ANALYSIS AND CLASSIFICATION OF WATER PURITY BASED ON PARAMATERS STATED BY CENTRAL POLLUTION CONTROL BOARD

[1] Bharath HG, [2] Greeshma S, [3] Kavya D, [4] Amith JS, [5] Aditya Pai H,

[1,2,3,4] Undergraduate student, Computer Science & Engineering, K S Institute of Technology,

[5] Associate Professor, Department of Computer Science & Engineering, K S Institute of Technology

Bangalore, India.

ABSTRACT Water from various sources are stored in lakes and used for multiple purposes after a prolonged storage time. The main purpose of this paper is to predict the stored water status using KNN algorithm technique. Quality of water affects public health, growth and yield of the plants. Water quality is measured based on its chemical, physical, and biological parameters. Water quality of two samples one is Yediyur Lake in Bengaluru and underground water in Raghuvanahalli region in Bengaluru where TDS and pH values are analyzed. The paper discusses about the initial findings of the research to check the Total Dissolved Solvents and Potential of Hydrogen in the Yediyur lake region and Raghuvanahalli region

Keywords— TDS, pH, KNN algorithm.

1.INTRODUCTION

One of the vital water resources in India is rivers, lakes and ground water. Though the lake water is not used directly for drinking or any other domestic purpose, the quality of lake water affects the quality of ground water in the nearby areas. The groundwater is also being vastly used in many cities and villages for water consumption and other domestic purposes. This proposed model predicts the water quality based on the chemical, physical and biological parameters like PH and Total Dissolved Solvents using KNN algorithm. As the water quality changes after every day of storage in the lake due to exposure to sunlight, there is an inconsistency in the data. Thus a KNN algorithm is used for this application. The sample of data is collected at different time intervals and are analyzed for the difference in the pH value and TDS.

2.RELATED WORK

In [1] The investigation of water source quality in Taihu Lake is being examined. The author describes the development of drinking water treatment processes. According to the paper the water samples were taken from two water source areas in Taihu Lake and were analyzed. The results showed that ammonia, TN, TP and COD Mn were the dominant contaminants. The concentrations of those contaminants all exceeded “surface water environmental quality standards(GB3838-2002)” and seasonally varietied. Turbidity and density of algae displayed similar variation trends all year around and reached the peak values in July and August.

In [2] Based on the result of suitable filter screening, A recirculation filtration process with multi-medias was applied to investigate the reactor performance in treating reclaimed water resource supply-type lake water. The results showed that the pollutant removal capabilities for single species packing filter was in the following order: activated carbon>;composite inorganic medium>;volcanic rock>;quartz sand>;gravel. A pilot scale test has been carried out using the recirculating filtration process with volcanic rock, composite inorganic medium and activated carbon media. And the removal efficiencies of CODCr, TP, Chlorophyll a., UV254, UV410 and Turbidity attained 66%, 70%, 90%, 65%, 72% and 83% respectively, while the circle period was 14 days and the filtration rate was 10m/h~13m/h. Moreover, the composition of dissolved organic matter (DOM) was analyzed by the excitation-emission matrix (EEM). The fluorescent components in the lake water DOM were mainly aromatic protein (Peak A and Peak E), humic-like fluorophores (Peak C) and fulvic-like fluorophores (Peak D), in which the aromatic protein fluorophores A and E occurred at 230/344nm and 280/360nm, respectively. The fluorophores intensity of the characteristic fluorescence peaks varied significantly indicating that the concentration of DOM varied with the wastewater treatment process.

In [3] Intelligent IoT based Water Quality Monitoring System developed involves microcontroller and processing unit. In addition, sensors are deployed in water and connected to microcontroller to measure necessary parameters, these parameters are sent to processing unit for analysis where an ML model is loaded which helps in classifying the data.

