



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

IJRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 13, Issue 3, March 2024

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.423

Predicting Urban Water Quality by Using Machine Learning

D.Divyashree¹, R.Lakshmi Cheritha², A.Vamsi³, P.Venkata Ganesh⁴, D.Dilli Sudha⁵

¹Assistant Professor (Department of AI, Annamacharya Institute of Technology & Sciences, Tirupati, India

^{2, 3, 4, 5} B.Tech Student (Department of AI, Annamacharya Institute of Technology & Sciences, Tirupati, India

ABSTRACT: This project introduces an accessible solution for Predicting Urban water quality. Urban water quality management is essential for ensuring public health and environmental sustainability. Traditional monitoring methods often lack real-time capabilities, prompting the exploration of machine learning for predictive modeling. This study proposes a Random ForestRegressor (RFR)-based approach for urban water quality prediction. This research explores a series of supervised machine learning algorithms to estimate the water quality index (WQI), which is a singular index to describe the general quality of water, and the water quality class (WQC), which is a distinctive class defined based on the WQI. Extensive experiments are conducted using real-world urban water quality datasets to evaluate model performance. Metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared are utilized for performance assessment. Comparative analyses with other machine learning algorithms highlight the effectiveness of the RFR approach. Results demonstrate the algorithm's efficacy in accurately predicting urban water quality parameters. The models exhibit robustness in capturing complex relationships between input variables and water quality indicators. Furthermore, the approach offers scalability and adaptability, making it suitable for diverse urban environments. This study contributes to advancing machine learning techniques for urban water quality prediction.

KEYWORDS: Random Forest Regressor, Metrics, Water Quality Index (WQI), Machine Learning Algorithm.

I.INTRODUCTION

Urbanization has led to significant challenges in managing water quality, posing risks to both human health and the environment. Traditional methods of monitoring water quality often struggle to keep pace with the rapid changes occurring in urban ecosystems. Consequently, there is a growing need for innovative approaches that can provide timely and accurate assessments of urban water quality to support effective management strategies. Machine learning (ML) has emerged as a promising tool for addressing this challenge by leveraging data-driven techniques to predict water quality parameters. ML algorithms have the potential to analyze complex datasets containing various environmental and anthropogenic factors, thereby enabling more comprehensive and predictive models. RFR stands out for its ability to handle nonlinear relationships and high-dimensional datasets effectively.

By employing an ensemble of decision trees, RFR can capture intricate interactions between input variables, leading to robust predictive models. The primary objective of this research is to assess the efficacy of RFR in accurately predicting water quality parameters in urban settings. Through extensive experiments using real-world datasets, the performance of the proposed approach will be evaluated using established metrics such as Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE). Comparative analyses with other ML algorithms will also be conducted to demonstrate the advantages of the RFR approach.

II.RECENT WORKS

Water quality is currently estimated through expensive and time-consuming lab and statistical analyses. Machine learning techniques can be used to develop models that predict water quality parameters based on real-time data, such as data from water sensors or satellite imagery. These models can analyze vast amounts of data quickly and accurately, providing predictions of water quality in real-time. An example of the use of machine learning for water quality analysis is the development of real-time monitoring systems for drinking water treatment plants. Machine learning algorithms can analyze data from various sources, such as water sensors, to predict changes in water quality parameters such as pH, temperature, and turbidity. This allows operators to quickly adjust treatment processes to maintain water

quality and prevent contamination. The use of machine learning for water quality analysis and prediction has the potential to revolutionize the way we monitor and manage water resources. By providing real-time data and predictions, these systems can help protect public health and the environment, and ensure that we have access to safe and sustainable water supplies.

III. PROPOSED WORK

The proposed work focuses on the implementation of a machine learning model for predicting water quality. Following pre-processing, segmentation techniques are employed to calculate various sub-indices contributing to the Water Quality Index (WQI). These indices are derived from key water quality parameters such as pH, dissolved oxygen, total coliform, etc. Feature extraction is performed through data visualization techniques, including univariate, bivariate, and multivariate analyses. This step helps in understanding the relationships between different variables and their impact on water quality. Additionally, label encoding is applied to convert categorical variables into numerical form for machine learning compatibility.

