Water Quality Monitoring and Prediction of Water Quality at College Premises using Internet of Things

Jaihind G, Ezhilarasie R, Umamakeswari A

Abstract: IoT is becoming more popular and effective tool for any real time application. It has been involved for various water quality monitoring system to maintain the water hygiene level. The main objective is to build a system that regularly monitors the water quality and manages the sustainability. This system deals with specific standards like low cost background and system efficiency when compared to other studies. In this paper, IoT based real time monitoring of water quality system is implemented along with Machine learning techniques such as J48, Multilayer Perceptron (MLP), and Random Forest. These machine learning techniques are compared based on the hyper-parameters and the results were obtained. The attributes such as pH, Dissolved Oxygen (DO), turbidity, conductivity obtained from the corresponding sensors are used to create a prediction model which classifies the quality of water. Measurement of water quality and reporting system is implemented by using Arduino controller, GSM/GPRS module for gathering data in real time. The collected data are then analyzed using WEKA interface which is a visualization tool used for the analysis of data and prediction modeling.The Random forest technique outperforms J48 and Multilayer perceptron by giving 98.89% of correctly classified instances.

Index Terms: Internet of Things (IoT), Machine learning techniques, Random Forest, water quality, WEKA.

I. INTRODUCTION

Water is the first necessary thing for any living organisms in the world. More than 70% of our world is covered by water. Besides having too much of water, there still remains a water scarcity for both domestic and agriculture purpose. Earth contains only 2.6% of fresh water. Statistics says that death is increasing day by day by consuming untreated water. There is 200 million and above cases were reported about illness and spreading of diseases because of low quality water and uncleanliness surroundings. Drastic changes in water temperature and weather transforms the aquatic water quality. In India 90% of the rural people for their drinking water purpose depends on lakes, ponds or water supplied by government. So it is necessary to analyze the water quality before it reaches the consumers [1].

Conventional method for analyzing water quality needs to take several number of water samples from various location by human. Then the water samples should be taken to laboratory for testing water quality. This conventional method needs powerful equipment and experienced person to test the water samples collected at various sites. There are many obstacles and demerits in conventional method as it is a time taken process. Manually collecting water samples from various locations is also tedious process and total amount of cost invested on the system is too high. Another major demerit is that the real time water condition will be changed, as it takes longer time to collect water sample and to check in laboratory. Thus the system is unreliable and requires more human intervention [2].

By analyzing the demerits from conventional method and evaluating with various standards an IoT based analysis of water quality monitoring and management system is proposed with low cost and reliability. IoT is emerging as the most reliable platform for real time application. Deploying various Wireless sensor network (WSN) based sensors in the location where the water quality should be tested and reliable data can be gathered by monitoring remotely. Water quality can be determined by chemical, biological and physical characteristics. After calculating various parameters then it to be compared with set of standards. This WSN based real time monitoring system will intimate an earlier alert to ensure quickly response to water contamination. WSN requires low startup and low maintenance. Thus the implementation of this system will surely overcome the conventional method by reducing human intervention using low cost equipment with high accuracy in data and it is reliable. This system will increase the performance and by using IoT, data can be put in to cloud it will reach the end user and the user can prevent from consuming dirty or untreated water [3]. Prediction model is used to analyze the data collected by the system which evaluate the uncertainty with standardized conditions. It is capable of finding values based on training and testing processes. The basic multilayer perceptron consist of an input, hidden and an output layer [4]. In this paper three machine learning models namely J48, Multilayer perceptron (MLP), Random Forest are compared and used to determine the type and quality of water.

II. RELATED WORK

Basically there are various parameters are to be considered to analyze the quality of water such as pH, turbidity, conductivity, amount of dissolved oxygen, salinity in water. By calculating physical and chemical parameters water quality can be tested and the drinking water quality can be assured with hygienic level.Likewise building water quality monitoring system IoT described in [5] which provides frequent detailed reports of gathered data from locations.

Manuscript published on 30 July 2019.
***** Correspondence Author (s)
Jaihind.G, Embedded Systems, SASTRA University, Thanjavur, India.
Ezhilarasie R, CSE Department, SASTRA University, Thanjavur, India.
Umamakeswari A, CSE Department, SASTRA University, Thanjavur, India.

