Journal of Science and Technology
ISSN: 2456-5660 Volume 8, Issue 07 (July -2023)
www.jst.org.in *DOI:https://doi.org/10.46243/jst.2023.v8.i07.pp58 - 63*

Water net:A Network For Monitoring And Assessing Water Quality

Mrs.Supriya Belkar [1],Algot Yogitha [2] ,Alluri shruthi [3] ,Chintala Sindhuta [4] ,Sowmya [5]
[1]Assistant Professor,Computer Science and Engineering,Malla Reddy Engineering College for Women,Secundrabad,India.
[2 3 4] UG Scholar,Department of CSE, Malla Reddy Engineering College for Women, Hyderabad, Telangana, India
Supriyabelkar27@gmail.com
algotyogitha@gmail.com, allurishruthireddy@gmail.com,
sindhuthareddy@gmail.com, chowdaboinasowmya17@gmail.com

To Cite this Article
Mrs.Supriya Belkar ,Algot Yogitha ,Alluri shruthi ,Chintala Sindhuta ,**Sowmya , Water net:A Network For Monitoring AndAssessing Water Quality**" *Journal of Science and Technology, Vol. 08, Issue 07,-July 2023, pp58-63*
Article Info
Received: 26-06-2023 Revised: 28-06-2023 Accepted: 10-07-2023 Published: 18-07-2023

***Abstract*—** Water is a fundamental requirement for human, animal, and plant survival. Despite its importance, quality water is not always fit for drinking, domestic and/or industrial use. Numerous factors such as industrialization, mining, pollution, and natural occurrences impact the quality of water, as they introduce or alter various parameters present therein, thus, affecting its suitability for human consumption or general use. The World Health Organization has guidelines which stipulate the threshold levels of various parameters present in water samples intended for consumption or irrigation. The Water Quality Index (WQI) and Irrigation WQI (IWQI) are metrics used to express the level of these parameters to determine the overall water quality. Collecting water samples from different sources, measuring the various parameters present, and bench-marking these measurements against pre-set standards, while adhering to various guidelines during transportation and measurement can be extremely daunting

I. INRODUCTION

Access to water is a critical component of human lives and is now considered a basic human right. Water is also important in agriculture and food production. Recent statistics shows that about 10% of the world population is malnourished, with developing countries being hit the hardest, with starvation resulting in about 45% of infant mortality. There are several sources of water for both drinking and irrigation use, including rivers, streams, rain, and groundwater. Several models have been developed to assess water quality, all of which consider various parameters, including chemical (such as hydrogen potential (pH), calcium, oxygen, sulphate levels etc.), microbial (such as E. coli, rotaviruses, Entamoeba etc.), and physical (temperature and clarity). These models producea unit metric, known as the Water Quality Index (WQI), as output.

The output of these processes indicates if the water sample is potable or non-potable. In this work, we propose a Cyber- physical network architecture for real-time monitoring of water parameters across a city and an alternative model based on machine learning to determine potability of water samples.

our work also only focuses on the physical and chemical parameters of water, while ignoring the biological. This is because our model is meant to be sensor based (in the context of the Internet of Things), and to our knowledge, there are no physical sensors for measuring biological parameters.

II. LITERATURE SURVEY

Water monitoring network: In a network for measuring and monitoring water parameters in a metal producing city in Brazil was developed. Twelve water monitoring stations were setup to measure several physico-chemical water parameters, including pH, dissolved solids, Zinc, Lead etc. When assessing the quality of drinking water, the Water Quality Index (WQI) has been the de facto metric. It is a unitless numeric value that gauges the suitability of water for human consumption or general usage. As stated earlier, several models exist for calculating WQI depending on the location and environmental conditions in such locations. Irrigation water is a vital part of food production, especially crop farming. The quality of water can affect crop yield, hence concerted efforts need to be made to ensure proper water quality standards . Like with drinking water, several classical techniques exist for ascertaining the quality of irrigation water ,however most are either tailored to drinking water alone or not economically viable for local farmers as theyrequire many parameters.

