location is potable or not. If the sensor is giving the readings which are within the ranges as 10 defined by the code, the water is potable else the water is not potable.

3. Implementation

Pseudocode (Hardware):

```
1. FUNCTION process_data_point(data_point)
2. CONVERT data_point values (pH, TDS, turbidity) to positive floats
3. FOR EACH value in data_point
4. CALL function to process individual value (e.g., process_ph(pH)) IF ALL processed values are within acceptable ranges RETURN 1 (good water quality)
5. ELSE
   RETURN 0 (bad water quality)
6. READ data from CSV file
7. FOR EACH row in data
8. IF row has at least 3 columns processed_result = process_data_point(row) ADD processed_result to results list
   PRINT individual value processing results
   PRINT final water quality assessment (good/bad based on processed result)
9. CALCULATE average of results list (if results exist)
10. PRINT average water quality assessment
```

Pseudo Code for Water Quality Prediction

```
1. Load and Preprocess Data
2. READ data from CSV file CREATE DataFrame data to store loaded data   # Explore data (optional)
   - Print data shape
   - Print information about target variable (e.g., number of samples per class)  # Separate features and target variable
3. features = data.drop("target_variable", axis=1)

   # Replace "target_variable" with actual target column name
4. target_variable = data["target_variable"]

   # Normalize features (optional)
5. CREATE Normalizer object (e.g., StandardScaler) - Normalize features using the Normalizer

   # Divide the data into training subset and testing subset
6. SPLIT the data into training subset and testing subset (by using train_test_split)
```

```
Train and Evaluate Random Forest Model
# Model Definition CREATE RandomForestClassifier object model_rf # Train the Model TRAIN model_rf on training data (features, target_variable) # Make PredictionsREDICT on testing data using model_rf # Evaluate the Model CALCULATE classification report and accuracy score # Perform Grid Search CV to find best hyperparameters
```

```
Train and Evaluate Other Models
# Define and Train Additional Models (e.g., LightGBM, SVM, AdaBoost)
1. CREATE model object for each model type (e.g., LGBMClassifier)
2. TRAIN each model on training data
3. PREDICT on testing data using each model
4. EVALUATE each model using classification report and accuracy score
Model Comparison and Visualization
# CREATE DataFrame to store model names and performance metrics
1. Include columns for model name, accuracy score, etc.
2. ADD information for each trained model # Visualize Performance (optional)
3. USE plotting library (e.g., Seaborn) to create a chart
4. PLOT model names on one axis and performance metric on another
```

```
K-Fold Cross Validation
# Define K-Fold object (e.g., KFold) with desired number of folds (e.g., 10) # Loop through models (e.g., SVM, Random Forest, LightGBM)
1. CREATE empty list to store validation scores
2.   FOR EACH fold in K-Fold object
     - SPLIT data into training and validation sets using the fold
     - TRAIN the model on the training set
     - EVALUATE the model on the validation set and store the score
     - CALCULATE average score across all folds for the current model
     - PRINT the average score for the current model
F1 Score Evaluation
# Retrain Models (optional)RAIN each model again on the entire training data # Calculate F1 Score for Each Model
1. FOR EACH model
   - PREDICT on testing data using the model
   - CALCULATE F1 score between predicted and actual target variable
   - PRINT F1 score for each model
END
```