By using plug and play method the device are controlled and data can be transferred with host system. In [6], system were designed in Raspberry pi 3 model B which runs on Linux kernel, DS18B20 digital temperature sensor used to gather the temperature of water which provides 9 to 12 bits temperature readings. In [7], the water meter collects the data about consumption of water at different locations. GPS/sensor is implemented to gather data and to transfer the data to server. The data is gathered by RFID, sensor. For data designing Expanded Theory of planned behavior (ETPB) is used. In [8], two-step approach 1.Dynamics of selected aquatic species, 2. Sensors particularly reports unsettling reason, system will assess the dynamics of fish using 3D array ultrasound transducers. In [9], Controller has been developed using Arduino Uno, which manages four sensors (temperature, pH, turbidity, conductivity) then controller connected to data concentrator. Analyzed water quality transferred to cloud through Wi-Fi transceiver. In [10], neural network modeling consists of collection of data, pre-processing of data and pruning of data. The training involves preference of architecture, functions, algorithms and hyper-parameters used, testing the trained model.

III. EXPERIMENTAL SETUP

Fig 1 gives the explanation of hardware setup of the system. To detect dissolved oxygen present in water Gas sensor MQ-2' is deployed, it will sense the data from environment and gather the information, transfer the values to Arduino controller. Similarly, the sensors of conductivity, Turbidity and pH values are obtained and interfaced with Arduino controller. After gathering the required data from various sensors Arduino controller transfers the data to IoT module (GSM/GPRS) to put the gathered data in cloud and the results are monitored in local host in real time. IoT module receive pin is interfaced with the Arduino controller transmission pin and IoT module transmission pin is interfaced with Arduino controller receive pin for the process of information transmission.

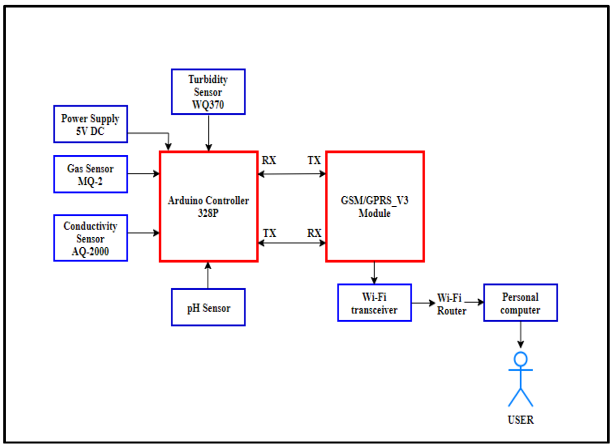

Fig 1: Architecture of IoT based water quality monitoring system

A. Turbidity Sensor- WQ370:

The central light beam is directed into the water monitoring sample. The light beam passes through the water sample and reflects the particles in it. The photo detector of the turbidity sensor is positioned at 90 degree to the light beam which is used to measure the light intensity.

B. Gas sensor MQ-2:

The MQ-2 sensor is used to detect gases in the surrounding of the water sample. Propane, CO, Methane are some of the gases can be detected using this sensor. The sensors are kept ON for preheating before it actually works.TTL driven analog pins are used to weigh the detected gas in ppm.

*C. Conductivity sensor AQ-2000***:**

Conductivity sensor made of a probe and it is deployed in the water samples to be monitored. It forms voltage among the electrodes used to measure the resistance of water, then it is to be transferred as conductivity.

D. PH sensor:

Three-way pH sensor used to calculate the level of pH, it can be used as soil moisture and light intensity testing meter also. Then all the sensors were connected to four analog pins of Arduino controller. Gas sensor is connected to pin A0 of the controller, pH sensor is connected to pin A1 of the controller, Turbidity sensor is connected to pin A2 and Conductivity sensor is connected to pin A3. Then the controller is interfaced with GSM/GPRS module and the experimental setup shown in Fig 2 and Fig 3 shows the quality of water monitored from the university premises.

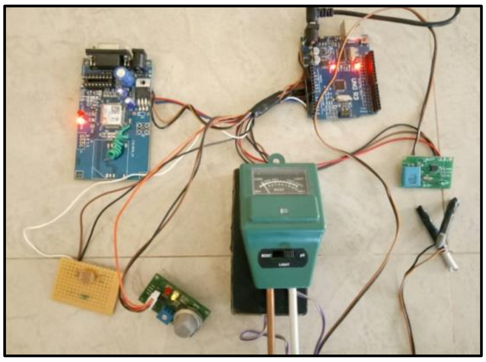

Fig 2: Sample setup of water quality Monitoring

Fig 3: Real time water quality Monitoring from the sump

IV. METHODOLOGY

A. Collection of data:

In this work four main parameters were calculated for analyzing the water quality such as Dissolved oxygen (DO), pH, Turbidity, Conductivity. These sensors were interfaced with Arduino controller to monitor the corresponding data every 15s.