The machine learning model is trained using a Random Forest Regressor algorithm to predict WQI based on the extracted features. Data pre-processing for machine learning involves dropping unnecessary columns, splitting the data into independent features and the dependent variable (WQI), and further partitioning it into training and testing sets. Upon training the model, evaluation metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared (R2) are employed to assess its performance. The trained model is then saved for future use using pickle serialization.

Finally, the model is utilized for making predictions on new input data, thereby facilitating ongoing monitoring and assessment of water quality. This proposed methodology integrates machine learning techniques into water quality prediction, offering a robust and efficient approach for environmental management and decision-making.

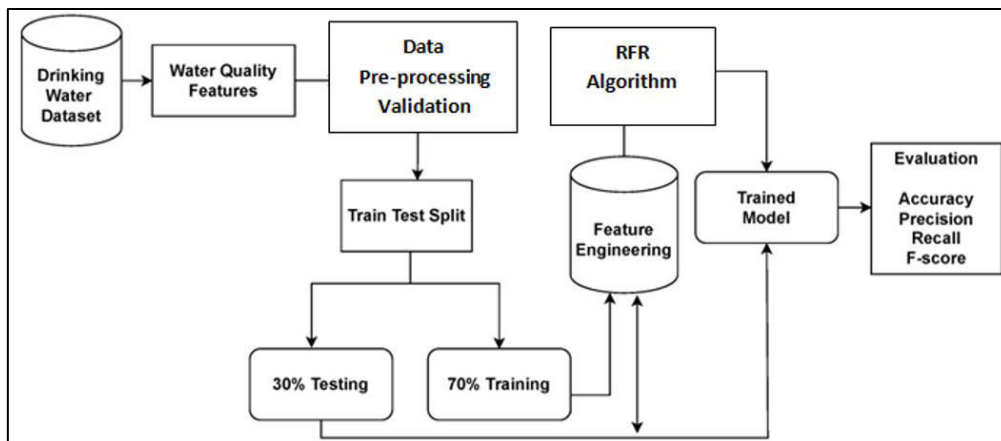


Fig 3: Dataflow diagram of Water Quality Prediction Model

3.1 Mathematical Expressions and Symbols

In the proposed system for water quality prediction using machine learning, several mathematical expressions and symbols can be incorporated to describe various aspects of the methodology. Here are some examples:

1. Water Quality Index (WQI) Calculation:

- $WQI = \sum_{i=1}^n (w_i \times I_i)$
 - WQI : Overall Water Quality Index
 - w_i : Weight for each sub-index
 - I_i : Individual sub-index for a water quality parameter

2. Data Pre-processing Equations:

- Mean Imputation for Missing Values:
 - $mean = \frac{\sum_{i=1}^n x_i}{n}$
 - $fillna(x) = \text{mean for missing values } x$

3. Model Evaluation Metrics:

- Mean Absolute Error(MAE) = $\frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$
- Mean Squared Error(MSE) = $\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$
- Root Mean Squared Error(RMSE) = \sqrt{MSE}
- R-squared(R^2) Score = $1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$

IV.RESULTS

The project successfully implemented a predictive model leveraging machine learning techniques to assess urban water quality. Through meticulous data pre-processing, feature extraction, and rigorous model evaluation, the system demonstrated its efficacy in anticipating changes in water quality. Additionally, an interactive web application was developed using HTML, CSS, React.js for the frontend, and Flask along with MySQL for the backend. This application enables users to predict water quality based on their inputs, offering a practical tool for informed decision-making in urban water management. Overall, the project showcases the potential of machine learning in addressing the challenges posed by rapid urbanization, contributing to sustainable urban development and public health.