4. Results

Table 1. Reservoir Dataset

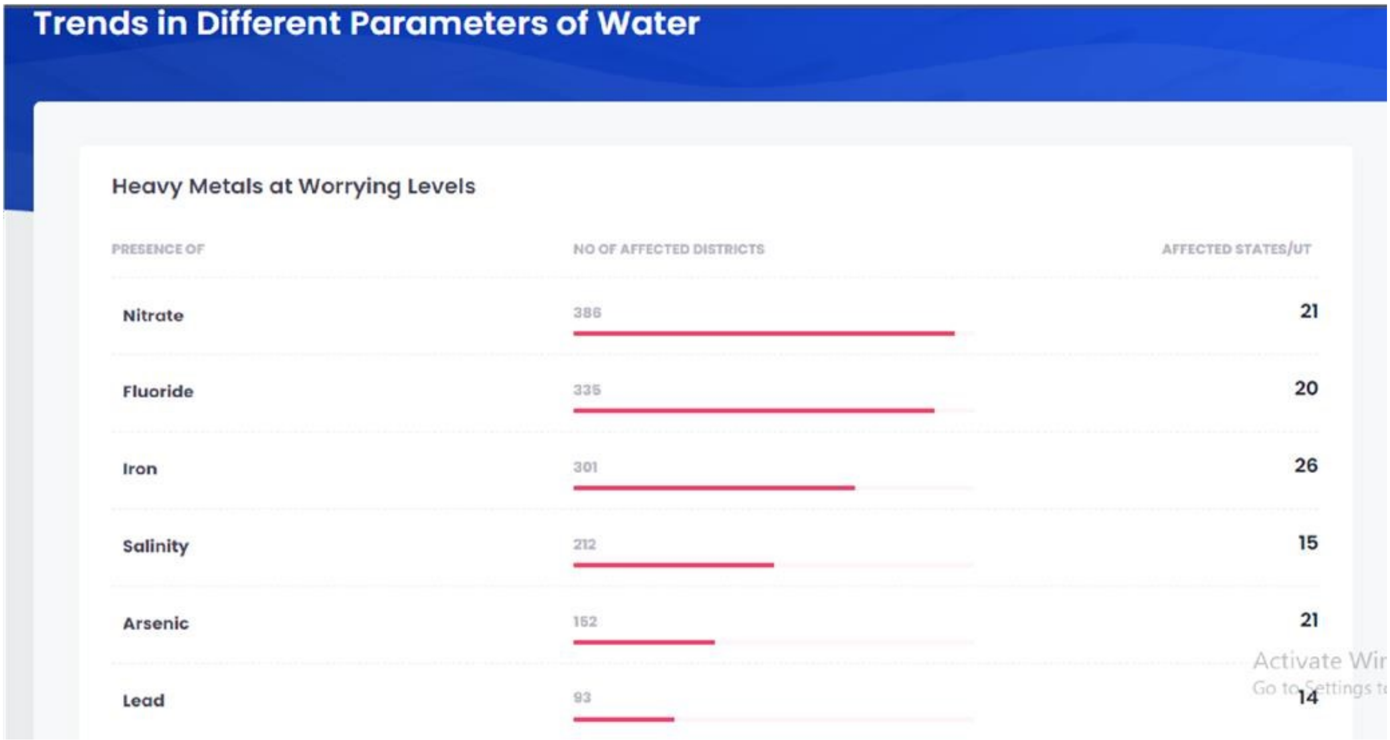
Algorithms	Testing Accuracy	Validation Accuracy	F1 score
SVM	48.8	65.9	68.5

Random Forest	73.1	72.1	73.6
LightGBM	75.5	72.6	75.5

Table 2. Dam Dataset

Algorithms	Testing Accuracy	Validation Accuracy	F1 score
SVM	48.2	98.4	94.7
Random Forest	98.3	94.8	97.9
LightGBM	97.5	98.3	97.8

The Table 1 and Table 2 represents how well the three algorithms performed on the reservoir and dam datasets. LightGBM has the best testing accuracy (75.5%), F1 score (75.5%), and validation accuracy (72.6%) for the Reservoir dataset. With a testing accuracy of 73.1%, validation accuracy of 72.1%, and F1 score of 73.6%, Random Forest fared well as well. With testing accuracy of 48.8%, validation accuracy of 65.9%, and F1 score of 68.5%, SVM scored poorly. Random Forest obtained the maximum testing accuracy (98.3%), F1 score (97.9%), and validation accuracy (94.8%) on the Dam dataset. With testing accuracy of 97.5%, validation accuracy of 98.3%, and F1 score of 97.8%, LightGBM trailed closely behind. SVM had a poorer testing accuracy (48.2%), but it did well in terms of validation accuracy (98.4%) and F1 score (94.7%).



