WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Predictive Modelling of Water Quality Using Machine Learning With pH, Turbidity and Temperature

***Nallamilli SatyaBhavani*[1]**

B. Tech Student, Department of IT, GMR Institute of Technology, Rajam-532127, Andhra Pradesh, India
Email: 21341A1284@gmrit.edu.in[1]

A B S T R A C T

Ensuring the availability of safe drinking water is essential for human well-being, and accurately predicting water quality is crucial for maintaining a pure water supply. Traditional methods often rely on statistical models, which may struggle to capture the intricate relationships among different water quality parameters. This investigation centers on utilizing machine learning algorithms, namely RF (Random Forest) and SVM (Support Vector Machine) to predict the quality of drinking water. This prediction relies on input parameters such as pH, turbidity, and temperature. The dataset, compiled from various sources, underwent thorough preprocessing for machine learning analysis. The RF and SVM models were then trained and assessed, revealing impressive accuracy and resilience in handling noise and outliers. These algorithms prove versatile for monitoring water quality, identifying contaminants, and facilitating decision-making in water treatment processes. Through the utilization of machine learning, this study contributes to advancing methodologies that consistently ensure the delivery of safe and high-quality drinking water, thereby protecting public health.

Keywords: Random Forest, Support Vector Machine, pH, Turbidity, Temperature.

Introduction

The exploration of predictive modeling for water quality using machine learning, specifically focusing on the influential parameters of turbidity and temperature, marks a significant stride in the realms of environmental science and public health. The importance of ensuring access to clean and safe drinking water underscores the need for accurate assessments, prompting a departure from traditional methods that may struggle to encompass the nuanced relationships within various water quality factors. In response to these challenges, this research delves into the innovative realm of machine learning algorithms, with a particular emphasis on RF and SVM, as tools for predicting water quality. Turbidity, gauging water clarity, and temperature, a fundamental environmental variable, are central to this predictive modeling approach. The study begins by compiling a diverse dataset drawn from various environments, ensuring a comprehensive representation of water quality scenarios. Thorough preprocessing techniques are then applied to refine the dataset, preparing it for analysis using RF and SVM algorithms. The goal is to attain a high level of accuracy in predicting water quality, showcasing the adaptability of these machine learning models in handling intricate datasets.

The significance of this research lies not only in advancing the scientific understanding of water quality dynamics but also in offering practical insights for the monitoring and management of drinking water supplies. The anticipated outcomes aim to inform decision-makers, guide water treatment processes, and contribute to the formulation of effective environmental policies. Ultimately, the study seeks to play a pivotal role in ensuring the consistent delivery of safe and high-quality drinking water, thereby safeguarding public health.