Then the Arduino controller connected with GSM/GPRS module to send the data to the web interface. The results can be seen in the local host in the browser.

The sensors are deployed in various environments to analyze the water quality. Then the logs of the data are converted into a comma separated values via tera-term. The water quality analyzed and measured in and around the surroundings of SASTRA University. Both contaminated and non-contaminated water were measured.

B. Pre-processing of data:

Pre-processing data helps to achieve efficiency and quality of the obtained data. These collected data may contain erroneous or noisy data which reduces the accuracy of the system. Therefore the erroneous and inconsistent data should be eliminated. In this paper, WEKA tool is used for the modelling of data. It is used for the transformation and preparation of data. The machine learning model used 66% of data as training set and the remaining data is used for testing.

C. Modelling of data:

The machine learning model uses training, validation, testing and application of the data. In this paper, three models namely: J48, Multilayer Perceptron (MLP), Random forest were used. The collected data are then fed into these machine learning models. The hyper-parameters are tuned and altered to get the desired result. The hyper-parameters are learning rate, hybrid models, Number of epochs, GUI interface. For multilayer perceptron hidden layers are defined. J48 often applies to unsupervised whereas both MLP and random forest are supervised learning algorithms.

D. J48 Decision tree:

Decision tree helps to understand about the attributes used over the number of instances. The amount of correctly classified instances helps to understand the accuracy and success rate of prediction model. Accordance to the training instances, classification of newly resulted instances are established. This set of rules generates the regulations for the prediction of the goal variable. J48 is expansion to Iterative Dichotomiser (ID3) and their features being rule based derivations, range of data, attributes, pruning of data and lacking values. WEKA tool uses java for the execution of ID3. The WEKA tool offers some of options related to tree pruning. In different algorithms the classification is achieved recursively until each non-coherent leaf is pure, that is the category of the statistics ought to be as best as feasible. This algorithm it generates the regulations from which unique identity of that records is generated. The objective is regularly generalization of a choice tree until it profits equilibrium of flexibility and accuracy.

D. Multilayer Perceptron (MLP):

To classify the number of correctly classified instances MLP uses back-propagation method. Sigmoid networks are used in MLP except the samples of data are numeric in nature. During training time, the hidden layers are modified and the results are obtained. Back-propagation algorithm plays a vital role in getting the desired output from an inter-network of processing the components. It gradually learns iteratively from averaging the weights of samples. It consists of input, hidden and output layer. MLP with feed forward system is shown in Fig 4.

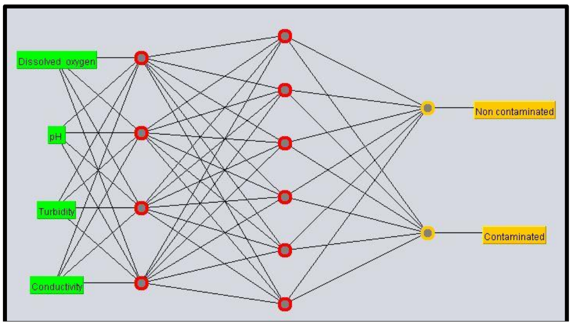

Fig 4: Neural network of MLP with two hidden layers

E. Random Forest

Random Forest (RF) constructs many individual decision trees at training. Predictions from all trees are pooled to make the final prediction; the mode of the classes for classification or the mean prediction for regression. As they use a collection of results to make a final decision, they are referred to as Ensemble techniques. Random Forest algorithm is said to be a black-box model in which the processing technique is unknown and rules for prediction is difficult to understand as they are having higher number of tress.

F. Feature Importance:

Feature importance is calculated as the decrease in node impurity weighted by the probability of reaching that node. The node probability can be calculated by the number of samples that reach the node, divided by the total number of samples. The higher the value is the more important the feature.

Bootstrap aggregating concepts are used for the training algorithm of random forest and it is also called as bagging. Let us consider a training set of X= $x_{1,\ldots,}x_n$, with outputs Y=$y_1,\ldots,y_n$, bootstrapping recurring A times chooses the sample with an arbitrary test with replacement. For a=1,…A:

1. Test sets with replacement, *n* no.of training sets from X, Y; call these X_a, Y_a.
2. Classification tree is trained f_a on X_a, Y_a.