Trends in Different Parameters of Water

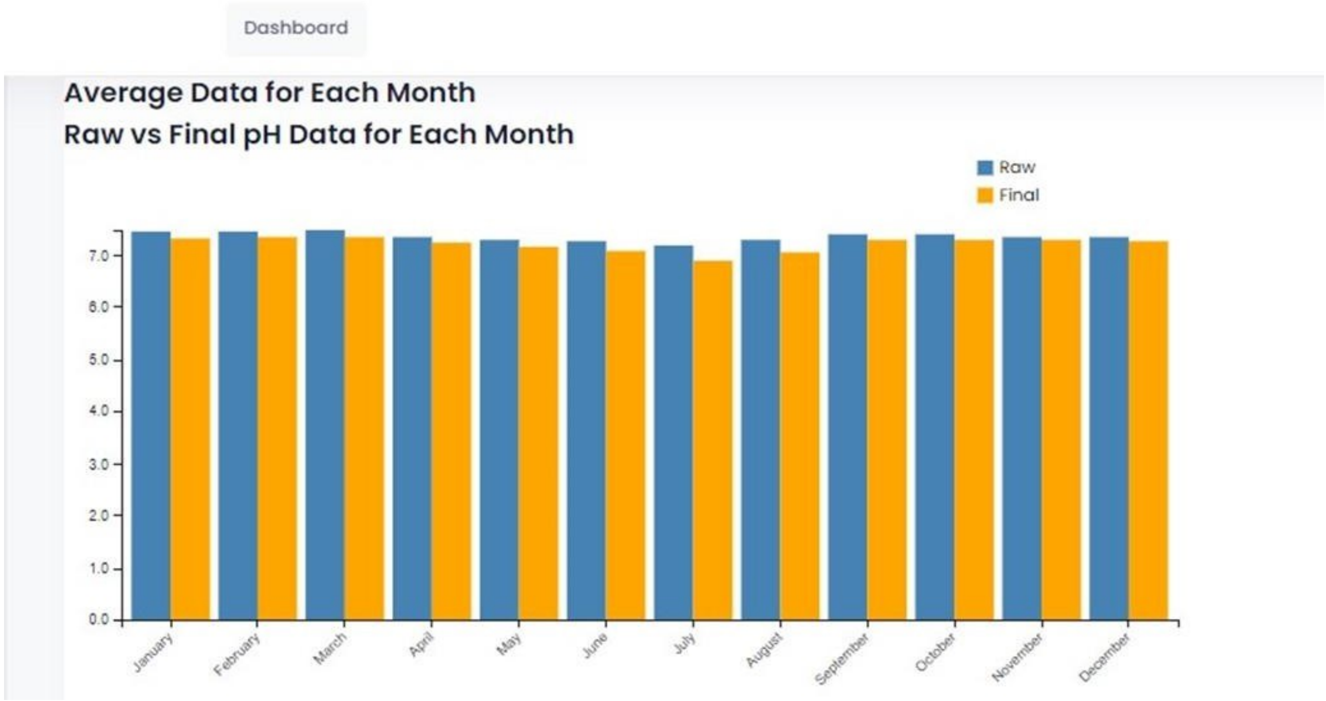
Heavy Metals at Worrying Levels

PRESENCE OF	NO OF AFFECTED DISTRICTS	AFFECTED STATES/UT
Nitrate	386	21
Fluoride	335	20
Iron	301	26
Salinity	212	15
Arsenic	152	21
Lead	93	14

Fig. 6.1 Dashboard1 – Water Quality

The above Fig. 6.1 shows the percentage of different heavy metals found in water.

