

# EFFICIENT WATER QUALITY ANALYSIS AND PREDICTION USING MACHINE LEARNING

**Team Leader : NithyaSree N**  
**Team Member 1 : Sathiyapriya M**  
**Team Member 2 : Shri Janani M**  
**Team Member 3 : Yogashree D**

## LITERATURE SURVEY

**Title: Water Quality Prediction Based on Machine Learning  
Techniques**

*Author: Zhao Fu, Cheng Mei Yang, Jacimaria Batista and Yingtao  
Jiang*

Published in January 2020

### **Abstract**

The ANN model also has shortcomings though it can accurately predict water quality in some scenarios. The adaptive neuro-fuzzy inference system (ANFIS) has been proven to be an effective tool in formulating the complicated linear and non-linear relationship hidden in datasets. In this dissertation, several methods have been proposed to improve the performance of ANFIS-based water quality prediction models. Stratified sampling is employed to cover different kinds of data distribution in the training and testing datasets. Lastly, intelligence algorithms are used to optimize the parameters of membership functions in the ANFIS model to promote the prediction accuracy.

## **Title: Efficient Water Quality Prediction Using Supervised Machine Learning**

*Author: Umair Ahmed, Rafia Mumtaz, Hirra Anwar, Asad A. Shah, Rabia Irfan and José García-Nieto*

Published in October 2019

### **Abstract**

The alarming consequences of poor water quality necessitate an alternative method, which is quicker and inexpensive. With this motivation, 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 on the basis of the WQI. The proposed methodology employs four input parameters, namely, temperature, turbidity, pH and total dissolved solids. The proposed methodology achieves reasonable accuracy using a minimal number of parameters to validate the possibility of its use in real time water quality detection systems.

## **Title: Evaluation of Multivariate Linear Regression And Artificial Neural Networks in Prediction of Quality Parameters**

*Author: Hamid Zare Abyaneh*

Published in January 2014

### **Abstract**

This paper examined the efficiency of multivariate linear regression (MLR) and artificial neural network (ANN) models in prediction of two major water quality parameters in a wastewater treatment plant. Biochemical oxygen demand (BOD) and chemical oxygen demand (COD) as well as indirect indicators of organic matters are representative parameters for sewer water quality. Performance of the ANN models was evaluated using coefficient of correlation ( $r$ ), root mean square error (RMSE) and bias values. The computed values of BOD and COD by model, ANN method and regression analysis were in close agreement with their respective measured values. Results showed that the ANN performance model was better than the MLR model.

## **Title: Prediction of Water Quality System for Aquaculture using Machine Learning**

*Author: Kiran babu T S, Manoj Challa*

Published in June 2019

### **Abstract**

Water quality is a critical factor in the processing of aquatic organisms. In this paper, we created an application using ML to predict water boundaries and monitor fish ponds. The water level is predicted in an hourly manner to ensure the growth and survival of aquatic life. The web application is built using Flask to alert the user to critical situations. The impact of water parameter changes can be effectively treated if the information is analysed, and water quality is expected ahead of time.

## **Title: Water Quality Prediction Models Based on Machine Learning**

*Author: Rongli Gai, Jiahui Yang*

Published in May 2022

### **Abstract**

Water quality prediction is a research hotspot in the field of ecological environment, which is of great significance to the prevention of water pollution and the construction of automatic water quality monitoring network. Firstly, the background of water quality prediction and the development and research trends of water quality models at home and abroad are systematically introduced. Then, the water quality prediction method based on machine learning is mainly introduced, focusing on time series prediction method, regression analysis method, neural network method and combination prediction method. The applicability and limitations of the model are analysed respectively. Finally, according to the research history and present situation of water quality prediction model, the development trend of water quality prediction model is prospected.

