INTERNATIONAL JOURNAL OF RESEARCH CULTURE SOCIETY ISSN: 2456-6683 Volume - 4, Issue - 3, Mar – 2020
Monthly, Peer-Reviewed, Refereed, Indexed Journal Scientific Journal Impact Factor: 5.245
Received on : 11/03/2020 Accepted on : 24/03/2020 Publication Date: 31/03/2020

An Automated System For Effective Monitoring of Water Quality

[1]Dr.G.Fathima, [2]G.Sandhya, [3]G.Keerthana, [4]S.Sandhiya

[1]Professor, [2, 3, 4] Student,
[1, 2, 3, 4] Computer Science and Engineering, Adhiyamaan College of Engineering(Autonomous),
[1, 2, 3, 4] Affiliated to Anna University, Hosur, India
Email – [1]fathima.ace@gmail.com, [2]sandhya.g.24.12@gmail.com, [3]gkeerthana258@gmail.com,
[4]sandhiyas2406@gmail.com

Abstract: *Water is the essential part of all living creatures in the world. Due to depletion of water resources many living creatures have a possible chance to drink impure water. The impure water would have larger parametric values and therefore the risk would also be very high. Many animals without knowing the hazardous impact of impure water would drink the water and would be subjected to various anomalies. The quality of water which we drink also plays a very major role in maintaining our health. Hence, these factors makes it essential to monitor the quality of water. Using emerging technologies such as IoT, Machine Learning and cloud technologies we can automate the water quality monitoring process. The proposed model collects data from sensors, stores the value in cloud and with the help of the developed machine learning model the quality of water is predicted. The proposed machine learning model implements Decision tree algorithm and has an accuracy score of 0.972. Finally the predicted outcomes are being displayed in the web page with visualization of various parametric values over a period of time.*

Key Words: *Water, Quality Monitoring, IoT, Cloud, Machine Learning, Decision Tree Model*

1. INTRODUCTION:

Water is an integral part of our life. Without water no individual can survive in this world. The water which we intake should be of good form in order to lead an healthy life. Due to fast growth of population, there are many cautions for water depletion. The pure form of water would be available only in the reservoirs, rivers and dams. But, many of the industrial wastes are being mixed in the water making it polluted. Without knowing that the water is polluted, many people and other living beings would consume the water. There are various hazards of drinking impure water. In this fast moving world, many of us fail to check the quality of water which we drink and as a result of this we often feel sick. Many hazardous diseases such as Cholera are spread by drinking the contaminated water. There is a need to monitor the quality of water which we intake. Traditional Water Quality monitoring is not automated and it involves testing the water quality parameters manually and finding the quality of water. This traditional water quality monitoring takes much time and it is difficult to take early action to avoid water contamination. Hence it is essential to automate the water quality monitoring system using technologies such as IoT, Cloud and Machine Learning . By using the proposed system the alert message can be sent to the authorities in correct time so that actions can be taken immediately whenever there is a contamination in water. The proposed system first uses sensors such as Temperature, pH, Turbidity, Conductivity/TDS sensors to measure the amount of various parameters present in water . These sensors are in turn connected to the Arduino Mega – 2560 board which has an in built AtMega Microprocessor. The Wi-Fi Module and the GSM Module are in turn connected to the Arduino Mega board. The data collected by the Microcontroller is passed to the firebase cloud . The data from the firebase cloud is collected and is processed by the Machine Learning model to check the quality of water by developing a web page. The analyzed result is displayed along with the graphical information regarding the measurements variations over the period of time. Whenever the quality of water is not good , the authorities are intimated about the quality of water through Short Messaging Service(SMS).

1.2. LITERATURE REVIEW:

The application of Machine Learning models in accordance with IoT and cloud computing is introduced in paper[1]. The system collects data from sensors, stores the data in cloud and performs analysis and displays the results in the web portal. The Machine Learning model helps in the automated turning on/off of heater and cooler. This paves way in maintaining the temperature of the water.Another system utilizes Raspberry-Pi module and sensors such as PH and Turbidity sensors to measure the values of the parameters present in the water. These values were passed to the Raspberry-Pi module, where a Machine Learning Algorithm called K-Means is applied to cluster the data into groups based on their mean distances. This work is proposed in paper [2].The prediction of the water quality by utilizing gradient boosting and polynomial regression algorithms are applied and finding the algorithm that best fit in predicting the quality of water was performed. The gradient boost and polynomial regression with degree 2 were found to best suit the water quality prediction with high accuracy rate and mean absolute error (MAE) of 1.9642

