WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Data Visualization and Prediction of Water Quality Monitoring System using IoT and Machine Learning.

Prof. R.S. Lavhe[1], Chaitanya Dahake[2], Akash Dive[3], Anjali Gaikwad[4], Namrata Munde[5]

[1]*Assistant Professor, Department of Information Technology, ABMSP's Anantrao Pawar College of Engineering & Research, Pune*
[2]*UG Student, Department of Information Technology, ABMSP's Anantrao Pawar College of Engineering & Research, Pune*
[3]*UG Student Department of Information Technology, ABMSP's Anantrao Pawar College of Engineering & Research, Pune*
[4]*UG Student, Department of Information Technology, ABMSP's Anantrao Pawar College of Engineering & Research, Pune*
[5]*UG Student, Department of Information Technology, ABMSP's Anantrao Pawar College of Engineering & Research, Pune*

A B S T R A C T

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. 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. The system makes use of IoT and Machine Learning technology. To develop an edge device for sensing the water quality parameters to detect changes in the water quality with respect to base line parameter using a machine learning approach at the edge device itself to generate the signals when water quality parameters go beyond its threshold value and to classify different types of contamination and analyses them for identifying possible contamination types. A network architecture to collect data on water parameters in real-time an accurate prediction and visualization of water quality related parameters is considered as pivotal decisive tool in sustainable water resources management. Five different ensemble machine learning models including Quantile regression forest, Random Forest, radial support vector machine, Stochastic Gradient Boosting and Gradient Boosting Machines were developed to predict the periodically biochemical oxygen demand values of water. An average data of water parameters such as temperature, Turbidity, Humidity, pH, total dissolved solids. Using Machine Learning tools to automatically determine suitability of water samples for drinking and irrigation purposes. Recursive feature elimination was then combined with the three ML models to reveal which of the water parameters had the greatest influence on the classification accuracies of the model.

Keywords: Sensors, IOT Based System, Machine Learning Algorithms, Data Visualization and Prediction

INTRODUCTION

By the World Health Organization (WHO) records, water related challenges such as deprival of safe and clean water for domestic purposes, increasing urban pollution and the scarcity of water are growing at an alarming rate. Hence, the decade 2018-2028 is announced as International Decade for Action, "Water for Sustainable Development" by the United Nations General Assembly. The hike in the population rate is another factor for the scarcity of water. The quality of water is controlled by its certain constituent parameters, so when the effluents are unloaded into the water the concentration of these water parameters change which results in the decrease of water quality. Over the past few years, a lot of investigative efforts have been done in the scrutiny of water quality and a variety of water quality models have been originated. Conventional methods include the collection of data manually and its statistical evaluation.

The quality of water is determined by various levels of different parameters. We proposed a fuzzy model that inputs water parameters (such as conductivity, pH and hardness). Dataset is generated from several water samples gathered from the IoT model. Artificial neural network (ANN) algorithms have been largely used for the prognostication of water quality; one such example of an algorithm is the Back Propagation (BP) algorithm. An issue in this BP algorithm is the low accuracy percentage, so an improved artificial bee colony (IABC) algorithm has been presented. Comparing both the algorithms, there has been a surge of 25% more accuracy in the model. In this algorithm, the connection weight values between network layers and the threshold values of each layer are previously enhanced. Analysing the results of the three algorithms, SVM model has a better performance when compared to accuracy. We suggested a proposal to resolve classification, prediction challenge of non-linearity and inadequate data using the SVM. But, the practicality of SVM is affected due to the difficulty in choosing the suitable SVM parameters. This paper exhibits a fusion of SVM and particle swarm optimization to decide SVM free parameters for better accuracy of the model.

A comparison of five classification algorithms such as Naïve Bayes, K star, Bagging, J48 and Conjunctive rule has been done to find the important factors that assisted in classifying water quality. Out of the five models, the Lazy model using the K Star algorithm was found out to be the best algorithm with 86.67%. of accuracy.