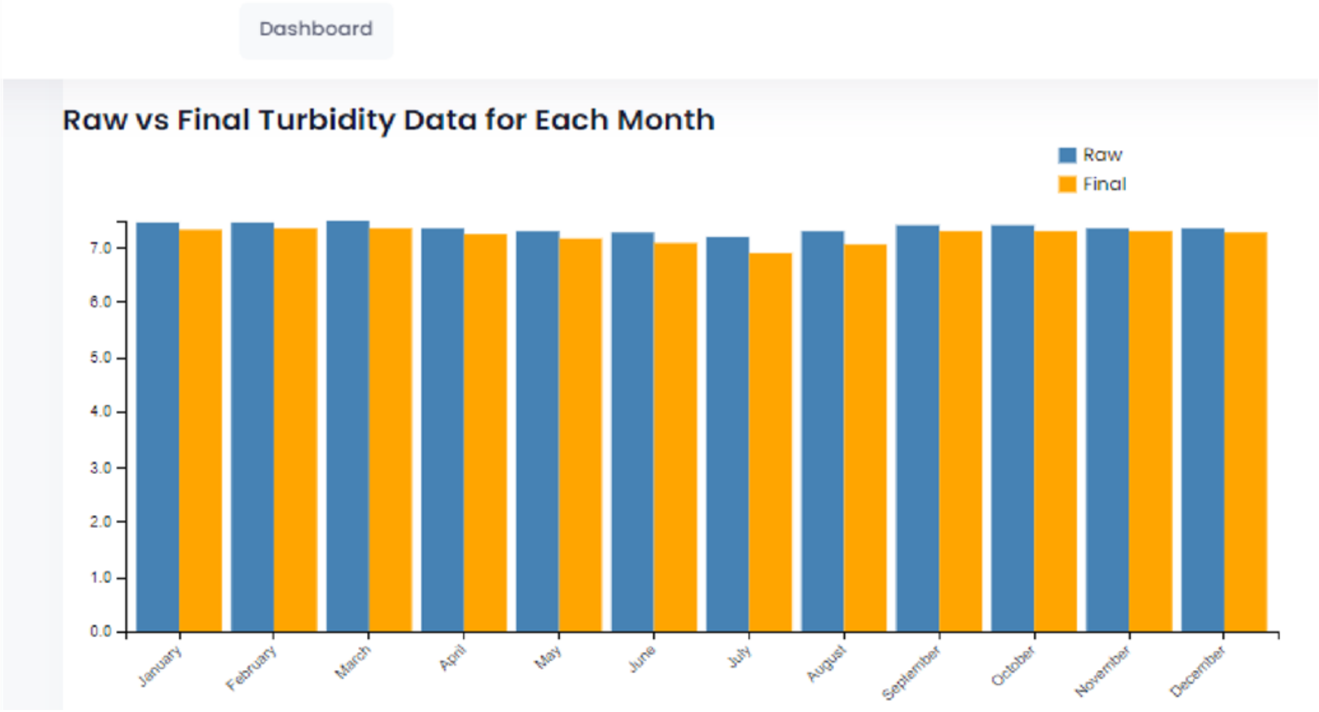
Fig. 6.2 Water Quality – pH analysis before and after filtration water



Average Data for Each Month

Raw vs Final pH Data for Each Month

The above figure shows the trends in pH present in before filtration and after filtration water for different months of the year.



Raw vs Final Turbidity Data for Each Month

Fig. 6.3 Water Quality – Turbidity analysis before and after filtration water

The above Fig 6.3 shows the trends in turbidity present in before filtration and after filtration water for different months of the year.