III. METHODOLOGY

It readitionally water quality monitoring has relied on manual sampling and laboratory analysis.However,this approach has limitations including high costs,time delays in obtaining results,and the inability to capture real-time changes in water quality.Environment protection,public Health,Resource Management,Early warning systems,policy and regulations. Network can provide real-time data,improve the efficiency of water quality monitoring and contribute to the sustainable use and protection of our water resources. Build a network for real-time collection and monitoring of water quality across water storage dams in the city of Cape Town. This network takes into consideration the unique geographical features of Cape Town, such as mountains and elevations that might obstruct radio frequency propagation.
Curate ample sized datasets on drinking and irrigation water that can be used to train (and test) machine learning models to automatically determine the `fitness for use" of a sample of water for drinking and/or irrigation purposes. Build models that determine the most critical parameters that influence the accuracy of machine learning models in analyzing water for drinking or irrigation

An Artificial Neural Network (ANN) which contains multiple layers between the input and output layer is called Deep Neural Network.

C. KNN CLASSIFIER

Simple, but a very powerful classification algorithm

- Classifies based on a similarity measure
- Non-parametric
- Lazy learning
- Does not “learn” until the test example isgiven

Whenever we have a new data to classify, we find itsK-nearest neighbors from the training data

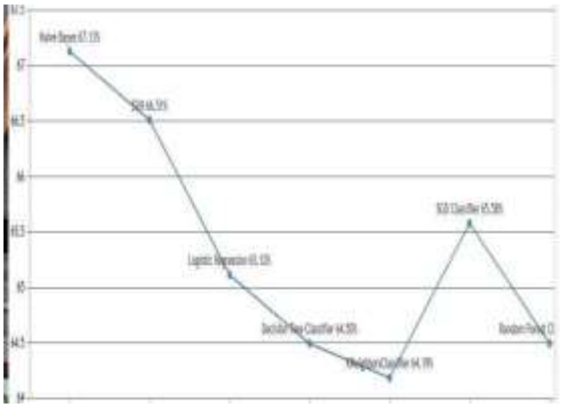


FIG 1: proposed methodology

IV. TYPES OF CLASSIFICATION ALGORITHMS

Many machine learning algorithms are being used in various

Example

- Training dataset consists of k-closest examplesin feature space
- Fature space means, space with categorization variables (non-metric variables)
- Learning based on instances, and thus also works lazily because instance close to the input vector for test or prediction may take time to occur in the training dataset

fields of research to help in solving the real-world problems. Mostly used machine learning classification algorithms are discussed below:

A. SUPPORT VECTOR MACHINE(SVM)

In classification tasks a discriminant machine learning technique aims at finding, based on an *independent and identically distributed* (*iid*) training dataset, a discriminant function that can correctly predict labels for newly acquired instances. Unlike generative machine learning approaches, which require computations of conditional probability distributions, a discriminant classification function takes a data point *x* and assigns it to one of the different classes that are a part of the classification task. Less powerful than generative approaches, which are mostly used when prediction involves outlier detection, discriminant approaches require fewer computational resources and less training data, especially for a multidimensional feature space and when only posterior probabilities are needed. From a geometric perspective, learning a classifier is equivalent to finding the equation for a multidimensional surface that best separates the different classes in the feature space.

B. DECISION TREE CLASSIFIER

Decision tree classifiers are used successfully in many diverse areas. Their most important feature is the capability of capturing descriptive decision-making knowledge from the supplied data. Decision tree can be generated from training sets. The procedure for such generation based on theset of objects (S), each belonging to one of the classes C1, C2, …, Ck is as follows:

Step 1. If all the objects in S belong to the same class, for example Ci, the decision tree for S consists of a leaf labeled with this class

Step 2. Otherwise, let T be some test with possible outcomesO1, O2…, On. Each object in S has one outcome for T so thetest partitions S into subsets S1, S2… Sn where each object in Si has outcome Oi for T. T becomes the root of the decisiontree and for each outcome Oi, we build a subsidiary decision tree by invoking the same procedure recursively on the set Si.

D.*NAIVE-BAYES*

The naive bayes approach is a supervised learning method which is based on a simplistic hypothesis: it assumes that the presence (or absence) of a particular feature of a class is unrelated to the presence (or absence) of any other feature .