and 2.7273 respectively. These ideas were proposed in paper [3].The data collection from various sensors such as pH, DO, Conductivity and storing it in the database and finally analysing it with the neural networks in order to predict the contamination of water. The system uses fixed standard rate for each parameter and compares it with the obtained value from the sensors by using neural network analysis and based on the result obtained, the user is intimated about changes in the parametric value was introduced in the paper[4]. The water quality monitoring data taken from water quality monitoring stations were used and a predicted model was developed with the utilization of Long short time Memory deep neural network in order to predict the quality of water was proposed in paper[5].

2. PROPOSED MODEL:

The proposed model is divided into three main blocks such as,

- Hardware Setup Module
- Cloud storage Module
- Analysis Module

- *Hardware Setup Module:*

The Hardware setup module involves integration of hardware components. The sensors such as Temperature, Turbidity, pH, Conductivity/TDS are first integrated to Arduino Mega-2560 Microcontroller board. The positive and negative terminals of the sensor are connected to the breadboard that is provided with 9V DC power supply. The data pin of the sensors are connected to the Arduino Mega board. The Wi-Fi module and GSM module are integrated to the Arduino Mega board[6]. The Wi-Fi Module involves the use of ESP8266 Wi-Fi Module that is used in the transmission of sensor data to the cloud. The GSM Module involves the use SIM-800A GSM Module to transfer the alert messages to the authorities about the quality of water that is supplied.

- *Cloud Storage Module:*

The cloud storage module stores the information collected from the sensors in the Firebase cloud[7]. The Firebase's Real Time Database is used for storing the real time sensor data. Firebase's Real Time database fetches the data with the help of Message Queuing Telemetry Transformation (MQTT) Protocol from the Arduino Mega with the help of ESP8266. The data collected are stored in the JSON format. The firebase database also helps in the creation of web application or mobile application according to our need.

- *Analysis Module:*

The Analysis Module involves analyzing data that is being stored in the cloud by the application of Machine Learning algorithms. An Web application is created through which the authorities can login and check the quality of water by the application of Machine Learning model that includes decision tree algorithm to predict the quality of water. The dashboard in the web page would help the authorities with the present and past visualizations of parametric measurements of water.

2.1. IoT ARCHITECTURE:

This section explains the architecture of the IoT platform that is being developed. The Fig.1 indicates the reference model of the IoT module. The IoT reference model from world forum consists of seven layers[6].

Layer 1: The layer1 of the IoT model is the hardware component sensors such as Temperature, Turbidity, pH and Conductivity/TDS Sensor.

Layer 2 : The data collected from the sensors are passed to Arduino Mega Microcontroller . The ESP8266 Wi-Fi Module sends the data to the cloud Firebase Real-time Database.

Layer 3: The data sent through the wi-fi module are to be formatted in JSON. Since the Firebase database accepts data in the form of JSON.

Layer 4: The JSON formatted data are stored in the firebase real time database with the help of MQTT Protocol. The database gets updated when ever a new value is sensed by the sensor and passed via the Wi-Fi Module.

Layer 5: The data from the firebase database are fetched to predict the quality of water by application of Algorithms that is suitable for prediction using sci-kit learn and pandas library of python.

Layer 6: The data are visualized with the help of powerful visualization tool. This visualization helps authorities understand the level of parameters during various time periods.

Layer 7: The results of the water quality prediction and visualization results are published in the web application. The respected authorities would be able to see the readings and quality of water parameters and water.