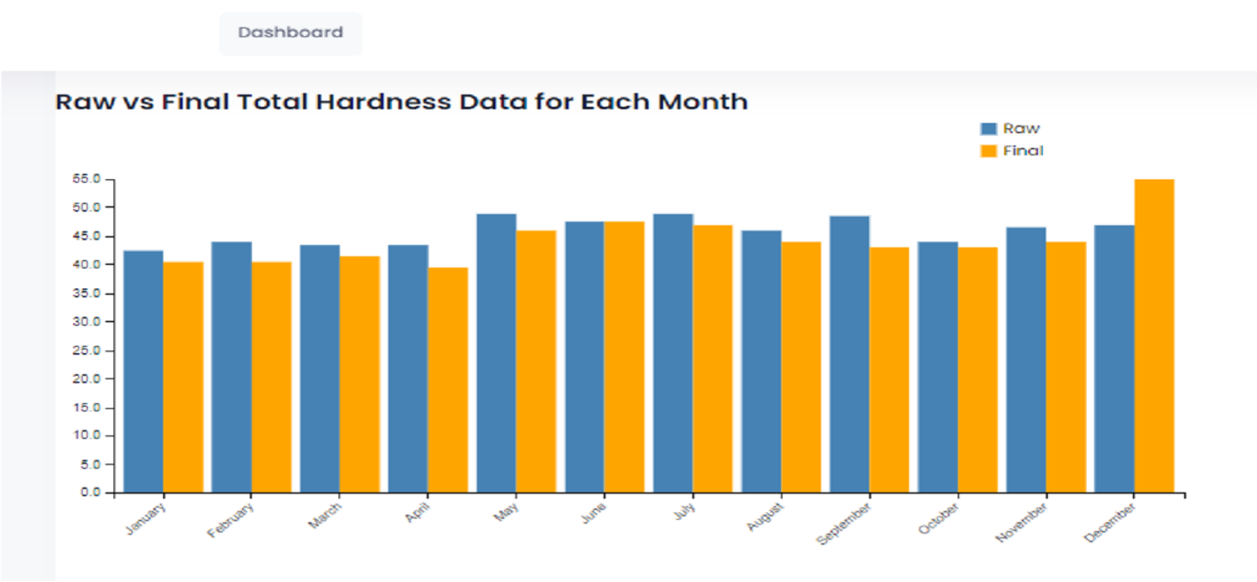


Fig. 6.4 Water Quality: Total Hardness analysis before and after filtration water

The above Fig. 6.4 shows the trends in Total Hardness present in of the year.
before filtration and after filtration water for different months

