ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 11, Issue 8, August 2024

Predicting Water Quality Parameters Using Machine Learning

[1] Tadikamalla Bhavana, [2] Bheema Shyam Kumar, [3] Dr. M. Karthikeyan

[1] [2] [3] Department of Computing Technologies, SRM University, Chennai, Tamil Nadu, India
Corresponding Author Email: [1] tb2349@srmist.edu.in, [2] bk0317@srmist.edu.in, [3] karthikm@srmist.edu.in

Abstract— "Predicting Water Quality Parameters using Machine Learning" project is a Water is always an essential component of daily existence. Water conservation and management are essential to human survival because of the state of the environment worldwide. Humanitarian initiatives driven by consumers that may be swiftly produced using Internet of Things (IoT) technology have been more popular in recent times. We propose in this study an real-time water level using Iot system. Our prototype is predicated on the notion that water purity can be a critical factor. The desired parameter is detected by a water level sensor, and a real-time signal is fed in if the water PH level reaches the parameter. One of the cloud servers was set up as a data repository.

Index Terms— Water Quality, Physical Parameters, Turbidity, Chemical Parameters, Temperature, PH, Dissolved oxygen.

I. INTRODUCTION

Water is a precious and essential resource that sustains life on our planet. Our need for clean, safe water is growing along with the world's population. Ensuring the health of human populations and the environment requires regular monitoring and maintenance of water quality. Historically, manual sampling and costly, time-consuming laboratory analyses have been the mainstays of water quality evaluations. But there is a chance to completely change how we track and forecast water quality thanks to technological developments, especially around machine learning. In this project, we investigate the use of Machine learning methods to forecast parameters related to water quality. Our goal is to create models that can predict the useful water quality indicators like temperature, Ph, Turbidity, and dissolved oxygen by utilizing data analytics. This makes monitoring easier and more affordable, and it also makes it possible to react quickly to any threats to the quality of the water, Data collection, preprocessing, feature selection, and the application of various machine learning algorithms are all part of this journey. The ultimate objective is to develop models that advance our knowledge of the intricate relationships found in aquatic ecosystems while also accurately predicting key water quality parameters. Come along as we explore the world of predictive analytics to protect water, one of our most important resources. Our goal is to improve our capacity to oversee, maintain, and protect water quality for present and future generations by using machine learning.

A. Machine Learning - Overview

Among the many important reasons why machine learning is so popular is that it can automatically extract meaningful insights, identify previously unidentified patterns, and without explicit programming instructions, create incredibly accurate predicting models from data. This high-level knowledge is crucial if one is ever going to be engaged in making decisions about the use of machine learning, how it might help accomplish project and company objectives, which machine learning methods to utilize, possible hazards, and how to interpret the results.

B. Problem Statement

The task is to develop machine learning models that can accurately predict various water quality parameters, such as Temperature, conductivity, turbidity, pH, dissolved oxygen (DO), and different contaminants (e.g., organic compounds, heavy metals), given historical data. The goal is to create models that can assist with efficient water resource management and monitoring, the detection of possible contaminant events, and the assurance of water safety for use in industry, agriculture, and drinking. Every city authority faces a challenge in providing safe drinking water due to the rapidly increasing urbanization. Polluted water can occur at any time. Therefore, it's possible that the water we set aside in the water tank on our apartment's roof or in our building's basement is unsafe. Most people in India still use basic water purifiers, which is insufficient to ensure that the water is pure. A general-purpose water purifier may not be able to remove hazardous particles or chemicals from the water due to certain circumstances. Furthermore, it is not feasible to manually check the water's quality every time. Therefore, in order to keep an eye on the condition of the water reserved in our apartment or society's water tank, an automated real-time monitoring system is needed. If there is an issue with the reserved water, it will automatically alert us. Additionally, we can check the water's quality from anywhere at any time. We specifically designed this system for residential areas with this in mind. For ecosystems and communities to remain safe and healthy, water quality is essential. To detect possible causes of pollution, evaluate the effects on the environment, and make well-informed decisions about the management of water resources, it is imperative to monitor water quality indices, including pH, dissolved oxygen, turbidity, and nutrient levels.