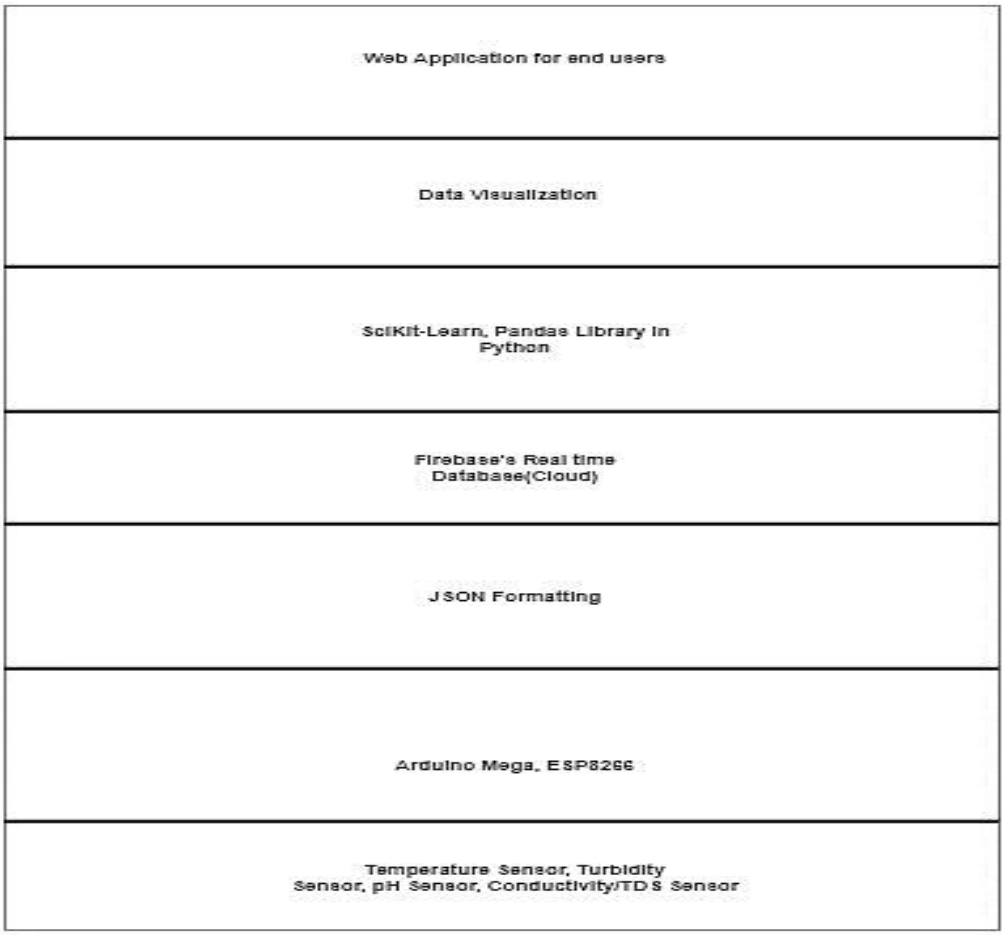

Figure 1. IoT Architecture

2.2. MACHINE LEARNING:

This section explains the Machine Learning part of the proposed system.

- *Data Set:*

The dataset was collected from data.gov.in where it includes parameters such as Temperature, pH, Turbidity, Conductivity, Faecal Coliform levels, Total Coliform Levels and Nitrate levels collected in the year 2014[8].

- *Data Cleaning and Data Manipulation*:

The data cleaning part involves cleaning of data using various strategies such as finding the null values, replacing the null values with mean values or other techniques such as replacing the value with percentile values[3]. The dataset collected from the data.gov.in were imported in the Jupyter notebook and data cleaning was done to replace the Null values with mean values of the particular column. And Matplotlib and seaborn were used for visualizing the flow of data in the dataset. The duplicate values were also checked. If duplicates were found that duplicate value is eliminated. Data Manipulation is performed by removing columns which are not significant for analysis. The label encoder is used to convert the classification labels to binary values.

- *Correlation Analysis:*

The correlation analysis involves analysing the data parameters in order to predict which of the parameters are closely related and which are unrelated. Heat Maps are used to visualize how well the parameters are correlated with each other[9]. The darker the color the more correlated the variables are. In this system correlation analysis is performed by using correlation matrix to check the parameters are correlated to each other. If the correlation value is negative then there is no correlation among those variables. If the correlation value is more than 0.5 then those variables are correlated with each other. Based on the correlation value, the most redundant features are to be removed. After removing redundant features, the heatmap should be symmetric[8]. Fig 2 represents the correlation values after the removal of redundant features. Thus correlation analysis helps in reducing the redundant features, which help in providing better performance of algorithm.

Figure 2. Heat Map representing the correlation among variables

- *Data Preparation for Model Building:*

Data Preparation for model building splits the data for training and testing. Almost 70% of the data are used for training the model and remaining 30% are used for testing. Finally, the training data are fitted to the model. After training the model is tested with testing data.

- *Hyper parameter tuning:*

Hyper parameter tuning refers to tuning of variables to make sure that the variables don't overfit or underfit. In this system, since the algorithm used in Decision tree , the Hyper parameter tuning helps in finding the max depth for implementing the algorithm. By using Hyper parameter tuning , the max depth is found to be 2 with best score of 0.956.