LITERATURE SURVEY

Nikhil Kumar Koditala [1]

Nikhil K Koditala, Dr. Purnendu S Pandey entitiled "Water Quality Monitoring System using IoT and Machine Learning" (2018). This paper having real time system which monitors water quality through sensors such as pH, turbidity and temperature and updates those values in Cloud service. This system consists of sensors which measure the chemical composition of water. These sensor values are then passed to Node MCU micro controller which has inbuilt WiFi module, using which the data is passed over to Azure Event Hub.

Rizqi Putri Nourma Budiarti [2]

Rizqi Budiarti, Anang Tjahjono and Mochamad Hariadi entitled "Development of IoT for Automated Water Quality Monitoring System"(2019). By implementing the use of Internet of Things (IoT) technology supported by data retrieval method using sensors, embedded systems and the use of remote communication technology can help simplify the water quality parameters. Water quality sensor connected to the raspberry pi type B and the sensor is able to send data by using MQTT protocols, we called the active sensors. To be able to store incoming data into MQTT Broker into the database, required a connection into the database server.

Ahmad Zaini [3]

Ahmad Zaini, Diah P. Wulandari, Retno Wulandari entitled "Data Visualization on Shrimp Pond Monitoring System Based on Temperature, pH, and Dissolved Oxygen with IoT"(2020). The first step in the work of this research is data acquisition by reading sensors. The sensors that will be used are temperature, pH, and DO sensors to determine each value of the parameter in the pool. Next is storing data into the database. Data that has been read by the sensor is then stored in the database to simplify its processing. And the last one is the mobile application which will function to visualize the results of the previous data processing.

Lakshmi Nandakumar [4]

Lakshmi Nandakumar's" Real Time Water Contamination Monitor using Cloud IoT and Embedded Platforms" The system

was developed with the intention of implementation mainly in those households, colonies or offices where the main supply of water is from the corporation. Also, office buildings and residential flats in metros rely on water supply provided and transported by tankers. This IoT - embedded system would counter a big social problem and help large number of people to stay healthy. This system was incorporated with sensors like pH, conductivity and turbidity to sense the respective parameters. Firebase and microcontroller boards Arduino and storing and post processing of the acquired sensor data. The system gave accurate warnings on Contamination of water samples which were validated through standard lab tests. The user interface, cloud storage enables the system to be conveniently used by both technical expert and common man.

Hauwa Mohammed Mustafa [5]

Hauwa M. Mustafa, Aisha Mustapha, Gasim Hayder, Abdullahi Salisu entitled "Applications of IoT and Artificial Intelligence in Water Quality Monitoring and Prediction"(2021). This paper contains on recent studies of IoT and Artificial intelligence (AI) tools in water quality monitoring. In hydrology, IoT devices are used to collect real- time data from a given water sample or station. At the same time, AI deals with the evaluation, simulation and prediction of data for easy interpretation and future use. Also, the review focuses on artificial neural network (ANN), particularly recent applications, advantages and limitations of ANN in the field of hydrology.

METHODOLOGY

Water quality monitoring is a critical process for ensuring the safety of water resources for human consumption and aquatic life. The following steps outline a general methodology for conducting water quality monitoring:

1. Define the objective: The first step in water quality monitoring is to determine the purpose and objectives of the monitoring program. This includes identifying the parameters to be monitored, the frequency of monitoring, and the type of data needed.

2. Select monitoring sites: The next step is to choose the sites where water quality will be monitored. These sites should be representative of the water resources being monitored and must take into account factors such as access, safety, and logistics.

3. Sample collection: Water samples are collected from the identified sites using appropriate sampling techniques. The sampling procedures should be standardized to ensure consistency in the results.

4. Laboratory analysis: The collected water samples are then analyzed in the laboratory to determine the levels of different parameters such as pH, temperature, dissolved oxygen, nutrients, and pathogens.

5. Data analysis and interpretation: The results obtained from laboratory analysis are analyzed and interpreted to determine the water quality status. Statistical analysis techniques can also be used to identify trends and patterns in the water quality data.