3.METHODOLOGY

To begin with, we make an application wherein the user needs to select if the water being tested is drinking water, irrigation water, or any other consumption classes. This is done to alert the user in case of a change in water parameters while testing. If the user is unaware of the water consumption class, he can skip the step. To set up the project, we need to get a microcontroller that is capable of directing values from measuring probes to a processing unit. We are going to use Arduino UNO as our microcontroller. We use different probes that are available in market to detect the values for parameters such as pH, BOD, Nitrate as NO3, Ammonia as N, and Boron as B. All these values in specific combination is classified under different category such as drinkable water, bathing water, irrigation use, etc, as per Centre Pollution Control Board (CPCB). We have a central processing unit, preferably a Raspberry pi board which is loaded with trained ML model that can classify the querying data sent by

Arduino. ML model is trained using KNN Algorithm using the set of values that already exist in the CPCB database. Each category/consumption class of water is identified using a set of proportionate parameters. Each parameter is plotted on its own axis. That is, if we have 5 parameters, then we have 5 different axes for them.

From the training data, we plot a point on the 5-dimensional plane and label it with the concurrent consumption class. Now when the querying data is loaded onto processing board, the latter plots the point on the 5-dimensional plane and calculates the distance with its K neighbours. The majority/repetitive consumption class among the neighbours will be selected as the label for querying data and hence gets classified. The classified value along with the querying data is stored in NoSQL database along with a timestamp for future use. If the classification turns out to be other than what the water is intended to be used for, then a notification is sent out in the form of an SMS and a mail to the user who filled the application in the beginning.

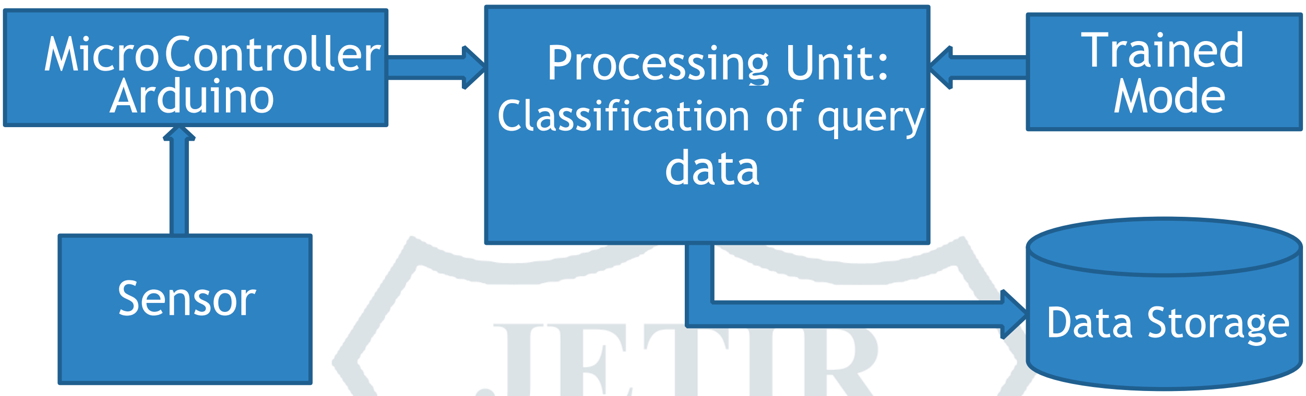
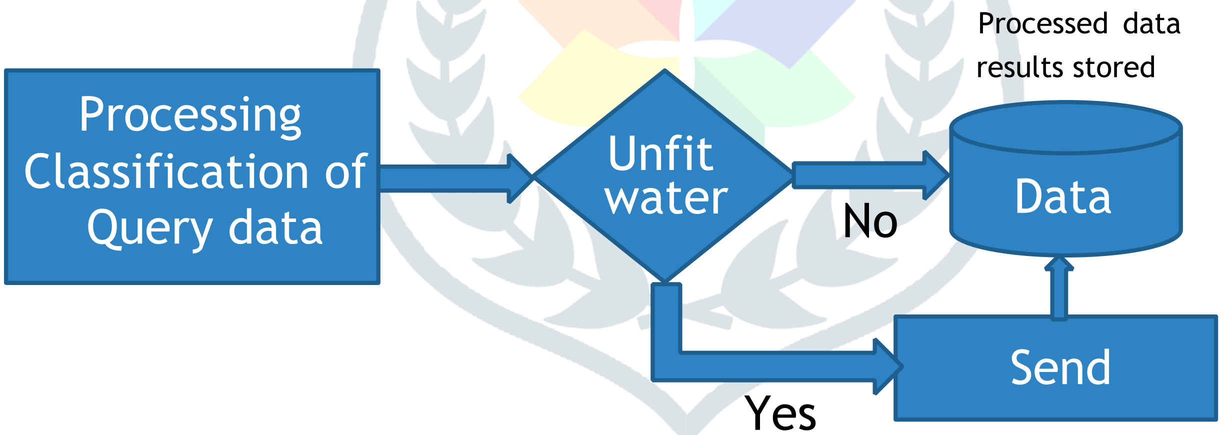


