Prediction of Water Quality Index Using Machine Learning Algorithms

by

Ayse B Sengul

Water is an essential component of human well-being and socia-economic development [1]. According to the 2021 World Water Development Report released by UNESCO, global use of freshwater has increased sixfold over the past 100 years and continues to grow at the rate of 1% each year. This increase is attributed to a combination of population growth, economic development and shifting consumption patterns [2]. With the increase of water consumption, there are significant difficulties emerging regarding maintaining and preserving the quality of water. Industrialization, agricultural activities, and urbanization have led to harmful impacts on the environment, resulting in the degradation and pollution of water bodies. These adverse effects have significant implications for both human health and the sustainable progress of societies. Therefore, monitoring and assessing water quality are crucial for protecting both human health and the environment, while also facilitating effective and sustainable water management [3].

Water Quality Index (WQI) is a mathematical model by converting large quantities of water quality data into a single value, making it easier to understand and monitor the overall quality of water resources [4, 5]. The WQI comprises four processes that include parameter selection, sub-Index calculation, weighting factor determination, and sub-Indices aggregation. First process involves choosing which water quality parameters to include in the evaluation. Once the parameters are selected, actual water quality data is collected. For each chosen parameter, the concentrations found in the water samples are converted into a single numerical value. These values help create sub-indices, which are essentially scores or ratings for each parameter. Third, weighting factor is assigned to each parameter to reflect their significance. Parameters with more impact on water quality receive higher weights in the calculation. Finally, an aggregation function combines all the sub-indices by using the assigned weighting factors for each parameter. This process generates a final single numerical value - the Water Quality Index. [6-9].

However, WQI might not provide an entirely accurate assessment of water quality. The reason being that each index is typically designed for certain locations or types of water, making it biased towards those specific conditions. Additionally, these indices can be highly sensitive to specific parameters concentrations or heavily reliant on the assigned weights for these parameters, which can affect the overall assessment of water quality [4, 6]. To address these problems, there's a critical need for an alternative approach that ensures both computational efficiency and accuracy in estimating the WQI [10]. In recent years, machine learning (ML) techniques provides an alternate method to estimate the WQI based on existing data [4, 11].

The main goal of this project is to develop an efficient model to predict WQI based on water quality parameters such as temperature, conductivity, dissolved oxygen, pH.

The Environmental Resources Division (ERD) of the Public Works Department of the City of Cape Coral conducts routine monitoring of aquatic systems in the city. The water quality characteristics routinely measured at each station. Laboratory analyses are provided by the Chemistry Group of the Cape Coral Water Reclamation Plant and by staff of ERS. Staff members of ERD collected field instrument data. The relevant water quality data for Cape Coral were obtained from https://hub.arcgis.com/datasets/CapeGIS::water-quality-data/explore .

References

1. Rockström, J., et al., *Water resilience for human prosperity*. 2014: Cambridge University Press.
2. UNESCO, *The United Nations World Water Development Report 2021: Valuing Water*. 2021: United Nations.
3. Bui, D.T., et al., *Improving prediction of water quality indices using novel hybrid machine-learning algorithms.* Science of the Total Environment, 2020. **721**: p. 137612.
4. Ahmed, M., R. Mumtaz, and S.M. Hassan Zaidi, *Analysis of water quality indices and machine learning techniques for rating water pollution: A case study of Rawal Dam, Pakistan.* Water Supply, 2021. **21**(6): p. 3225-3250.
5. Lap, B.Q., et al., *Predicting water quality index (WQI) by feature selection and machine learning: a case study of An Kim Hai irrigation system.* Ecological Informatics, 2023. **74**: p. 101991.
6. Ahmed, M., R. Mumtaz, and Z. Anwar, *An Enhanced Water Quality Index for Water Quality Monitoring Using Remote Sensing and Machine Learning.* Applied Sciences, 2022. **12**(24): p. 12787.
7. Uddin, M.G., S. Nash, and A.I. Olbert, *A review of water quality index models and their use for assessing surface water quality.* Ecological Indicators, 2021. **122**: p. 107218.
8. Uddin, M.G., et al., *A comprehensive method for improvement of water quality index (WQI) models for coastal water quality assessment.* Water Research, 2022. **219**: p. 118532.
9. Uddin, M.G., et al., *A novel approach for estimating and predicting uncertainty in water quality index model using machine learning approaches.* Water Research, 2023. **229**: p. 119422.
10. Khoi, D.N., et al., *Using machine learning models for predicting the water quality index in the La Buong River, Vietnam.* Water, 2022. **14**(10): p. 1552.
11. Hassan, M.M., et al., *Efficient prediction of water quality index (WQI) using machine learning algorithms.* Human-Centric Intelligent Systems, 2021. **1**(3-4): p. 86-97.