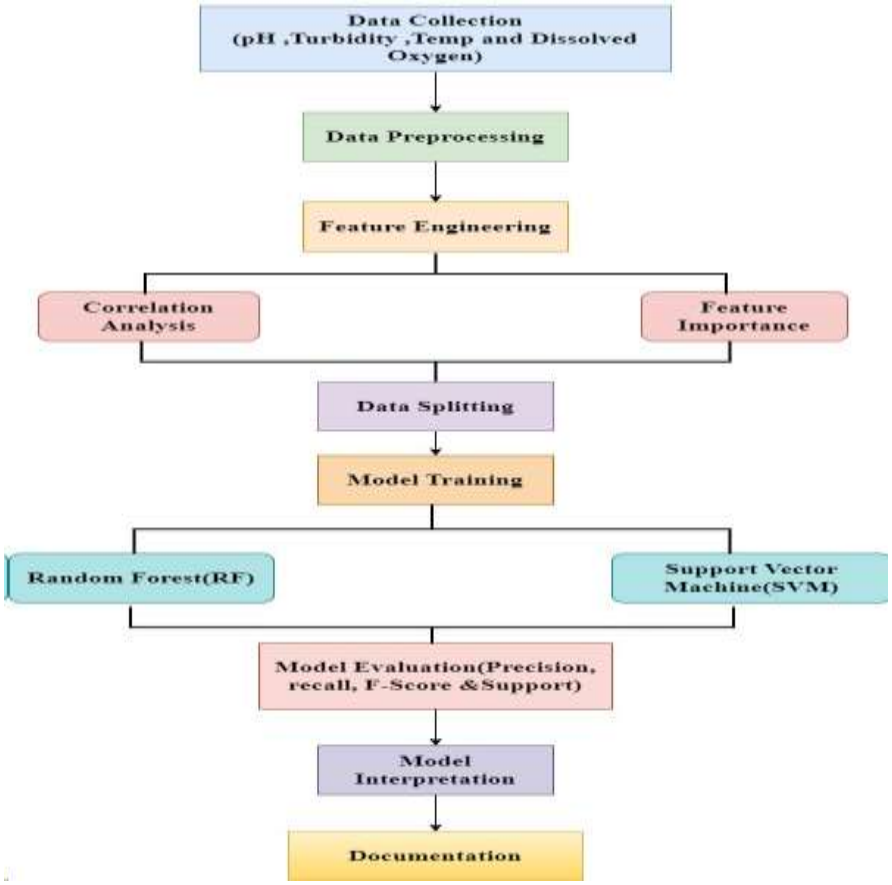
Literature Review

Covering an extensive array of studies and discussions, the review centers on the analysis and anticipation of water quality, highlighting its pivotal role in upholding public health and environmental safety. It accentuates the necessity for precise and effective assessment techniques. The discourse delves into the utilization of machine learning such as SVM(Support Vector Machine), DT(Decision Trees) and RF(Random Forest) for water quality prediction, with a specific focus on their application in analyzing well water. Moreover, the survey explores the potential of advanced artificial intelligence (AI) algorithms, including artificial neural network models and machine learning algorithms, anticipating the WQI(water quality index) and determining WQC(water quality classification) stand as pivotal tasks. These models convincingly demonstrate their effectiveness in predicting a range of water quality parameters.

Additionally, the review scrutinizes the applications of rough set theory in handling ambiguity and uncertainty in decision-making, particularly in data mining and extracting decision rules from data. Emphasizing the importance of water quality assessment standards and procedures, it offers a

comprehensive overview of the guidelines and methodologies applied in water quality assessment. This places the research in the broader context of water quality analysis.

METHODOLOGY



1. Data Collection:

The process commences with the thorough gathering of data related to four essential water quality parameters: pH, turbidity, temperature, and dissolved oxygen. This data forms the basis for subsequent analysis and modeling.

2. Data Preprocessing:

The collected data undergoes a meticulous preprocessing phase to ensure its reliability and suitability for analysis. This includes addressing missing values, rectifying inconsistencies, standardizing formats, and mitigating the impact of outliers. This stage establishes a cleansed and prepared dataset ready for further exploration.

3. Feature Engineering:

In this crucial step, innovative techniques are employed to derive enhanced features from the existing data. This may involve creating new features to enhance predictive capabilities, reducing dimensionality through feature selection, or transforming features for better model understanding. The objective is to increase the informativeness of the data and enhance overall model performance.

4. Feature Importance Assessment:

To prioritize the most influential features, their relative importance is thoroughly assessed. This analysis highlights features with the most significant impact on predictions, guiding feature selection and improving model understanding. It enables informed decisions regarding which features to emphasize during model training.

5. Data Splitting:

The dataset is strategically divided into two distinct subsets:

- Training Set: This portion is used to meticulously construct and calibrate the machine learning models.

- Testing Set: This independent segment is exclusively reserved for evaluating the models' performance and generalizability, ensuring an unbiased assessment.

6. Model Training:

Two distinct machine learning models are carefully trained on the prepared dataset:

- Random Forest (RF): This robust model leverages the collective wisdom of an ensemble of decision trees, promoting resilience to overfitting and often yielding exceptional predictive accuracy.
- Support Vector Machine (SVM): This powerful algorithm excels in navigating high-dimensional data spaces, adeptly constructing decision boundaries that effectively categorize data points.