6. Reporting: The final step in water quality monitoring is to report the results to relevant stakeholders such as water resource managers, policymakers, and the public. Reports should be clear, concise, and include recommendations for improvements if necessary. Overall, water quality monitoring is an ongoing and dynamic process that requires regular evaluation and improvement to ensure the continued safety and sustainability of water resources.

EXISTING SYSTEM

The existing water quality monitoring system typically involves the following components and processes:

1. Water sampling: Water samples are collected from various sources such as rivers, lakes, streams, groundwater wells, and treated water systems.
2. Laboratory analysis: The collected water samples are analyzed in a laboratory to measure various parameters such as pH, turbidity, total dissolved solids (TDS), conductivity, total coliform, fecal coliform, and other parameters that indicate the quality of water.

3. Data interpretation: The results obtained from the laboratory analysis are interpreted by experts to determine the health and environmental risks associated with the water quality.

4. Reporting: The interpretation of data is shared with policymakers and the public through an online database or reports that are published periodically.
5. Response: If any potential risk is detected, appropriate measures are taken to address the problem and prevent further degradation of water quality.
6. Regulatory compliance: The regulatory agencies such as the Environmental Protection Agency (EPA) monitor the data generated by water quality monitoring programs to enforce laws and regulations related to the protection of public health and the environment.

PROPOSED SYSTEM

- Semi-automated Segregator which is less expensive, easy to use solution for a segregation system at households so that it can be sent directly for processing.
- Use of water parameter sensors for water quality and sends the sensing value to cloud storage and process the data to suggest to decreases the water pollution methods.
- Apply machine learning techniques and data analysis algorithms to data visualize and prediction of water parameters data.

Water Source Layer:

This layer contains the natural sources of a water source such as a river, a lake, a stream, or any other water source which is used to take samples for our prescribed system. The water source may also be a man-made water supply, such as a well, a water distribution network, or any other such as water tanks.

IOT / Sensing Layer:

As depicted in the figure, the sensing layer interacts directly with the water samples in a river, stream, dam etc. to measure water parameters. It is built into a vertical pole tagged “sensor probe” and consists of numerous sensors bundled together. These sensors might include pH, conductivity, turbidity, temperature or residual chlorine etc., similar to those offered by Labellum. All telemetry data measured by these sensors are sent to the Fog Nodes (FNs), wired or wirelessly, via the sending unit. In scenarios were installing sensors in water source(s) is extremely difficult or when the required sensors are not readily available, water parameter readings can be collected from the associated water treatment plants.

This layer consists of low-end processing devices (edge modules), such as single board computers (e.g., Raspberry Pi or Nvidia Jetson), or microcontrollers (e.g., Arduino, ESP32). These devices act as i.).

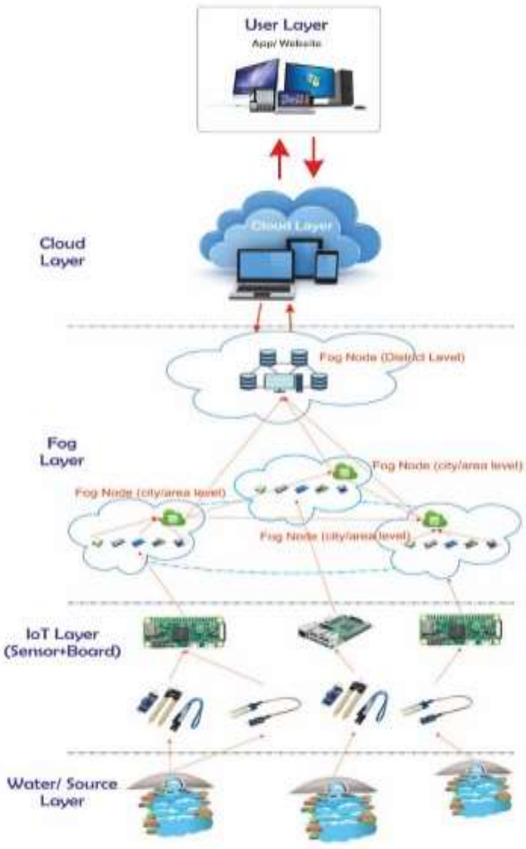

Fig. Architecture of water quality monitoring system using IoT and Machine Learning

