Ideation Phase
Literature Survey on The Selected Project & Information Gathering

Date	7 September 2022
Team ID	PNT2022TMID21972
Project Name	Project- Efficient Water Quality Analysis and Prediction using Machine Learning
Maximum Marks	2 Marks

TABLE:

S.NO	AUTHOR'S NAME	PROPOSED WORK
1.	Prasad, D. Venkata Vara, P. Senthil Kumar, Lokeswari Y. Venkataramana, G. Prasannamedha, S. Harshana, S. Jahnavi Srividya, K. Harrinei, and Sravya Indraganti.	Gathered water samples from different areas of Pakistan and tested them against different parameters using a manual lab analysis and found a high presence of E. coli and fecal coliform due to industrial and sewerage waste.
2.	Ajayi, Olasupo O., Antoine B. Bagula, Hloniphani C. Maluleke, Zaheed Gaffoor, Nebo Jovanovic, and Kevin C. Pietersen	Tested 46 different samples from Orange town, Karachi, using manual lab analysis and found them tone high in and total fecal coliform count. After getting familiar with the water quality research concerning Pakistan, we explored research employing machine learning methodologies in the realm of water quality
3.	Kenchannavar, Harish H.,	Estimated water quality

	Prasad M. Pujar, Raviraj M. Kulkarni, and Umakant P. Kulkarni	using classical machine learning algorithms namely, Support Vector Machines (SVM), Neural Networks (NN), Deep Neural Networks (Deep NN) and k Nearest Neighbors (KNN), with the highest accuracy of 93% with Deep. 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.
4.	Fonseca-Campos, Jorge, Israel Reyes-Ramirez, Lev Guzman-Vargas, Leonardo FonsecaRuiz, Jorge Alberto Mendoza-Perez, and P. F. Rodriguez-Espinosa.	Employed single feed forward neural networks and a combination of multiple neural networks to estimate the WQI. They used 25 waters quality parameters as the input. Using a combination of backward elimination and forward selection selective combination methods, they achieved an R2 and MSE of 0.9270, 0.9390 and 0.1200, 0.1158, respectively. The use of 25 parameters makes their solution a little immoderate in terms of an inexpensive real time system, given the price of the parameter sensors.

5.	Khan, Yafra, and Chai Soo See	Predicted the WQI using 16 water quality parameters and ANN with Bayesian regularization.
6.	Qian, Xueqing, Zhen Li, Zhaozong Meng, Nan Gao, and Zonghua Zhang	Predicted the dissolved oxygen (DO) using a feedforward neural network (FNN). They used 10 parameters to predict the DO, which again defeats the purpose if it has to be used for a Real time WQI estimation of an IOT system.

[1] Prasad, D. Venkata Vara, P. Senthil Kumar, Lokeswari Y. Venkataramana, G. Prasannamedha, S. Harshana, S. Jahnavi Srividya, K. Harrinei, and Sravya Indraganti. "Automating water quality analysis using ML and auto ML techniques." Environmental Research 202 (2021): 111720. This paper evaluates traditional and AutoML techniques within the avenue of water quality analysis by collecting the dataset from the Korattur Lake, Chennai. The dataset consists of observations of over a ten-year period, starting from 2009 until 2019. Under 9 parameters, around 5000 records are existent. The 9 parameters specified are Total Dissolved Solids (TDS), Turbidity, pH, Chemical Oxygen Demand (COD), Iron, Phosphate, Sodium, Chloride and Nitrate. From the preliminary stages, data proved to have a profound impact upon the both models. Use of SMOTE increased accuracy, reinforcing the fact that AutoML, efficient as it might be, provides better results when data is cleaned, handled and molded to suit the purpose. The factors such as time taken, academic experience required are all extremely less in the case of AutoML.

[2] Ajayi, Olasupo O., Antoine B. Bagula, Hloniphani C. Maluleke, Zaheed Gaffoor, Nebo Jovanovic, and Kevin C. Pietersen. "Water Net: A Network for Monitoring and Assessing Water Quality for Drinking and Irrigation Purposes." IEEE Access 10 (2022): 48318-48337. In this paper they have expressed water quality in terms of WQI (Water Quality Index) and IWQI (Irrigation Water

Quality Index). Collecting water samples from different sources, measuring the various parameters present, and bench-marking these measurements against preset standards, while adhering to various guidelines during transportation and measurement can be extremely daunting. This uses network architecture to collect data on water parameters in real-time and use Machine Learning (ML) tools to automatically determine suitability of water samples for drinking and irrigation purposes. The developed monitoring network is based on LoRa and takes the land topology into consideration. Results of the test showed that LR performed best for drinking water, as it gave the highest classification accuracy and lowest false positive and negative values, while SVM was better suited for irrigation water.

