LITERATURE SURVEY

Date	19 September 2022
Team ID	PNT2022TMID10672
Project Name	Project - Efficient Water Quality Analysis & Prediction using Machine Learning
Maximum Marks	4 Marks

TEAM MEMBERS:

- Mohammed Mujahid M(TL)
- Karthikeyan K
- Kishor M
- Rajesh V

ABSTRACT:

Water makes up about 70% of the earth's surface and is one of the most important sources vital to sustaining life. Rapid urbanization and industrialization have led to a deterioration of water quality at an alarming rate, resulting in harrowing diseases. Water quality has been conventionally estimated through expensive and time-consuming lab and statistical analyses, which render the contemporary notion of real-time monitoring moot. 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. Of all the employed algorithms, gradient boosting, with a learning rate of 0.1 and polynomial regression,

with a degree of 2, predict the WQI most efficiently, having a mean absolute error (MAE) of 1.9642 and 2.7273, respectively. Whereas multi-layer perceptron (MLP), with a configuration of (3, 7), classifies the WQC most efficiently, with an accuracy of 0.8507. 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.

LITERATURE SURVEY

[1] Drinking Water Quality in Capital City of Pakistan Shahid Mehmood1, Asif Ahmad1, Anwaar Ahmed1, Nauman Khalid2 and Tariq Javed3.

This study was designed to evaluate the quality of drinking water in selected areas of capital of Pakistan. Its adjacent city Rawalpindi. Drinking water samples collected from selected localities of Rawalpindi and Islamabad are analyzed for different water quality parameters such as pH, alkalinity, hardness, total dissolved solids, chloride, bicarbonates, sodium, potassium, calcium, magnesium, sulphates, phosphates, nitrates lead, copper, cadmium, cobalt, iron and zinc. Total viable count, coliforms, fecal coliforms and Escherichia coli were also part of the study.

Advantage:

• Needed in rapidly change in system.

Disadvantage:

• Drinking water quality is poorly managed and monitored.

• water pressure is low in Pakistan supply systems.

[2] Gazzaz, N.M.; Yusoff, M.K.; Aris, A.Z.; Juahir, H.; Ramli, M.F. Artificial neural network modeling of the water quality index for Kinta River (Malaysia) using water quality variables as predictors. Mar. Pollut. Bull.2012, 64, 2409–2420.

This article describes design and application of feed-forward, fully-connected, three-layer perceptron neural network model for computing the water quality index (WQI)(1) for Kinta River (Malaysia). The modeling efforts showed that the optimal network architecture was 23-34-1 and that the best WQI predictions were associated with the quick propagation (QP) training algorithm; a learning rate of 0.06; and a QP coefficient of 1.75. The WQI predictions of this model had significant, positive, very high correlation ($r=0.977$, $p<0.01$) with the measured WQI values, implying that the model predictions explain around 95.4% of the variation in the measured WQI values.

Advantage:

• Saving in bulk water requirement.

Disadvantage:

• possible long term degradation of water quality for possible saline intrusion .

Tirabassi

[3] Sakizadeh, M. Artificial intelligence for the prediction of water quality index in groundwater systems. .Model. Earth Syst. Environ. 2016.

A study was initiated to predict water quality index (WQI) using artificial neural networks (ANNs) with respect to the concentrations of 16 groundwater quality variables collected from 47 wells and springs in Andimeshk during 2006–2013 by the Iran's Ministry of Energy. Such a prediction has the potential to reduce the computation time and effort and the possibility of error in the calculations. For this purpose, three ANN's algorithms including ANNs with early stopping, Ensemble of ANNs and ANNs with Bayesian regularization were utilized. The application of these algorithms for this purpose is the first study in its type in Iran.

Advantage:

• Non toxic humans ,non residue left behind.

Disadvantage:

• Strong oxidizer may causes hydro genic minerals nutrients to precipitate, reducing bioavailability.

[4] Zhang, H. The optimality of naive Bayes. AA 2004.

Naive Bayes is one of the most efficient and effective inductive learning algorithms for machine learning and data mining. Its competitive performance in classification is surprising, because the conditional independence assumption on which it is based, is rarely true in real world applications. An open question is: what is the true reason for the surprisingly good performance of naive Bayes in classification? In this paper, we propose a novel explanation on the superb classification performance of naive Bayes.

Advantage:

• Naive Bayes is suitable for solving multi-class prediction problems.

• If its assumption of the independence of features holds true, it can perform better than other models and requires much less training data.

Disadvantage:

• Naive Bayes assumes that all predictors (or features) are independent, rarely happening in real life. This limits the applicability of this algorithm in real-world use cases.

, coliforms, fecal coliforms and Escherichia coli were also part of the study.

CONCLUSION

One of the most important resources for survival is water, and WQI measures the quality of water. Traditionally, one must undergo an expensive and time-consuming lab analysis to test the purity of the water. This study investigated a different machine learning approach to forecast water quality using basic, readily accessible water quality data.