## **Title: Water quality prediction based on Naive Bayes algorithm**

*Author: M. Ilic, Z.Srdjevic, B.Srdjevic*

Published in January 2022

### **Abstract**

In the fast-changing world with increased water demand, water pollution, environmental problems, and related data, information on water quality and suitability for any purpose should be prompt and reliable. One of such approaches is machine-learning (ML) based prediction. This paper presents the results of the application of the Naïve Bayes, a widely used ML method, in creating the prediction model. The proposed model is based on nine water quality parameters: temperature, pH value, electrical conductivity, oxygen saturation, biological oxygen demand, suspended solids, nitrogen oxides, orthophosphates, and ammonium. Forty-eight samples are used to train the model. Once trained, the Naïve Bayes model correctly predicted the class of water sample in 64 out of 68 cases, including cases with missing data. This recommends it as a trustful tool in the transition from traditional to digital water management.

## **Title: Multi-task learning framework for predicting water quality using non-linear machine learning technique**

*Author: D.Senthilkumar, D.George Washington, A.K.Reshmy, M.*

*Noornisha*

Published in April 2022

### **Abstract**

Predicting the quality of water is a very important issue in an ecosystem and it can be used to control the increase of water contamination. Existing water quality prediction model not focused on multi-target learning process simultaneously and not identifying the non-linear relationship between the features and target variables. Therefore, this study proposes a multi-task learning method dealing with multi-target regression using non-linear machine learning technique. Finally, experiments are conducted to build a prediction model based on the proposed methods to evaluate accuracy on water quality dataset. The experimental results indicate that our method increases the overall accuracy of the experimental dataset compared with the existing methods with the reduced number of significant features.

# **Title: Water Pollution Prediction Based on Deep Belief Network in Big Data of Water Environment Monitoring**

*Author: Li Liang*

Published in December 2021

## **Abstract**

Aiming at the problems that the traditional water quality prediction model is generally not high in prediction accuracy and robustness, a water pollution prediction using deep learning in water environment monitoring big data is proposed. Objective. To optimize and improve the prediction accuracy of the water quality prediction model. Firstly, in the water environment monitoring system, the Internet of Things big data technology is used to accurately sense and monitor the real-time data of sewage treatment equipment and sewage quality. Then, the deep belief network (DBN) is used to build the water pollution prediction model, and the collected sewage treatment data is analysed to predict the water quality status. Finally, particle swarm optimization algorithm is used to dynamically optimize the number of hidden layer neural units and learning rate in the DBN prediction model, which makes the prediction results more scientific and accurate.

# **Title: A study on water quality prediction by a hybrid CNN-LSTM model with attention mechanism**

*Author: Yurong Yang, Qingyu Xiong, Chao Wu, Qinghong Zou<sup>1</sup>, Yang Yu<sup>1</sup>, Hualing Yi<sup>1</sup>, Min Gao*

Published in June 2021

## **Abstract**

New approaches to solve such nonlinear problems need further research since the complexity of water quality data and they are easily affected by the noise. In this paper, we propose a water quality prediction model named CNN-LSTM with Attention (CLA) to predict the water quality variables. We conduct a case study on the water quality dataset of Beilun Estuary to predict pH and NH<sub>3</sub>-N. Linear interpolation and wavelet techniques are used for missing data filling and data denoising, respectively. The hybrid model CNN-LSTM is highly capable of resolving nonlinear time series prediction problems, and the attention mechanism captures longer time dependence. The experimental results show that our model outperforms other ones and can predict with different time lags in a stable manner.

# **Title: Smart Urban Water Quality Prediction System Using Machine Learning**

*Author: Bharath Singh J, Nirmitha S, Kaviya S S*

Published in August 2021

## **Abstract**

Quality of water serves as a powerful environmental determinant and a foundation for the prevention and control of waterborne diseases. The project aims to design a water quality prediction system using Machine learning based on the water standards suggested by BIS to prevent deaths due to water related diseases. The quality is predicted based on parameters such as pH, Temperature, TDS, Turbidity and Conductivity value. The dataset is pre-processed and split into test and training data. The data is fed into regression algorithm and been evaluated. Sensors that can measure the water parameters have also been implemented. A webpage interfaced with the Machine Learning model is created to upload sensor values and the corresponding water quality is predicted. This project can be used in urban areas to predict the quality of the drinking water thereby preventing the spread of diseases such as dysentery, typhoid, and cholera due to consumption of contaminated water.