C. Significance

The significance of Identifying water quality Parameter outcomes using ML cannot be overstated, as it offers several benefits:

Enhanced Aquaculture Management: This section explains how aqua culturists can transform aquaculture management by using (IoT) Internet of Things - based on monitoring Water Quality technologies to provide fast and accurate data on important environmental factors.

Increased Productivity and Profitability: It explains how improving fish health, growth rates, and overall productivity may be attained by IoT-enabled monitoring of water quality parameters, which in turn increases the financial sustainability of aquaculture operations.

Motivation and Significance: The introduction emphasizes how important it is to look after the water quality in aquaculture to make sure that the aquatic organisms are healthy and productive. It highlights the necessity of creative alternatives, like the Iot, to get beyond the drawbacks of monitoring techniques and boost the sustainability and effectiveness of aquaculture operations.

Warning System: By Predicting the Water Quality parameters, it automatically recognizes the sensor measurements and send report to the user phone using GSM in Arduino uno.

D. Machine Learning Tasks

This article outlines common machine learning tasks and methods for solving problems, with suggestions for improvement. It also includes a list of key machine learning tasks, which can be further briefed in the article. The article encourages comments and suggestions on important points and apologizes for any types.

Choose features
Classification by Regression
Estimating density and clustering
Decrease of dimensions.
Testin
Examining and coordinating.

E. Purpose

The purpose of the project is identifying water quality parameters using ml can be multifaceted and could include several objectives:

Environmental Monitoring and Management
Early Warning System
Public Health Protection
Disease control and Response
Research And Analysis
Accurate Results.

F. Objective

The Aim of project is to Create an IoT-based Monitoring System: Gathering, sending, and evaluating data on water quality parameters in aquaculture settings is the main goal of designing and creating an IoT-based monitoring system. the performance of the proposed IoT system in terms of data accuracy, reliability, scalability, and cost-effectiveness compared to existing monitoring methods.

G. Outcome

The outcome for this project is to predict Water Quality Parameters using Ph, Temperature, Humidity, Turbidity, and using machine learning technique to find out that water is consumable or not.

II. LITERATURE SURVEY

[1] Implementing a Wireless Sensor Network for Real-Time Water Quality Monitoring in Overhead Tanks (2007) This study presents the application of Wireless Sensor Network (WSN) technology for online, real-time water quality monitoring. A wireless sensor network is made up of many networking-capable sensor nodes that are placed at various overhead tanks and bodies of water within a certain region to monitor the quality of the water.

[2] stresses both horizontally and vertically produced by water waves as they move are monitored using fiber optic pressure sensing arrays: (2019) It has been established that distributed pressure sensor arrays made of fiber Bragg gratings can monitor dynamic subsurface pressures under water waves in a wave tank in real time. Periodic wave trains passing above were used to measure the horizontal and vertical pressures inside the tank using two sensor arrays. With the 90 and 35 sensing elements in the horizontal and vertical arrays, correspondingly, separated by 1-cm intervals, allowing for very accurate spatial resolution in both directions. Video image analysis and commercial piezo-electric pressure sensors were used to validate the pressures measured with the Fiber optic array. Programming was done on the wave tank paddle to produce wave-trains with different peak-to-trough lengths ranging from 5 to 30 cm.

[3] Metro's ZigBee-based WSN-based overhead tank monitoring system (2019) This research presents a ZigBee-based wireless sensor network (WSN) monitoring system for metro water overhead tanks. WSN applications make wired network installation more feasible. The suggested system is intended to monitor the distribution of water and Based on the water level that is available, either fill the water tanks or stop pumping.To regulate the pumping motor and check the water level, a prototype remote node was created. The controller node received this set of parameters. We may use a controller unit to operate tanks across a large region by using ZigBee-based WSN.