[3] Kenchannavar, Harish H., Prasad M. Pujar, Raviraj M. Kulkarni, and Umakant P. Kulkarni. "Evaluation and Analysis of Goodness of Fit for Water Quality Parameters using Linear Regression through the Internet of Things (IoT) based Water Quality Monitoring System." IEEE Internet of Things Journal (2021). In this paper they have used IoT help to obtain real-time data, in the river basin region. To implement this, we make use of WQM system. WQM consists of sensors such as T, pH, dissolved oxygen (DO), electrical conductivity (EC), biochemical oxygen demand (BOD), NO3, and total dissolved solids (TDSs) to monitor water quality. The Smart WQM is used for ecological monitoring of the water environment. An IoT system based on low-cost Raspberry Pi for WQM that controls the flow of water. The monitored parameters are physicochemical parameters. The WQM uses linear regression that helps to estimate the relationship between two parameters. After linear regression apply one-way ANOVA to the samples. It is used to compare two or more sample means by the F distribution method. Overall, we can see that all of the water quality parameters are within the normal range prescribed, and the water can be used for daily purposes.

[4] Fonseca-Campos, Jorge, Israel Reyes-Ramirez, Lev Guzman-Vargas, Leonardo FonsecaRuiz, Jorge Alberto Mendoza-Perez, and P. F. Rodriguez-Espinosa. "Multiparametric System for Measuring Physicochemical Variables Associated to Water Quality Based on the Arduino Platform." IEEE Access 10 (2022): 69700-69713. In this paper they have used pH, ORP, turbidity and TDS sensors provide an analog output. A 16-bit-ADC module increases the resolution of 10 bits offered by the Arduino Mega native ADC. ORP is an electrochemical parameter that is measured similarly to pH, but the electrode uses a noble metal as a measurement

element. TDS provides a measure of the water salinity, and it is related to the EC of water. Turbidity is an optical property describing how much light is scattered for a water sample. An IR light source like LED, sends light into a water sample. The Arduino Mega has an ADC of 10 bits. Incorporating the module ADS1115 from Adafruit, having a 16-bit ADC with a programmable gain amplifier improves the system resolution. Dissolved oxygen provides the magnitude of the oxygen gas dissolved in water. Overall, the system exhibited a good performance with low-costand readily available elements.

[5] Khan, Yafra, and Chai Soo See. "Predicting and analyzing water quality using Machine Learning: a comprehensive model." In 2016 IEEE Long Island Systems, Applications and Technology Conference (LISAT), pp. 1-6. IEEE, 2016. In this paper they have developed a water quality prediction model with the help of water quality factors using Artificial Neural Network (ANN) with Nonlinear Autoregressive (NAR) time series. This time series has been used with Scaled Conjugate Gradient (SCG) as training algorithm. Time Series Data are collected from United States Geological Survey (USGS) online resource called NWIS. The samples include the data ranging from January to March 2014, with 6-minute time interval. Four water quality factors Turbidity, Dissolved Oxygen Concentration, Chlorophyll and Specific Conductance have been measured using four ANN models. The performance of 4 models have been analyzed using Mean Square Error (MSE) and Root Mean Square Error (RMSE). The ANN-NAR model providesbest accuracy with lowest MSE od 3.7x10^-4 for turbidity and best Regression Value for Specific Conductance(0.99).

[6] Qian, Xueqing, Zhen Li, Zhaozong Meng, Nan Gao, and Zonghua Zhang. "Flexible RFID tag for sensing the total minerals in drinking water via smartphone tapping." IEEE Sensors Journal 21, no. 21 (2021): 24749-24758. In this project, they have designed and implemented RFID sensor tag for evaluating total minerals in drinking water. The sensor reading can be obtained through smartphone tapping and the results are received in 1 second The reading range between smart phone andsensor tag is 1-3cm.The developed RFID sensors exhibits particular superiority in flexibility and convenience of use due to advantages in wireless power, data transfer, no added hardware and software for smartphones.