Yet, despite this, it appears robust and efficient. Its performance is comparable to other supervised learning techniques. Various reasons have been advanced in the literature. In this tutorial, we highlight an explanation based on the representation bias. The naive bayes classifier is a linear classifier, as well as linear discriminant analysis, logistic regression or linear SVM (support vector machine). The difference lies on the method of estimatingthe parameters of the classifier (the learning bias).

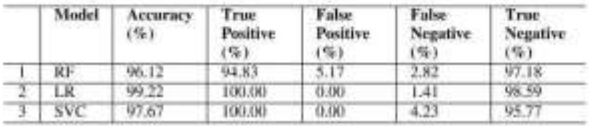
E. LOGISTIC-REGRESION-CLASSIFIERS

Logistic regression analysis studies the association between a categorical dependent variable and a set of independent (explanatory) variables. The name *logistic regression* is used when the dependent variable has only two values, such as 0 and 1 or Yes and No. The name *multinomial logistic regression* is usually reserved for the case when the dependent variable has three or more unique values, such as Married, Single, Divorced, or Widowed. Although the type of data used for the dependent variable is different from that of multiple regression, the practicaluse of the procedure is similar.

V. EXPERIMENTAL RESULTS

Table shows that all 3 models performed well w.r.t accuracy scores. RF had the least accuracy at 96.12% and, though impressive, had the highest False Negative (FP) rate at 5.17%.This implies that RF misclassified hazardous water samples assafe for drinking about 5% of the time. LR and SVC on the other hand resulted in FP values of 0% and are thus better alternatives for RF. However, SVC had a False Negative (FN)rate of 4.23%, implying that it misclassified some potable water samples as not drinkable. LR gave the best results of the3 models with 99.22% classification accuracy and 1.41% FN. In essence, LR only misclassified safe drinking water as non- potable about 1.5% of the time.

DRINKING WATER:



	Model	Accuracy (%)	True Positive (%)	False Positive (%)	False Negative (%)	True Negative (%)
1	RF	96.12	94.83	5.17	2.82	97.18
2	LR	99.22	100.00	0.00	1.41	98.59
3	SVC	97.67	100.00	0.00	4.23	95.77

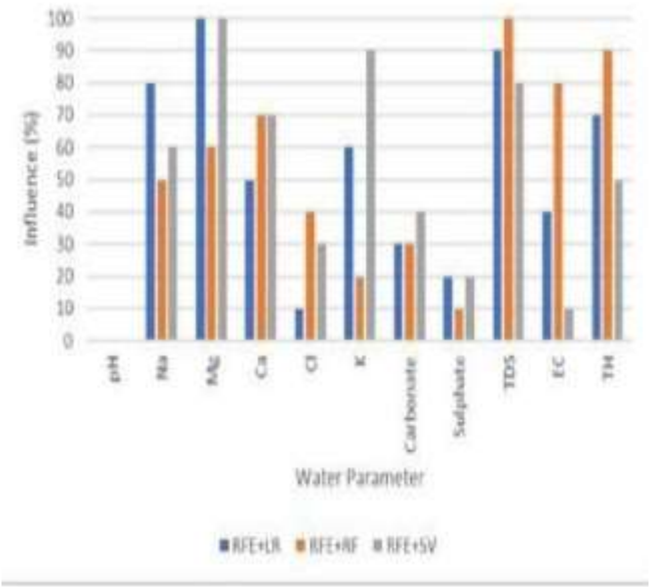
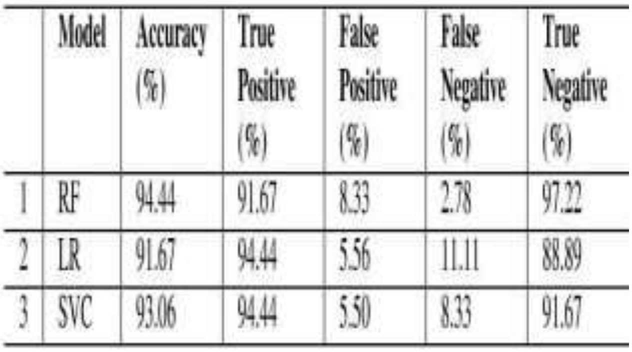


Figure shows a graphical depiction of the result of carrying out RFE on each of the models considered, that is, RFE on LR (RFE+LR), RFE on RF (RFE+RF), and RFE on SVC
(RFE+SV). The result, though non-uniform, revealed that pHwas the least influential parameter across board.