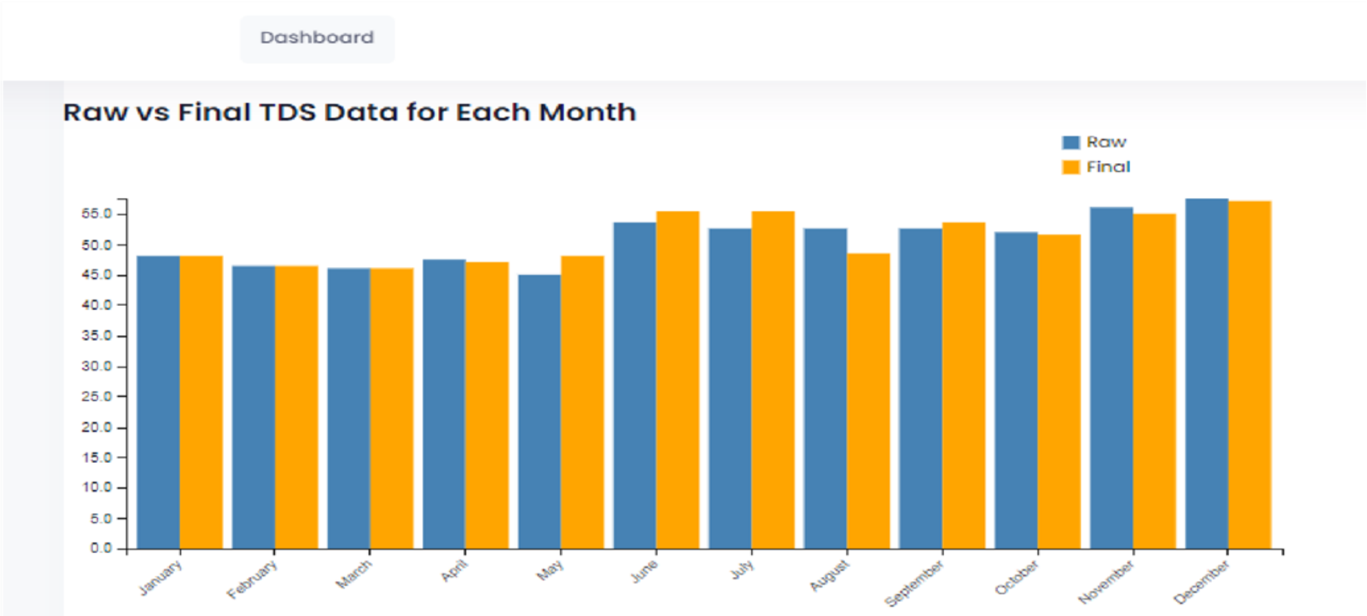
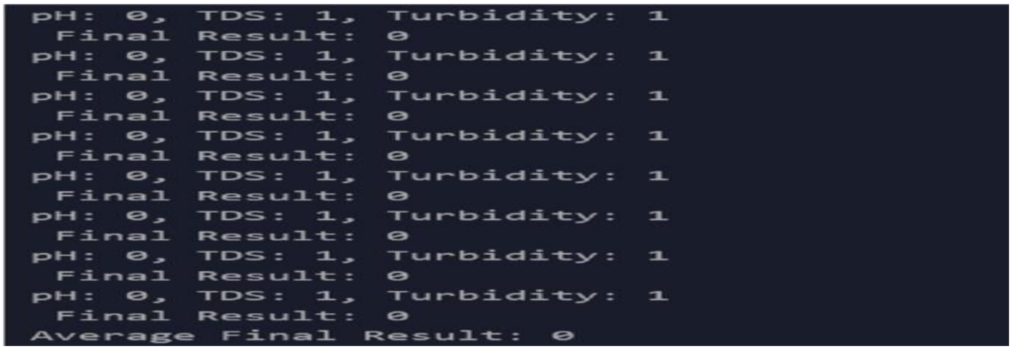


Fig. 6.5 Water Quality – TDS analysis before and after filtration water

The above Fig 6.5 shows the trends in TDS present in before filtration and after filtration water for different months of the year.



```
pH: 0, TDS: 1, Turbidity: 1
 Final Result: 0
pH: 0, TDS: 1, Turbidity: 1
 Final Result: 0
pH: 0, TDS: 1, Turbidity: 1
 Final Result: 0
pH: 0, TDS: 1, Turbidity: 1
 Final Result: 0
pH: 0, TDS: 1, Turbidity: 1
 Final Result: 0
pH: 0, TDS: 1, Turbidity: 1
 Final Result: 0
pH: 0, TDS: 1, Turbidity: 1
 Final Result: 0
pH: 0, TDS: 1, Turbidity: 1
 Final Result: 0
Average Final Result: 0
```

Fig. 6.6 Sensor Result

CONCLUSION

In conclusion, AquaSage integrates the hardware sensors with machine learning algorithms that provide real-time analysis of indicators of water quality. This is a step in water-quality monitoring. Being able to collect information not only from the reservoir but also from the Bhandup Complex Water Treatment Plant, it aids in proactive detection of when water contamination occurs while securing both environmental sustainability and public health. Improving and expanding the capacity of AquaSage becomes necessary for ensuring that everyone has access to clean and safe water as well as taking up emerging challenges and fostering collaboration across sectors, paving the way for a more sustainable and healthier future together.