- *Algorithm:*

Algorithms play a major role in implementing the Machine Learning. The algorithm best suitable must be taken into considerations. We incorporate supervised learning in our system. K-Nearest Neighbor and Decision tree algorithms were implemented and were best checked for their accuracy rate and parameters that help in providing better performance. Since the water quality monitoring is a classification problem, K-Nearest Neighbor and Decision Tree algorithms best fit the analysis. The K-Nearest Neighbor algorithm finds the nearest neighbor's to a particular value by finding the Euclidian distances and the most less value of the Euclidian distance is considered as the nearest neighbor. The Decision Tree constructs a tree structure with leaf nodes containing the classification values. It is noted that **Decision Tree Algorithm** has better accuracy rate than K-Nearest Neighbor with 0.972 as accuracy rate. Table I shows algorithm performance metrics by testing with various performance testing parameters. The real time data that is fetched from the firebase's real time database are subjected to the calculation of Water Quality Index(WQI). The WQI is calculated as follows:

$$WQI = \sum_{i=1}^{n} q_i \; w_i \; / \sum_{i=1}^{n} w_i \qquad (1)$$

In the equation(1),

- n is the number of parameters.
- q_i = The quality of the i^{th} parametric value that is calculated as

$$q_i = 100 \; (v_{measured} - v_{ideal} / v_{standard} - v_{ideal})$$

$v_{measured}$ = Measured value of the parameter.

v_{ideal} = Ideal value of the parameter. The ideal of all parameters remains zero except pH and D.O. The ideal value of pH is 8 and ideal value of D.O is 14.6mg/l.

$v_{standard}$ = Standard value specified by World Health Organisation(WHO).

- w_i = The unit weight of i^{th} parameter is calculated as the inverse of standard value of the parameter[15].

Table I : Algorithm Performance Metrics

Algorithm	Precision	Recall	fl-score	Support	Accuracy
KNN	0.97	0.97	0.97	108	0.888
Decision Tree	0.97	0.97	0.97	108	0.972

The Standard values used in measuring WQI are according to the World Health Organisation(WHO) standards. These WQI value along with the various parametric value are passed to the Algorithm and the algorithm predicts the output class. There are five classes. If the predicted outcome is class I , then the water quality is excellent and would have the WQI range between 0-25. If class II is the outcome predicted, then it indicates that the water quality is good and the WQI range would be between 25-50. If the outcome predicted is class III, it indicates that the water quality is medium and the WQI range would be between 50-75. If the predicted outcome is class IV then it indicates that the water quality is bad and the WQI range would be between 75-100. If the predicted outcome is class V then it indicates that the water quality is very poor and the ranges of WQI will be greater than 100 [15].When ever the outcome is predicted as Class IV or Class V, the message would be sent to the authorities immediately, so that timely action could be taken.

Table I shows the various values of Precision, Recall, f1-score, support and accuracy. A correlation matrix plays an important role in predicting the values of algorithm's performance metrics.

Precision is calculated as True Positive value obtained that is divided by actual true positive and false positive. The precision value of KNN and Decision Tree obtained for this system is 0.97.

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive + False Positive}}$$

Recall parameter suggests that the ratio obtained for the correctly predicted positive observations to each and every observation in the original class. Recall value that is obtained for the Decision Tree and KNN algorithm used in this system is 0.97

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive + False Negative}}$$

f1 score is calculated to check the balance of precision and recall. The f1 score calculated for KNN and decision tree algorithms used in this system is 0.97.

f1 score = 2*((precision_value*recall_value) / (precision_value+ recall_value))

Accuracy is the ratio of number of correctly predicted outcome for given input samples. The Accuracy of KNN and decision tree are respectively 0.888 and 0.972.

$$\text{Accuracy} = \frac{\text{True Positive} + \text{True Negative}}{\text{True Positive + False Positive +True Negative + False negative}}$$

3. EXPERIMENTAL SETUP:

The system has to integrated with the Hardware and Software components in order to function effectively. This section involves providing details regarding experimental setup of the system.

- *Architectural Design*

Fig 3 represents the entire architectural diagram of this system. The various types of sensors are connected unidirectionally to the Arduino Mega Micro controller board.