7. Model Evaluation:

The trained models undergo rigorous assessment using a comprehensive suite of performance metrics, including precision, recall, F-score, and support. These metrics collectively measure the models' ability to accurately classify or predict outcomes, providing a quantitative basis for model comparison and selection.

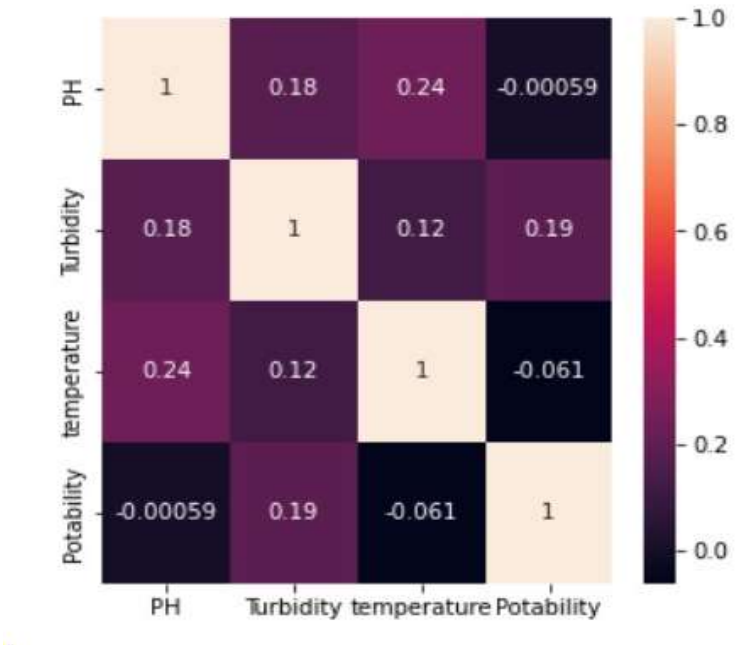
8. Model Interpretation:

To gain deeper insights into the models' decision-making processes and enhance explainability, techniques such as feature importance analysis and visualization are employed. This stage unveils the underlying logic and reasoning behind model predictions, enhancing understanding and trust in the outcomes.

9. Documentation:

In the concluding phase, a thorough documentation process captures the entire methodology, details results comprehensively, and articulates key findings. This comprehensive documentation promotes reproducibility, facilitates knowledge sharing within the scientific community, and enables future research to build upon these findings.

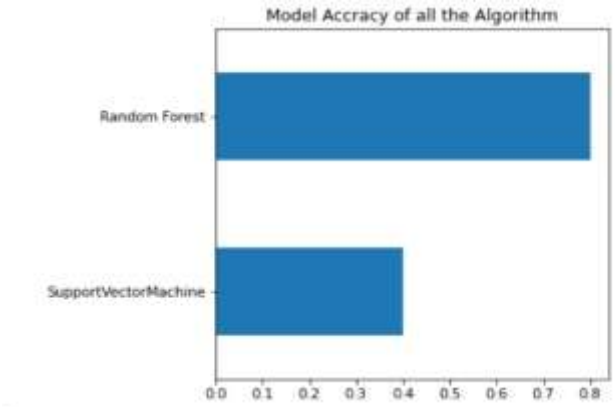
Results



Relationship Between Temperature and Turbidity in Water Quality

	precision	recall	f1-score	support
0	1.00	0.33	0.50	3
1	0.78	1.00	0.88	7
accuracy			0.80	10
macro avg	0.89	0.67	0.69	10
weighted avg	0.84	0.80	0.76	10

	precision	recall	f1-score	support
0	0.29	0.67	0.40	3
1	0.67	0.29	0.40	7
accuracy			0.40	10
macro avg	0.48	0.48	0.40	10
weighted avg	0.55	0.40	0.40	10



Model Accuracy Comparison of Machine Learning Algorithms for Water Quality Assessment

Conclusion

In summary, leveraging machine learning for water quality predictive modeling, focusing on pH, turbidity, and temperature, proves highly effective. The systematic approach, from data collection to model training (utilizing Random Forest and Support Vector Machine), yields robust models. Feature importance analysis emphasizes the critical role of pH, turbidity, and temperature, guiding future refinements. Strategic data splitting ensures rigorous evaluation, and performance metrics affirm model accuracy. Interpretation techniques enhance explainability, unveiling the influence of key parameters. Meticulous documentation promotes transparency and knowledge dissemination. This research advances water quality analysis, offering a reliable framework for utilizing machine learning in safeguarding and enhancing water resources.