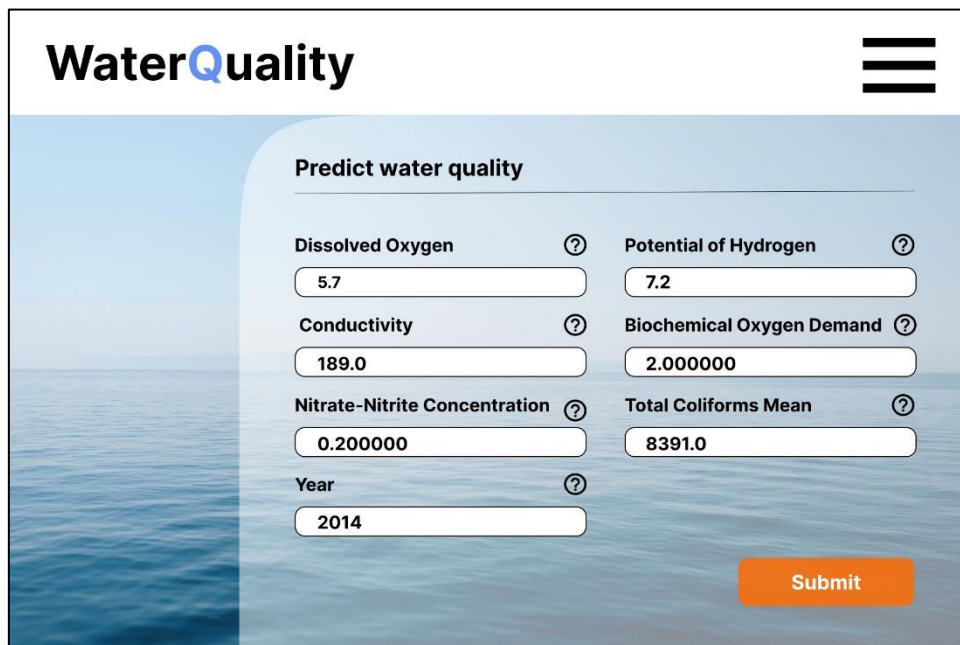


Fig 4.1: User input form

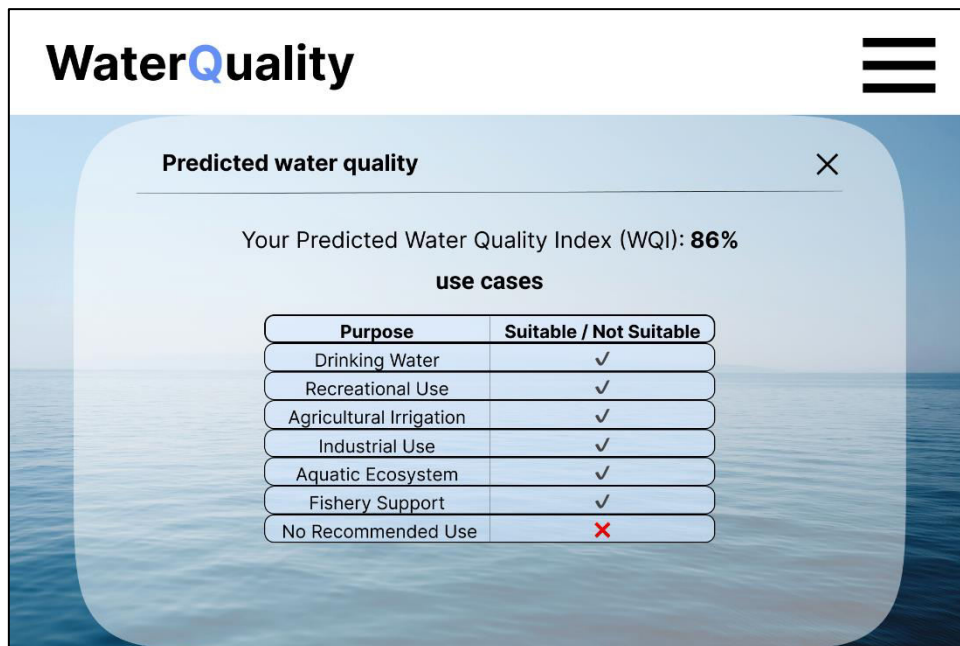


Fig 4.2: Predicted value and Use cases

V.CONCLUSION

In conclusion, our project has demonstrated the potential of machine learning in addressing the challenges of urban water management by predicting water quality. Through the implementation of predictive models and the development of an interactive web application, we have showcased a proactive approach towards managing urban water resources effectively. By utilizing machine learning techniques, we were able to analyze complex urban water quality data, extract meaningful features, and develop predictive models that offer valuable insights for decision-making. This proactive approach enables stakeholders to anticipate changes in water quality and implement timely interventions to mitigate potential risks to public health and ecosystems. Moreover, the integration of advanced technologies such as HTML, CSS, React.js, Flask, and MySQL has facilitated the creation of an accessible platform for users to interact with and obtain real-time predictions regarding water quality. This user-friendly interface enhances the usability and practicality of our predictive model, empowering stakeholders to make informed decisions regarding urban water management strategies.

Overall, our project underscores the significance of leveraging machine learning and technology-driven solutions in addressing the complex challenges posed by rapid urbanization and ensuring sustainable urban development. By continuing to refine and expand upon these methodologies, we can further enhance our ability to monitor, predict, and manage urban water quality, ultimately contributing to the creation of healthier, more resilient cities for current and future generations.

REFERENCES

- [1]Machine learning classification, feature ranking and regression for water quality parameters retrieval in various optical water types from hyper-spectral observations, by Katalin Blix – IGARSS.