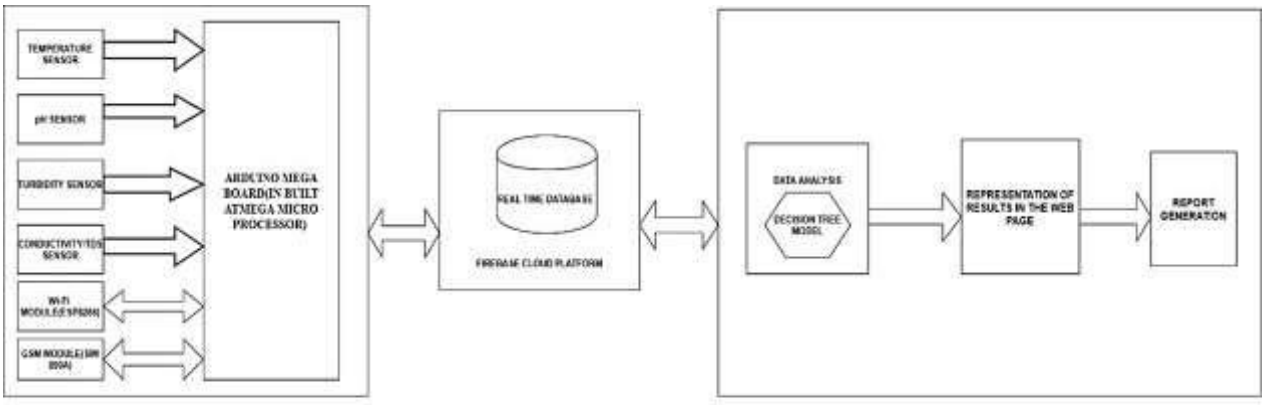

Fig 3: ARCHITECTURAL DESIGN

The Wi-Fi module and the GSM module are connected bi-directionally to the Micro controller board. The sensors are placed in the water to sense their respective parametric values. With the help of Wi-Fi Module these data are passed to the Firebase Cloud. These data are stored in the Firebase's Real Time Database. The data are fetched from the cloud where the Machine Learning Algorithms are applied in order to predict the quality of water and provide various visualization strategies for better understanding of water quality for various time periods.

- *Hardware and Software Components:*

The hardware and software components used in this project are listed below.

✓ Arduino Mega:

Arduino Mega is a micro-controller board that has 54 digital input/output pins and 16 analog input/output pins. Due to large amount of input/output pins we can connect many sensors to the micro controller board. In this system, there are 4 sensors, a Wi-Fi module and a GSM module connected to the micro controller board[10].

✓ Temperature Sensor:

The temperature sensor used in this system is DS18 B20 water proof temperature sensor. There is a need of temperature sensor in water quality monitoring since the temperature might affect the oxygen levels in the water. Hence temperature sensor is used. The temperature sensor used is digital[1].

✓ Turbidity Sensor:

The turbidity sensor used in this project is IGZD turbidity sensor. Turbidity Sensor is used to measure the suspended particles in water. The less the turbidity range the water quality would be high. The data pin of Turbidity Sensor is connected to the Analog pin A3 of the Arduino board[11].

✓ pH Sensor:

pH Sensor is used to measure the acidic content in water. It is one of the most important parameters in measuring the water quality. The normal pH range of pure water is 7. pH sensor used in this project is Analog. The pH sensor data pin is connected to the Analog pin A0 of the Arduino board[12].

✓ Conductivity/TDS Sensor:

The conductivity/TDS sensor used in this project is "DF ROBOT TDS sensor meter". The TDS is one of the important parameter since measures the salt content in water. The less the conductivity the more the quality of water. The conductivity sensor is connected to the Analog pin A2.

✓ Wi-Fi Module:

The Wi-Fi Module used in this project is "ESP8266" Node MCU . This Wi-Fi Module is used to transmit the sensor data to the cloud. The Wi-Fi is connected serially to the Arduino Mega board through Rx and Tx pins connected to both boards[13].

✓ GSM Module:

The GSM Module used in this project is "SIM800A".The The purpose GSM Module is to transmit the message about the water quality to the authorities whenever the quality is bad. The GSM module is connected to the Arduino Mega board[4].

✓ Firebase Cloud:

The data that are collected are passed through the Wi-Fi Module to the firebase cloud. The firebase stores the data in the Real Time Database. Whenever new value is collected it is stored in the database[14].