[4] A Communication Gateway with Reconfigurable Properties for Distributed Embedded Control Systems (2012) This paper's major goal is to suggest a communication gateway device that complies with domain standards and is programmable and adaptable. Therefore, this method is based on the IEC 61499 reference model for distributed and reconfigurable automation and utilizes the concept of service interface function blocks together with high-level communication patterns to create a hardware-independent access to communication services.

[5] Khan et al. (2019) developed a water quality monitoring and management system based on Iot that integrates various sensors and utilizes machine learning for predictive analysis. This study suggests an Iot system used for managing and water quality monitoring. integrating various sensors including pH, temperature, and turbidity. Machine learning algorithms are applied for predictive analysis to determine water quality.

[6] A. H. Abdulwahid focused on developing an IoT system for managing and monitoring water quality is proposed in this research. specifically tailored for rural areas, employing pH, temperature, and turbidity sensors alongside machine learning for predictive modeling. The authors offer an Internet of Things (IoT)-based turbidity, temperature, and pH sensor-based water quality monitoring system intended for rural regions. Machine learning methods are used to forecast the state of the water quality.

[7] Pasika, Sathish, and Sai Teja Gandla (2020) developed a "Smart Water Quality Monitoring System with Cost-Effective Using IoT," as published in Helion. The system aims to provide a cost efficiency solution for monitoring water quality by utilizing IoT technology. The paper likely discusses the integration of various sensors into an IoT framework for continuous monitoring of water parameters. The focus is likely on leveraging IoT capabilities to enable real-time data collection and analysis, thus enhancing water quality management efforts. The system's cost-effectiveness may stem from the use of affordable IoT devices and open-source software solutions.

[8] Mukta, Islam, Barman, Reza, and Hossain Khan (2019) presented an "IoT based Smart Water Quality Monitoring System" In the course of the 4th International Conference on Computer and Communication Systems (ICCCS) organized by IEEE in 2019. The system likely integrates IoT technology with water quality monitoring to enable real-time data collection and analysis. This paper may discuss the deployment of various sensors such as pH, temperature, and turbidity sensors to continuously monitor water parameters. The focus is likely on leveraging IoT capabilities for remote monitoring and management of water quality, aiming to improve efficiency and accuracy in identifying water quality issues. The system may also incorporate communication technologies to enable data transmission and visualization for stakeholders.

[9] Amruta and Satish (2013) presented a "Solar powered water quality monitoring system using wireless sensor network" at the Multi-Conference on Compressed Sensing, Automation, Computing, Communication, and Control (iMac4s). The system likely utilizes solar power as its energy source, making it sustainable and environmentally friendly. It is designed to monitor water quality parameters using a WSN, enabling remote and continuous monitoring of water quality in various locations. The paper may discuss the integration of sensors such as pH, temperature, and turbidity sensors to measure key water quality indicators. Additionally, it may address the implementation of communication protocols within the WSN for data transmission and retrieval. This system could offer a practical solution for remote areas or regions with limited access to conventional power sources, ensuring continuous monitoring of water quality for environmental and public health purposes.

[10] A work titled "Smart Water Quality Monitoring System for Real Time Applications" was published by Tha. Sugapriyaa, S. Rakshaya, K. Ramyadevi, M. Ramya, and P.G. Rashmi (2018) in the International Journal of Pure and Applied Mathematics, Volume 118. The creation of an intelligent water quality monitoring system intended for real-time applications is probably covered in the article. In order to provide continuous monitoring of water quality indicators like pH, temperature, and turbidity, it could concentrate on combining several sensors and technologies. Real-time data gathering and analysis is perhaps the system's main focus, allowing for the early identification of problems with water quality and rapid response. The use of data visualization and communication protocols to provide consumers easily accessible and useful insights may also be included in the paper.

