Efficient Water Quality Analysis and Prediction using Machine learning

Introduction

Water is the most important of sources, vital for sustaining all kinds of life; however, it is in constant threat of pollution by life itself. Water is one of the most communicable mediums with a far reach. Rapid industrialization has consequently led to deterioration of water quality at an alarming rate. Poor water quality results have been known to be one of the major factors of escalation of harrowing diseases. As reported, in developing countries, 80% of the diseases are water borne diseases, which have led to 5 million deaths and 2.5 billion illnesses. The most common of these diseases in are diarrhoea, typhoid, gastroenteritis, cryptosporidium infections, some forms of hepatitis and giardiasis intestinal worms Water quality is currently estimated through expensive and time-consuming lab and statistical analysis, which require sample collection, transport to labs, and a considerable amount of time and calculation, which is quite ineffective. This research explores the methodologies that have been employed to help solve problems related to water quality. Also, employ various machine learning methodologies to predict water quality using minimal parameters and discuss the results of regression and classification algorithms, in terms of error rates and classification precision.

Literature Survey

[1] Shafi et al[1]estimated water quality using classical machine learning algorithms namely, Support Vector Machines (SVM), Neural Networks (NN), Deep Neural Networks (Deep NN) and K Nearest Neighbours (KNN), with the highest accuracy of 93% with Deep NN. The estimated water quality in their work is based on only three parameters: turbidity, temperature and pH, which are tested according to World Health Organization (WHO) standards (Available online at URL https://www.who.int/airpollution/guidelines/en/). Using only three parameters and comparing them to standardized values is quite a limitation when predicting water quality.

[2]In this research paper, the advanced artificial intelligence algorithms, namely, NARNET and LSTM were used to predict the WQI. Also, machine learning algorithms such as SVM, KNN AND Naive Bayes were used to classify the WQI data.

Disadvantage: The efficiency and robustness is very poor in predicting the WQI.

[3]In this paper, ten indicator parameters like Solids, chloramines, sulphate, conductivity, organic carbon, trihalomethanes, turbidity, pH value, hardness and potability were used to predict the water quality. To estimate, a set of representative supervised machine learning algorithms like Decision tree, K-Nearest Neighbour were used.

[4]This paper presents a novel spatio-temporal multi-view multi-task learning framework to forecast the water quality of a station by fusing multiple sources of urban data. It also uses Spatio-temporal view alignment to work toward local information aggregation for each station and global

prediction alignment, which incorporates the spatial correlations among stations and performs co-prediction over all stations using these correlations.

Disadvantage: Problem will arise in water quality inference in the urban water distributed systems.

[5] This study's objective is to create a water quality prediction model utilising Artificial Neural Networks (ANN) and time-series analysis to incorporate water quality parameters. Historical data on water quality are used in this study. Mean-Squared Error (MSE), Root Mean Squared Error (RMSE), and Regression Analysis are the performance evaluation metrics used to gauge how well the model is doing. ANN has received widespread recognition as a tool for classifying complicated information, including those pertaining to environmental dynamics. It can effectively explain the non-linear relationship between the intricate water quality statistics.

[6] The proposed framework has settled to an Intelligent constant IoT based Water Quality Monitoring framework which depends on Machine to Machine correspondence through AI. Turbidity and the conductivity sensors are connected. The conductivity acts as a sensor gateway. The sensor input are sent to the pi4, an edge level processor(personal computer) where in the K Means, a machine learning algorithm is used for predicting the quality of water. The predicted water quality data are stored in Cloud server for future access. The predicted data is sent to the water controller unit for further action.

References

[1] Surface Water Pollution Detection using Internet of Things – Shafi

[2] Water Quality Prediction Using Artificial Intelligence Algorithms – Theyazn H.H Aldhyani

[3]Water Quality Prediction using Machine Learning – Sai Sreeja Kurra, Sambangi Geethika Naidu, Sravani Chowdala, Sree Chithra Yellanki, Dr.B.Esther Sunanda

[4] Urban Water Quality Prediction Based on Multi-task Multi-view Learning – Ye Liu, Yu Zheng, Yuxuan Liang, Shuming Liu, David S. Rosenblum

[5] Machine learning methods for better water quality prediction - Ali Najah Ahmeda

[6] A Real Time Water Quality Monitoring Using Machine Learning Algorithm - S. Angel Vergina, Dr.S.Kayalvizhi ,Dr. R.M. Bhavadharini, Kalpana Devi. S