Future Work

It is necessary to conduct additional research and development in the future to improve AquaSage's efficiency and practicality. This involves optimizing the user interface for increased accessibility and usability, exploration of further integration of other sensors to additionally determine more water quality parameters improves machine learning models in order to predict precisely, and conducts long-term monitoring studies in various environmental conditions to check the performance of AquaSage. Further endeavors regarding scalability initiatives, and community participation are necessary to guarantee the extensive implementation and efficaciousness of AquaSage in protecting water resources and public health.

REFERENCES

- Water Quality Prediction: A data-driven approach exploiting advanced machine learning algorithms with data augmentation, Journal of Water and Climate Change Vol 00 No 0,1,2023. https://doi.org/10.2166/wcc.2023.403
- S. A. A. Md.G.Uddin, "A novel approach for estimating and predicting uncertainty in water quality index model using machine learning approaches," Water Research, vol. 229, p. 119422, 2023.https://doi.org/10.1016/j.watres.2022.119422
- Mohamed Torky, Ali Bakhiet, Mohamed Bakrey, Ahmed Adel Ismail and Ahmed I. B. EL Seddawy, “Recognizing Safe Drinking Water and Predicting Water Quality Index using Machine Learning Framework” International Journal of AdvancedComputer Science and Applications(IJACSA), 14(1), 2023. http://dx.doi.org/10.14569/IJACSA.2023.0140103
- Madni HA, Umer M, Ishaq A, Abuzinadah N, Saidani O, Alsubai S, Hamdi M, Ashraf I. Water-Quality Prediction Based on H2O AutoML and Explainable AI Techniques. Water. 2023; 15(3):475. https://doi.org/10.3390/w15030475
- Dritsas E, Trigka M. Efficient Data-Driven Machine Learning Models for Water Quality Prediction. Computation. 2023; 11(2):16. https://doi.org/10.3390/computation11020016
- M. A. Rahu, A. F. Chandio, K. Aurangzeb, S. Karim, M. Alhussein and M. S. Anwar, "Toward Design of Internet of Things and Machine Learning-Enabled Frameworks for Analysis and Prediction of Water Quality," in IEEE Access, vol. 11, pp. 101055-101086, 2023, doi: 10.1109/ACCESS.2023.3315649.
- Zhou S, Song C, Zhang J, Chang W, Hou W, Yang L. A Hybrid Prediction Framework for Water Quality with Integrated W-ARIMA-GRU and LightGBM Methods. Water. 2022; 14(9):1322. https://doi.org/10.3390/w14091322
- R. S. Q. A. P. N.-A. S.I.Abba, " Integrating feature extraction approaches with hybrid emotional neural networks for water quality index modelling," Applied Soft Computing, vol. 114, p. 108036, 2022. https://doi.org/10.1016/j.asoc.2021.108036
- M.C.U.A.I. Aslam.B, "Water quality management using hybrid machine learning and data mining algorithms: An indexingapproach," IEEE Access, vol. 10, pp. 119692-119705, 2022. https://doi.org/ 10.1109/ACCESS.2022.3221430
- A.-M. R. F. K. Y. Al-Sulttani.A.O, "Proposition of new ensemble data-intelligence models for surface water quality prediction," IEEE Access, vol. 9, pp. 108527-108541, 2021. https://doi.org/10.1109/ACCESS.2021.3100490
- L. A. A. P. A. A. Abba.S.I, "Hybrid machine learning ensemble techniques for modelling dissolved oxygen concentration," IEEE Access, vol. 8, pp. 157218-157237, 2020. https://doi.org/10.1109/ACCESS.2020.3017743
- S. H. C. Z. Kuang.L, "An enhanced extreme learning machine for dissolved oxygen prediction in wireless sensor networks," IEEE Access, vol. 8, pp. 198730-198739, 2020. https://doi.org/10.1109/ACCESS.2020.3033455