Fog / Cloud Layer:

Fog Nodes (FNs): these are small sized distributed cloud computing nodes that bring computing and storage closer to the data source, thus reducing latency resulting from transmission delay to/from the remote Cloud. The FN is responsible for classification of water samples using machine learning models such as the ones proposed in this work. Due to the limited computing power at the Fog (compared to the Cloud), only the most influential parameters need to be considered when classifying water samples. This can be beneficial as less sensors would be required (since not all parameters are being measured) and by extension lower computing resources would be needed for the classification process. Furthermore, resource management, scheduling etc. can also be carried out on FNs. When long term storage and/or advanced computations are required, which are beyond the Fog's capacity, data are forwarded to the Cloud data centre. Cloud Data Centre: The Cloud is a remote high performance computing infrastructure, which provides computing on demand. In our system, the Cloud serves as a data warehouse as well as a platform for performing advanced data analytics, dashboarding, and hosting for relevant services and software.

Application Layer:

serves as an interface between users (water management authorities, end users / customers, other stakeholders) and software / services running in the Cloud. Relevant software for water parameter monitoring is hosted at this layer and made available to users through mobile and web platforms. The water monitoring network proposed in this work is to be deployed in the City of Cape Town in Western Cape, South Africa, with the intention of monitoring water parameters in water storage dams and/or water treatment plants across the city. Data gathered by the monitoring network are then passed through Machine Learning (ML) models to determine their suitability for consumption or irrigation purposes.

CONCLUSION

The paper shows a detailed survey on the tools and techniques employed in extant smart water quality monitoring systems. Also, a low cost and less complex water quality monitoring system is suggested. The implementation permit sensor to supply online data to consumers. The experimental setup can be grow the incorporating algorithms for anomaly detections in water quality.

REFERENCES

Koditala, Nikhil Kumar, and Purnendu Shekar Pandey. "Water quality monitoring system using IoT and machine learning." 2018 International Conference on Research in Intelligent and Computing in Engineering (RICE). IEEE, 2018.

Jalal, Dziri, and Tahar Ezzedine. "Performance analysis of machine learning algorithms for water quality monitoring system." 2019 International Conference on Internet of Things, Embedded Systems and Communications (IINTEC). IEEE, 2019.

Radhakrishnan, Neha, and Anju S. Pillai. "Comparison of water quality classification models using machine learning." 2020 5th International Conference on Communication and Electronics Systems (ICCES). IEEE, 2020.

Zaini, Ahmad, Diah Puspito Wulandari, and Retno Wulandari. "Data Visualization on Shrimp Pond Monitoring System Based on Temperature, pH, and DO (Dissolved Oxygen) with IoT." 2020 International Conference on Computer Engineering, Network, and Intelligent Multimedia (CENIM). IEEE, 2020.

Tewari, Naveen, and Mukesh Joshi. "Water quality prediction system (WQPS) and method using fog of thing (FoT)." 2020 9th International Conference System Modeling and Advancement in Research Trends (SMART). IEEE, 2020.

Kumar, Ganjikunta Raj, et al. "Waste contamination in Water—A Real-time Water Quality Monitoring System using IoT." 2021 International Conference on Computer Communication and Informatics (ICCCI). IEEE, 2021.

Mustafa, Hauwa Mohammed, et al. "Applications of IoT and Artificial Intelligence in Water Quality Monitoring and Prediction: A Review." 2021 6th International Conference on Inventive Computation Technologies (ICICT). IEEE, 2021.

Haq, KP Rasheed Abdul, and V. P. Harigovindan. "Water Quality Prediction for Smart Aquaculture using Hybrid Deep Learning Models." IEEE Access (2022).