Predictions for the missing test sets x' can be calculated by predicting the mean from all of the independent regression models on x':

$$\hat{f} = \frac{1}{A}\sum_{a=1}^{A} f_a(x')$$

otherwiseby getting the highest vote when in the instance of classification trees.

$$\sigma = \sqrt{\frac{\sum_{a=1}^{A}(f_a(x') - \hat{f})^2}{A-1}}$$

Bootstrapping model is the best fit because it reduces the variance of the given procedure. On a single training set, training all the trees gives best correlated trees. Bootstrap algorithm is used to decorrelate by training with different sets of data.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

V. RESULTS AND DISCUSSION

The samples are collected from different water bodies and they are observed using the experimental setup for a period of time. The data collected from the sensors can be viewed from the given local host http://iotcloudv3.000webhostapp.com/tank.php? which is shown in Fig 5. These data are then converted into a dataset with four different attributes collected from various samples. Then these data are then used for modeling the machine learning approach.

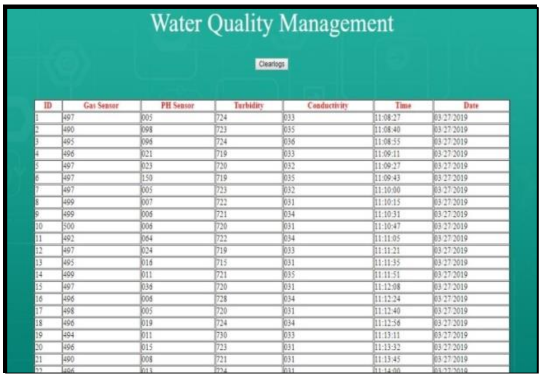

Fig 5: Collected data-set monitored in local host server

In this study, WEKA tool is used to measure various hyper-parameters which give the best prediction model among the three algorithms based on TP rate, FP rate, precision, Recall, F-measure, Receiver Operating Characteristics area, Precision Recall area, MCC. These hyper-parameters are compared with each other for all the algorithms which shown in table1 where Random forest outperforms the remaining models.

S.NO	Data Analysis	J48	MLP	Random Forest
1	TP rate	0.979	0.972	0.989
2	FP rate	0.052	0.101	0.002
3	Precision	0.980	0.972	0.990
4	Recall	0.979	0.972	0.989
5	F-Measure	0.980	0.972	0.989
6	MCC	0.910	0.886	0.957
7	ROC area	0.966	0.996	0.991
8	PRC area	0.978	0.996	0.987

Table 1: Data analysis compared with J48, MLP and Random Forest

It gives the higher TP rate of **0.989** and lowest FP rate of **0.002** and overall correctly classified instances of about **98.89%**. The J48 algorithm identified **96.68%**of correctly classified instancesand MLP algorithm identified **97.23%**. The root mean squared error is squared before averaging and larger the error value more it is influencive in nature. The RMSE value for random forest, j48, MLP algorithm comes to be **0.1127, 0.1758** and0.**1492** respectively. The best fit can be addressed when the RMSE value is closer to zero.

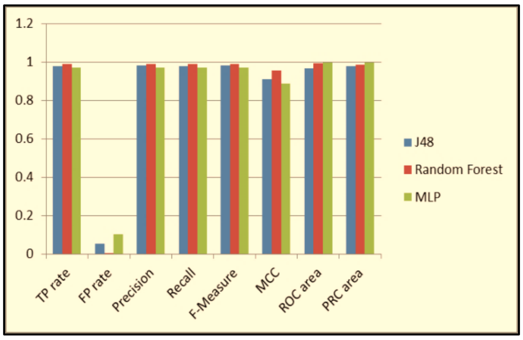

Fig 6: Graphical view of the data analysis compared with J4, MLP and Random Forest

VI. CONCLUSION

A real-time water quality monitoring system is developed using IoT. Data collection achieved by network sub-system established on a joined device of embedded prototype. Execution time of this system is less. This system forms the uncomplicated circuit and it is very easy to implement. Performance of this system is very effective when comparing to other models. Wireless sensor networks delivers effective and reliable infrastructure to domestic water quality monitoring and surveillance. But this system has limitations in resource, power and communication. Power consumption can be reduced by deploying low power gathering sensor nodes which gives low computational cost and high in efficiency with less executional time. In future this work can be implemented with additional parameters like temperature, oxidation reduction potential (ORP), salinity to get more effective water quality analysis. This work can be developed to interface with relay to control the water monitoring system and deploying low power gathering nodes to reduce the power consumption which increases the lifetime of the system. The Random Forest algorithm is very effective as it is a supervised learning algorithm. It gives the better accuracy model when comparing to others and it has the better learning capability.