III. EMBEDDED SYSTEM DESIGN

A. Overview of Embedded System

An embedded system is a kind of special computer system designed to do more than one specified tasks, often with limited real-time computing capabilities. it is integrated into a full gadget that also includes mechanical components and hardware. On the other hand, a personal computer or other general-purpose computer can do a wide range of functions based on its programming. Because they oversee so many everyday objects, embedded systems have grown in importance in recent years. Embedded systems provide serval functions:

- Monitor The Environment
- Control The Environment
- Transform The Information

ISSN (Online) 2394-2320
International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 11, Issue 8, August 2024

B. Block Diagram

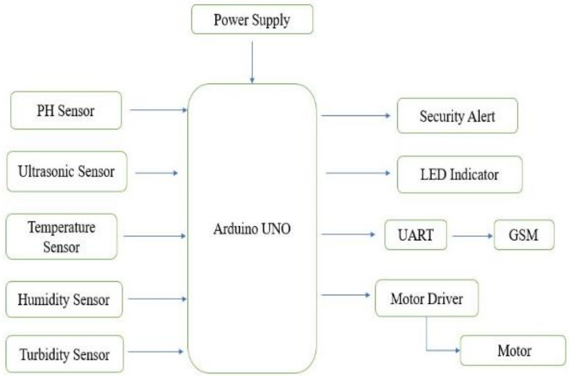

Fig3.1 Block Diagram for Iot

IV. METHODOLOGY

The Methodology involved in Water Quality Parameters prediction using machine Learning:

- Dataset Collection
- Pre-processing
  - Data cleaning
  - Data transformation
  - Data selection
- Algorithms - Random Forest, Linear Regression

A. Dataset Collection

Gathering data is one of the most important phases in building a machine learning model. we collect the information in accordance with published guidelines. There are also some unnecessary data in the dataset. Thus, to create the ideal data set for the algorithm, we must first preprocess the data. Dataset collection in this paper consists of Ph variables, Hardness, Temperature, Water level, Humidity, Turbidity, portability these are the set of data which are used in this implementation and for predicting the output we are using machine learning techniques to find out whether the water is Portable or not.

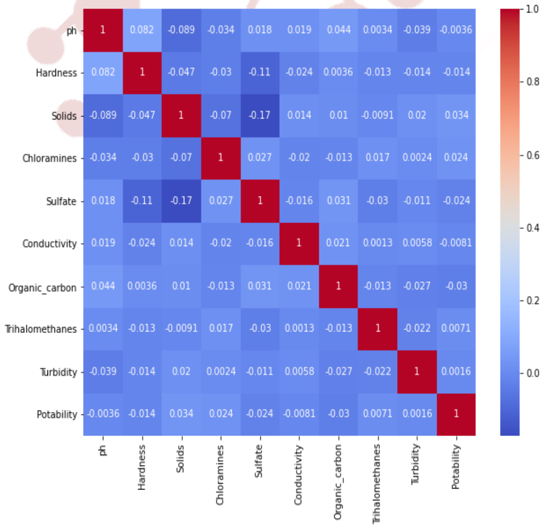

Fig 4.1. plot graph

B. Pre-processing

It involves gathering task-related data based on certain variables that are intended to be analyzed and yield useful results. Nevertheless, a portion of contained data might be noisy nose, meaning that it may include incomplete, erroneous, and inaccurate values.

Therefore, the data processing is essential before analyzing it and drawing conclusions. When preparing a dataset for machine learning algorithms to forecast water quality, data preprocessing is an essential step. It entails several steps intended to prepare the data for analysis by changing and cleaning it. Managing missing values, outliers, and inconsistent datasets is one of the main preprocessing issues. One way to deal with missing values is to either remove them or impute them using machine learning algorithms or statistical techniques. Statistical approaches are used to identify outliers, or data points that significantly different from rest of the sample. These outliers can then be transformed or removed. To avoid certain features from predominating over others during model training, numerical features are scaled to guarantee that they are on a similar scale.

It is necessary to convert categorical information into numerical values using methods like label or one-hot encoding. To minimize overfitting and enhance model performance, feature selection is done to choose the most pertinent features for model training. To meet the assumptions of machine learning techniques, data transformation techniques such as the logarithmic transformation or the Box-Cox transformation are used to alter the distribution or feature structure. Techniques for resampling or creating synthetic data are used to handle imbalanced datasets, which have unbalanced distributions among the classes. Finally, to accurately assess model performance, the dataset is divided into different steps train datasets, validate, and test datasets. In general, preprocessing data makes sure that the dataset is uniform, clean, and suitable for machine learning research, which eventually improves the precision and dependability of predictions about water quality.