References

[1] Ahmed, U., Mumtaz, R., Anwar, H., Shah, A. A., Irfan, R., & García-Nieto, J. (2019). Supervised machine learning is effectively utilized for accurate water quality prediction. Published in Water, 11(11), 2210.

[2] Patel and colleagues (2022) crafted a model for predicting water potability, employing a machine learning methodology that integrates the synthetic minority oversampling technique and explainable AI. The study is published in Computational Intelligence and Neuroscience: CIN in 2022.

[3] Kurra, S. S., Naidu, S. G., Chowdala, S., Yellanki, S. C., & Sunanda, D. B. E. (2022). Accurate water quality forecasting is achieved through the implementation of machine learning. Published in the International Research Journal of Modernization in Engineering Technology and Science, India.

[4] In Processes (2022), Najwa Mohd Rizal et al. conduct a thorough investigation comparing regression models, SVM(Support Vector Machine) and ANN(Artificial Neural Network) for their effectiveness in predicting the quality of river water.

[5] In their work scheduled for 2023, D. Venkata Vara Prasad, Suresh Jagannathan, and Santosh Sivan forecast water quality by employing machine learning algorithms alongside rough set theory.

[6] Al-Sulttani, A. O., Al-Mukhtar, M., Roomi, A. B., Farooque, A. A., Khedher, K. M., & Yaseen, Z. M. (2021). Proposes novel ensemble data-intelligence models for precise surface water quality prediction. Published in IEEE Access, 9, 108527-108541.

[7] El Bilali, A., & Taleb, A. (2020). Machine learning models successfully predict irrigation water quality parameters in a semi-arid environment. Published in the Journal of the Saudi Society of Agricultural Sciences, 19(7), 439-451.

[8] Sarkar, A., & Pandey, P. (2015). Artificial Neural Network is employed for River Water Quality Modeling.

[9] 16. P. Li and J. Wu, "Examines drinking water quality and public health," published in Exposure and Health, vol. 11, no. 2, pp. 73–79, 2019.

[10] Hassan, M. M., Hassan, M. M., Akter, L., Rahman, M. M., Zaman, S., Hasib, K. M., ... & Mollick, S. (2021). The effective forecasting of the water quality index (WQI) is achieved by employing machine learning algorithms. This research is featured in Human-Centric Intelligent Systems, 1(3-4), 86-97.

[11] Ahmed, A. N., Othman, F. B., Afan, H. A., Ibrahim, R. K., Fai, C. M., Hossain, M. S., ... & Elshafie, A. (2019). Utilizing machine learning techniques enhances the accuracy of water quality prediction. Published in the Journal of Hydrology, 578, 124084.

[12] Khadr, M., & Elshemy, M. (2017). Data-driven modeling is utilized for accurate forecasting the quality of water in the drainage system connected to Manzala Lake, Egypt. Published in Ain Shams Engineering Journal, 8(4), 549-557.

[13] B. Charbuty and A. M. Abdulazeez employ a decision tree algorithm for machine learning classification. Published in the Journal of Applied Science and Technology Trends, vol. 2, no.

[14] M. Hassan, L. Akter, et al., successfully forecast the water quality index (WQI) through the application of machine learning algorithms. The findings are published in Human-Centric Intelligent Systems. Volume 1, Numbers 3-4, pages 86–97, 2021.

[15] Haghiabi, A. H., Nasrolahi, A. H., & Parsaie, A. (2018). Machine learning methods are employed for accurate water quality prediction. Published in the Water Quality Research Journal, 53(1), 3-13.