Fig. 1. Explains the working of existing system

Expected Outcome:

We get the class of water and we label the sample accordingly. In case if unfit water is found to be supplied to residential places, code can raise alarms by sending mails/sms to concerned authority.



3.1 K-Nearest Neighbor (k-NN) approach

The k-NN approach was first described by Nemes et al. [4] in the prediction of solvent contents. For the convenience of the reader, the k-NN approach as summarized in Seybold et al. is presented here. There are no predefined mathematical functions that estimate TDS and pH. A "reference dataset" is searched for pH and TDS and is calculated:

$$d_i = \sqrt{\sum_{j=1}^{x} \Delta a_{ij}^2} \tag{1}$$

Where d_i is the distance of the ith water from the sample water, and Δa_{ij} is the difference of the ith water from the sample water in the j^{th} water attributes. In the present study, there are more than two input attributes (e.g., TDS and pH), so the attribute values are normalized before they are used to calculate "distance", which avoids bias towards one or more attributes. As a result, temporary variables are generated with a distribution having a zero mean and a standard deviation of one using the following transformation from:

$$a_{ij(temp)} = \left[\left(a_{ij}\right) - \bar{a}_j\right] / \sigma\left(a_j\right) \qquad (2)$$

Where a_{ij} represents the value of the j^{th} attribute of the ith soil, and $\bar{a}_j$ and $\sigma(a_j)$ represent the mean and standard deviation of the observed values of the jth attribute in the reference data set. Then, the minimum-maximum range of those temporary variables, are scaled to obtain zero mean and the same minimum-maximum range in the data of all attributes:

$$a_{ij(trans)} = a_{ij(temp)} \left\{ MAX \begin{bmatrix} range\left(a_{j=1(temp)}\right), \ldots, \\ range\left(a_{j=x(temp)}\right) \end{bmatrix} \right\} / range\left(a_{j(temp)}\right) \qquad (3)$$

Where $a_{j(temp)}$ represents the data of the jth water attributes normalized using Eq. 2; and $a_{ij(trans)}$ represents the final transformed value of the j^{th} attribute of the ith water that are to be used as input [4]. It should be noted that taxonomic order is a strata within the reference dataset and was not normalized. Within water order the continuous attributes were normalized.

The closest 10 water (k) in the reference dataset were then used to formulate the estimate of the output SBs. It was shown by Nemes et al. [4] that *k* was not very sensitive to reference data set size as long as *k* was above 8 or 9 (in their particular case). A *k* of 10 was successfully used in a k-NN model for predicting soil bulk density [4]. We felt there was no reason to alter *k*; therefore a *k* of 10 was used here in the present study.

Nemes et al. [4] presented the argument that a water closer to the target should have more weight in calculating the estimated value. Therefore, their distance-dependent weighting system was employed here to account for the distribution of the distances of the selected water samples to the target:

$$d_{i(rel)} = \left(\left[\sum_{i=1}^{k} d_i\right] / d_i\right)^p \qquad (4)$$

Where *k* is the number of neighbors considered, and w_i is the assigned weight, and $d_{i(rel)}$ is the relative distance of the ith selected neighbor, calculated as:

$$d_{i(rel)} = \left(\left[\sum_{i=1}^{k} d_i\right] / d_i\right)^p \qquad (5)$$

Where *k* is the number of neighbors considered, and *di* is the distance of the ith selected neighbor calculated using Eq. 1, and *p* is a power term that was set to one for the estimation of SBs in the present study. Nemes et al. has shown that *p* values remained around one and were generally insensitive to sample size (in their particular case), suggesting that "one" is a safe first approximation. Attribute weights were applied equally across the entire data space. The final result of collected water samples.