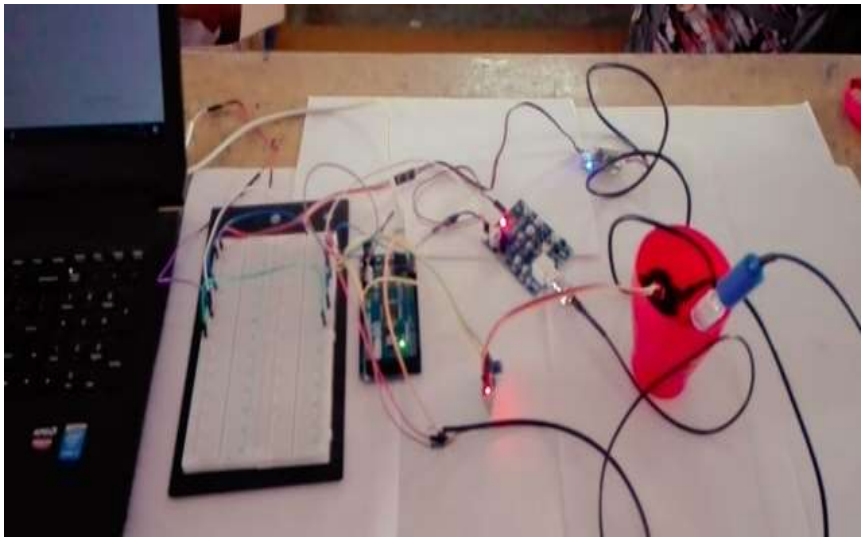

Figure 4. Hardware Setup

The Fig 4 represents the complete hardware setup which includes Breadboard, Arduino Mega, Turbidity Sensor, Temperature Sensor, pH Sensor, Conductivity Sensor , Wi-Fi Module and GSM Module.

4. CONCLUSION:

The water quality monitoring system was developed to automate the process of traditional water quality monitoring process. The proposed system is cost efficient and incorporates various technologies such as IoT, Cloud Computing and Machine Learning. The accuracy score obtained assures better quality monitoring of water. This system would be helpful in taking timely actions which would further help in preventing the spread of various anomalies that are associated with the impure water. The Report generation about the water quality would also provide better analysis to the authorities. This system thereby helps in automated monitoring of water quality with less or no human intervention.

REFERENCES:

1. Nikil Kumar Koditala, Dr.Purenendu Shekar Pandey(2018):"Water Quality Monitoring using IoT and Machine Learning" IEEE, DOI: 10.1109/RICE.2018.8509050
2. Soundarya Pappu, Pratuysha Vudatha, Niharika.A.V, Kathick T, Suresh Sankaranarayanan (2017):"Intelligent IoT Based Water Quality Monitoring" International Journal of Applied Engineering Research(IJAER), volume-12, issue-16, page no.5447-5454.
3. Umair Ahmed, Rafia Mumtaz, Hirra Anwar, Asad A.Shah, Rafia Mumtaz, Hirra Anwar, Asad A.shah , Rabia Irfan, Jose Garcia-Nieto (2019): "Efficient Water Quality Predication Using Supervised Machine Learning" MDPI.
4. C.Ashwini, Uday Pratap Singh, Ekta Pawar, Shristi (2019):"Water Quality Monitoring using Machine Learning and IoT" International Journal of Scientific & Technology Research(IJSTR), volume-8, issue-10, page no.1046-1048.
5. Ping Liu, Jiu Wang, Arun Kumar Sangaiah, Yang Xie and Xinxhun Yin (2019): "Analysis and Prediction of Water Quality Using LSTM Deep Neural Networks in IOT Environment" MDPI.

WEB REFERENCES:

- Internet of Things World Forum, www.iotwf.com/resources/72
- https://data.gov.in/dataset-group name/water-quality
- https://www.mdpi.com/1660-4601/15/12/2907/pdf
- https://medium.com/data-science-bootcamp/data-visualization-in-machine-learning-beyond-the-basics-baf2cbea8989
- https://store.arduino.cc/usa/mega-2560-r3
- https://www.electroschematics.com/diy-water-quality-meter-using-a-turbidity-sensor/
- https://www.dfrobot.com/product-1025.html
- https://www.electronicwings.com/arduino/esp8266-wifi-module-interfacing-with-arduino-uno
- https://firebase.google.com/
- http://www.ajer.org/papers/v5(10)/O0501096101.pdf