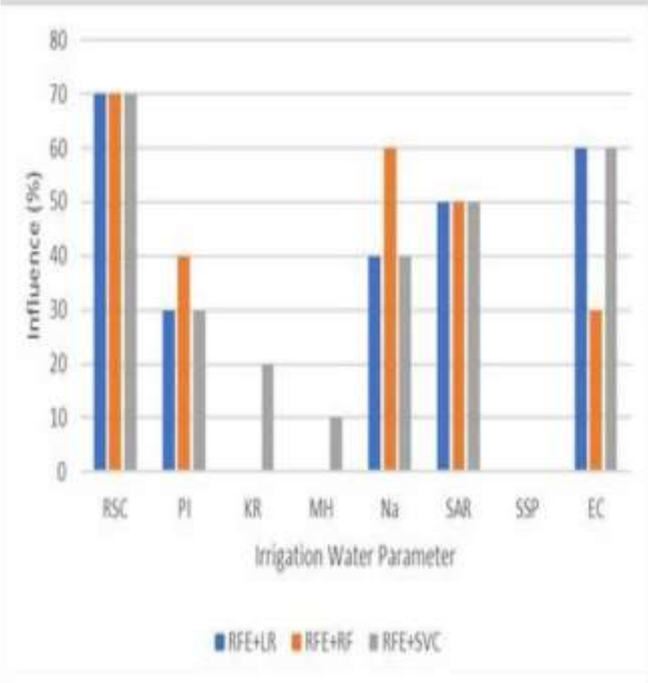
To further examine the influence of different combinations ofparameters on the classification accuracies of each model, weran iterative experiments using all possible combinations of parameters. For each iteration we held one parameter constantand cycled through the other 10. Table 7 summarizes the results of the top 40 combinations for LR, RF and SVC respectively. For each model, the table shows the resulting classification accuracies when at least two water parameters are removed from the dataset.

IRRIGATION WATER:



	Model	Accuracy (%)	True Positive (%)	False Positive (%)	False Negative (%)	True Negative (%)
1	RF	94.44	91.67	8.33	2.78	97.22
2	LR	91.67	94.44	5.56	11.11	88.89
3	SVC	93.06	94.44	5.50	8.33	91.67

:



The Similar to the results on Table, Table also shows that RF performed the worst of all three models w.r.t. to FP with a score of 8.33%. The same trend as in Table 6 is also observed for LR and SVC, with both having the lower FP rates of 5.56% and 5.50% respectively. However, in contrast to the results of the drinking water dataset, LR performed the worst w.r.t False Negative (FN) at 11.11%. The effect of FN are not as adverse on health as FP, hence, SVC would be considered the best option for irrigation water, as it gave acceptably high classification accuracyand the lowest False Positive value.

Graphical depiction of the results of recursive feature elimination (RFE+LR, RFE+RF, and RFE+SVC) on the irrigation water dataset. It reveals that SSP had the least influence on the classification accuracies of the models, while RSC was the most influential feature (water parameter). SAR and Na were also relatively influential across board. EC is RFE+LR and RFE+SVC but not with RFE+RF, yet the reverse is the case with Na. These contrasting influences are. most likely responsible for the lower false positive values observed with LR and SVC compared to RF, and the lower false negative values of RF compared to LR and SVC on Table . Table summarizes the results of the top 20 combinations of parameters influencing LR, RF and SVC when used on irrigation water.