3.2 Water quality parameters:

The following parameters were selected to analyse the water quality as per the guidelines given by the central pollution control board (CPCB).

3.2.1 pH

pH is a parameter used to measure the acidic content of a water. Every increase of one on the scale is actually an increase of 10 times more acid. Most plants and animals cannot survive in water that has a pH less than 5. It is too acidic. Ideal water pH is 6.7to 8.6.

3.2.2 TDS

TDS stands for **total dissolved solids**, and represents the total concentration of dissolved substances in **water**.Common inorganic salts that can be found in water include calcium, magnesium, potassium and sodium, which are all cations, and carbonates, nitrates, bicarbonates, chlorides and sulfates, which are all anions.

4.RESULTS AND DISCUSSION

Property	N	Min	Max	Median	Mean	Std. Dev.
pH (H_2O)	36,910	2.1	10.1	5.7	5.8	0.9
pH ($CaCl_2$)	33.838	2.2	9.4	5.2	5.2	0.9
CEC ($cmol_{(+)}$ kg^{-1})	36,910	0.1	196	13.8	16.5	13.2
ECEC ($cmol_{(+)}$ kg^{-1})	14,810	<0.1	152	6.0	8.7	8.0

5.CONCLUSION

Traditional water quality monitoring involves three steps namely water sampling, Testing and investigation. These are done manually by the scientists. This technique is not fully reliable and gives no indication before hand on quality of water.

Now with the advent of Machine to Machine Communication which leads to devices communicating among themselves and accordingly analyzing the data intelligently, we here have developed an "Intelligent IoT based water quality analysis system" pertaining to storage tanks being used by residential areas.

The goal of developing this model is to make aware of general public of the pureness of water they are drinking and to caution them whenever required.

We ensure that the quality of water is monitored so that the contaminated water does not affect the public health and growth

6.ACKNOWLEDGEMENT

The successful project execution would have not been possible without the people who made it possible and whose constant guidance crowned our effort with success. We take this opportunity to express our sincere gratitude to Management K S Institute of Technology, Bengaluru. We would like to express our gratitude to Dr. K.V.A. Balaji C.E.O. K.S. Institute of Technology, Bengaluru, for facilitating us to build and present the project. We would like to extend our gratitude to Dr.T.V.Govindaraju, Principal/Director,
K.S. Institute of Technology, Bengaluru, affiliated by VTU for providing opportunity to publish this paper.

We thank Dr. Rekha.B.Venkatapur, Professor and Head, Department of Computer Science and Engineering, K.S. Institute of Technology, Bengaluru, for her encouragement.
We would also like to thank, Mr. K. Venkata Rao, Associate Professor, Department of Computer Science and Engineering, K.S. Institute of Technology, Bengaluru, for his constant guidance and inputs.

We wholeheartedly thank our project mentor Mr.Aditya Pai H, Assistant Prof., Department of Computer Science and Engineering, K.S. Institute of Technology, Bengaluru, for his support and guidance.
Finally, we would like to thank all the teaching and non-teaching staff of the college for their cooperation.
Moreover, I thank all my family and friends for their invaluable support and cooperation.

REFERENCES

[1] "Water quality of water source area in Taihu lake and effect on water treatment process" Published in: 2011 International Conference on Consumer Electronics, Communications and Networks

[2] "Research on water purification for reclaimed water resource supply-type lakes by the method of recirculation filtration" Published in: 2011 International Conference on Remote Sensing, Environment and Transportation Engineering

[3] Karthick.T and Suresh Sankaranarayanan Department of Information Technology, SRM University, Kattankaluthur Campus, Chennai, India. International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, Number 16 (2017)

[4] "Soil Survey: Prediction of Sum of Bases Using K-Nearest Neighbor Approach" Published by: Cathy A Seybold*, Moustafa Elrashidi and Zamir Libohova