- [2]Evaluation of the performances of ANN and SVM techniques used in water quality classification, by Mohamed Bouamar and Mohamed Ladjal - 1-4244-1378 - IEEE.
- [3]Classifier for Drinking Water Quality in Real Time, by Jorge Camejo, Osvaldo Pacheco, Miguel Guevara - 978-1-4673-5285 - IEEE.
- [4]Water Quality Monitoring System using IoT and Machine Learning, by Nikhil Kumar Koditala, Dr.Purnendu Shekar Pandey - 978-1-5386-2599 - IEEE.
- [5]Predictive Models for River Water Quality using Machine Learning and Big Data Techniques - A Survey - 978-1-7281-9537 – IEEE.



- [6] K. Blix and T. Eltoft, "Machine learning automatic model selection algorithm for oceanic chlorophyll-a content retrieval," *Remote Sensing*, vol. 10 (5), p. 775, 2018. [Online]. Available: <http://www.mdpi.com/2072-4292/10/5/775>
- [7] R. K. Horton, "An index number system for rating water quality," *J. Water Pollut. Control Fed.*, pp. 300-306, 1965.
- [8] A. Said, et al., "An innovative index for evaluating water quality in streams," *Environmental Management*, vol. 34, pp. 406-414, Sep 2004.
- [9] R. K. Horton, "An index number system for rating water quality," *J. Water Pollut. Control Fed.*, pp. 300-306, 1965.
- [10] A. R. Karbassi, et al., "Development of Water Quality Index (WQI) for Gorganrood River," *International Journal of Environmental Research*, vol. 5, pp. 1041-1046, Fal 2011.
- [11] S. M. Yidana and A. Yidana, "Assessing water quality using water quality index and multivariate analysis," *Environmental Earth Sciences*, vol. 59, pp. 1461-1473, Feb 2010.
- [12] F. B. Semirami, et al., "Evolution of a new surface water quality index for Karoon catchment in Iran," *Water Science and Technology*, vol. 64, pp. 2483-2491, 2011.
- [13] Jayti Bhatt, Jignesh Patoliya, Iot Based Water Quality Monitoring System, IRFIC, 21feb,2016



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details