VI.CONCLUSION AND FUTURE WORK

This work focused on two major concept, firstly, the proposal of a real-time water monitoring network for gathering data on water parameters from water bodies. Secondly, the application of machine learning (ML) models as means of assessing water quality. The developed water monitoring network is based on LoRa, a low power long range protocol for data transmission, and was developed using the City of Cape Town as case study. Results of the simulation done in Radio Mobile, revealed a partial mesh network topology as the most adequate network to cover the city. Data gathered from this monitoring network would ideally be aggregated on a Cloud server, where ML models can then be applied to assess the water's fitness of use for drinking or irrigation purposes. Due to the absence of relevant datasets, two suitable datasets were built in this work and used to training and testing three ML models considered, which are Random Forest (RF), Logistic Regression (LR) and Support Vector Machine (SVM).

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. Finally, a model for identifying the most influential water parameter(s) w.r.t classification accuracies of the ML models was then explored using recursive feature elimination (RFE). Obtained results showed that pH, and total hardness were the least influential parameters in drinking water, while SSP was the least for irrigation water.

Though the authors acknowledge the possible application of deep learning models, these were not used in this work. In future works, deep learning models such as the various variants of neural networks could be considered as expansion to this work. Furthermore, water quality indices were manually calculated and used to assess the ‘‘fitness for use’’ of water, future works could explore the application of unsupervised ML models as alternatives to manually calculated water quality indices. In the same vein, rather than using RFE, other approaches such as multi criteria decision making could also be considered to identify influential parameters. Finally, incorporating usage prediction models and microbial monitoring into the water network as well as tracking sources of water contaminates could also be avenues

to further this work.

REFERENCE

[1] B. X. Lee, F. Kjaerulf, S. Turner, L. Cohen, P. D. Donnelly,

R. Muggah, R. Davis, A. Realini, B. Kieselbach, L. S. MacGregor, I. Waller, R. Gordon, M. Moloney-Kitts, G. Lee, and J. Gilligan, ‘‘Transforming our world: Implementing the 2030 agenda through sustainable development goal indicators,’’

J. Public Health Policy, vol. 37, no. S1, pp. 13–31, Sep. 2016.

[2] Integrated Approaches for Sustainable Development Goals Planning: The Case of Goal 6 on Water and Sanitation, U. ESCAP, Bangkok, Thailand,2017. [3] WHO. Water. Protection of the Human Environment. Accessed: Jan. 24, 2022. [Online]. Available: www.afro.who.int/health-topics/water [4] L. Ho, A. Alonso, M. A. E. Forio, M. Vanclooster, and P. L. M. Goethals, ‘‘Water research in support of the sustainable development goal 6: A case study in Belgium,’’ J. Cleaner Prod., vol. 277, Dec. 2020, Art. no. 124082. [5] Global Nutrition Report 2016: From Promise to Impact: Ending Malnutrition by 2030, International Food Policy Research Institute, Washington, DC, USA, 2016, doi: 10.2499/9780896295841. [6] N. Akhtar, M. I. S. Ishak, M. I. Ahmad, K. Umar, M. S. Md Yusuff, M. T. Anees, A. Qadir,and Y. K. A. Almanasir, ‘‘Modification of the water quality index (WQI) process for simple calculation using themulticriteria decision-making (MCDM) method: A review,’’Water, vol. 13, no. 7, p. 905, Mar. 2021. [7] World HealthOrganization. (1993). Guidelines for Drinking-Water Quality.World Health Organization. Accessed: Jan. 12, 2022. [Online].Available:http://apps.who.int/iris/bitstream/handle/ 10665/44584/9789241548151-eng.pdf [8] Standard Methodsfor the Examination of Water and Wastewater, Federation WE, APH Association, American Public Health Association(APHA), Washington, DC, USA, 2005. [9] L. S. Clesceri, A. E. Greenberg, and A. D. Eaton, ‘‘Standard methods for theexamination of water and wastewater,’’ Amer. Public HealthAssoc. (APHA), Washington, DC, USA. Tech. Rep.21, 2005.

[10] M. F. Howladar, M. A. Al Numanbakth, and M. O. Faruque, ‘‘An application of water quality index (WQI) and multivariate statistics to evaluate the water quality around Maddhapara granite mining industrial area, Dinajpur, Bangladesh,’’ Environ. Syst. Res., vol. 6, no. 1, pp. 1–8, Jan. 2018.