ACKNOWLEDGMENT

The authors are grateful to the Department of Science & Technology, New Delhi, India (SR/FST/ETI-371/2014). They also wish to acknowledge SASTRA University, Thanjavur, India for extending the infrastructural support to accomplish this work.

REFERENCES

1. Yiheng Chen. etal.,, “Water quality monitoring in smart city: A pilot project”, Automation in Construction, Vol.no: 89 (307–316), 2018.
2. Mompoloki Pule. etal.,, “Wireless sensor networks: A survey on monitoring water quality”, Journal of Applied Research and Technology, Vol.no:15 (562–570), 2017.
3. Dung Nguyen. etal.,, “A Reliable and Efficient Wireless Sensor Network System for Water Quality Monitoring”,13th International Conference on Intelligent Environments, 2017.
4. ArchanaSolanki. etal.,, “Predictive Analysis of Water Quality Parameters using Deep Learning”, International Journal of Computer Applications (0975 – 8887) Volume 125 – No.9, 2015.
5. Md. Omar Faruq. etal.,, “Design and Implementation of Cost Effective Water Quality Evaluation System”, IEEE Region 10 Humanitarian Technology Conference (R10-HTC) 21 - 23 Dec 2017.

6. Paul B. Bokingkito Jr. etal.,, "Design and Implementation of Real-Time Mobile-based Water Temperature Monitoring System", Procedia Computer Science, Vol.no: 124, (698–705), 2017.
7. Yang Fu. etal.,, "Behavioural informatics for improving water hygiene practice based on IoT environment, Journal of Biomedical Informatics", Vol.no: 78,(156–166), 2018.
8. Cristina Polonschiia, Eugen Gheorghiua, "A multitiered approach for monitoring water quality", Energy Procedia,Vol.no: 112, (510–518), 2017.
9. M. Parameswari, M. Balasingh Moses, "Online measurement of water quality and reporting system using prominent rule controller based on aquacare-IOT", Des Autom Embed Syst, DOI 10.1007/s10617-017-9187-7.
10. Vesna Rankovic. Etal.,,, "Neural network modeling of dissolved oxygen in the Gruža reservoir, Serbia", Ecological Modelling, Volume No: 221, (Pages 1239-1244) April 2010.
11. G. Indrajith and K.Vijayakumar, "Automatic Mathematical and Chronological Prediction in Smartphone Keyboard" International Journal of Engineering and Computer Science ISSN: 2319-7242Volume 5 Issue 5 May 2016, Page No. 16714-16718.
12. K. Vijayakumar and C. Arun, "A Survey on Assessment of Risks in Cloud Migration", International Journal of Applied Engineering Research, ISSN 0973-4562 Vol. 10 No.66 May 2015.
13. K. Vijayakumar and C. Arun, "Continuous Security Assessment of Applications in Cloud Environment", International Journal of Control Theory and Applications, ISSN: 0974-5645 volume No. 9(36), Sep 2016, Page No. 533-541.

AUTHORS PROFILE

Jaihind G is currently pursuing Masters of Technology in Embedded Systems at SASTRA university, Thanjavur. He has undergraduate degree in Electrical and electronics engineering at Easwari engineering college, Chennai. His main research interests are in the areas of Internet of things, Machine learning, Smart Embedded Services.

Ezhilarasie R is working as an Assistant professor II (AP- II), Computer science department at SASTRA University, Thanjavur. She is pursuing phD at SASTRA university. She has a Masters degree in Computer science engineering at SASTRA university and she has degree in Computer science engineering at periyarManiammai college, Thanjavur. Her main research interests are in the areas of virtualization in Embedded Systems, Edge Computing, Internet of Things, Wireless Sensor Networks, Compilers. Click the link for her publication details**https://orcid.org/0000-0001-8486-7290**

Dr. A. Umamakeswariis a DEAN of School of computing, SASTRA university, Thanjavur. . Her main research interests are in the areas of Microprocessors and Interfacing, Computer Architecture, Operating Systems, Software Engineering, Embedded systems, Parallel Architecture, Visual Computing, Internet-of-things. Click the link for her publication details https://orcid.org/0000-0001-9724-1115

Retrieval Number: E10140785S319/2019©BEIESP
DOI:10.35940/ijeat.E1014.0785S319
Journal Website: www.ijeat.org

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)