C. Data Cleaning

Data cleansing consists of Complete missing values, reduce noise in the data, find and eliminate outliers, and fix inconsistencies. Data Cleaning is like organizing a disorganized space before having visitors. It entails locating and correcting flaws and inconsistencies in your dataset, including formatting issues, duplicate entries, and missing values. By Data Cleaning, you can ensure that it is trustworthy, correct, and ready for analysis. It's like beginning again with your dataset: you must trust the results of your research to make better judgments based on the facts.

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 11, Issue 8, August 2024

Steps In data Cleaning:

Data Collection: The process begins with the collection of various data sources, including Ph values, demographic information, healthcare capacity, Water Usages. And more.

Data Inspection: The collected data is inspected to identify potential issues, such as missing values, outliers, duplicate entries, and data format inconsistencies.

Outlier Detection: The Data points which is significantly different from bulk of the data which are called outliers. In Water Quality it is used to predict the water can be drinkable or waste.

Data Format Consistency: Data from various sources may have different formats. Data cleaning includes converting data into a consistent format for analysis.

Normalization and Scaling: In some cases, To make sure that numerical characteristics fall on the same scale, they may need to be scaled or normalized. This is important for machine learning algorithms that are sensitive to feature scales, like Ph Values.

Dealing with Duplicate Data: Identifying and removing duplicate records ensures that the dataset is free from redundancies that could lead to biased analysis.

D. Data Transformation

Data transformation is essential for enhancing the quality of the data and getting it ready for analysis or modeling, reducing noise, and ensuring that it meets the requirements of the chosen analytical techniques. By applying appropriate transformation techniques, you can enhance the interpretability, efficiency, and effectiveness of your data analysis or modeling efforts.

E. Algorithms

KNN (K-Nearest Neighbors)

An Internet of Things (IoT) and machine learning (ML) water quality prediction system uses the K-Nearest Neighbors (KNN) method. Here’s a high-level overview:

Data collection: Gather information on temperature, pH, humidity, turbidity, and ultrasonic sensor readings using Internet of Things (IoT) sensors. These factors are essential in assessing the water's quality.

Data Analysis: After being gathered, the data is entered into a database and sent for examination. For this, you can apply the KNN algorithm. Since KNN is a non-parametric method, it doesn't make any assumptions about the data it uses to function.

Prediction: Based on the gathered sensor data, the KNN algorithm will forecast if the water is potable or not.

Alerts: The system notifies the user of an alarm if any of the parameters drop below the set values. This allows the consumer to know in advance if their home tanks are contaminated with water.

```
from sklearn.neighbors import KNeighborsClassifier
from sklearn.model_selection import train_test_split
import pandas as pd

# Load your dataset
# df = pd.read_csv('sensor_data.csv')

# Split the data into input 'X' and output 'y'
# X = df[['temperature', 'ph', 'humidity', 'turbidity', 'ultrasonic']]
# y = df['potable']

# Split the dataset into training and testing sets
# X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Initialize the model
# model = KNeighborsClassifier(n_neighbors=3)

# Train the model
# model.fit(X_train, y_train)

# Make predictions
# predictions = model.predict(X_test)
```

Fig 4.2. KNN Algorithm

Based on the following phases, the KNN algorithm operates:

1. Select the K Value: Choose the quantity of closest neighbors(K).
2. Determine Distance: Determine the separation between each new and all of the current data points. The most often used formula for The Euclidean distance, which may be stated as follows, is used to calculate distance.

d=root(x2−x1)^2+(y2−y1)^2.

3. Determine Neighbors: Using the computed distances, determine the K closest neighbors.
4. Assign Class: Place the newly added data point in the class where its K nearest neighbors have the greatest frequency.

Recall that the suitability of your ML model and the calibre of your data will determine how accurate your forecasts are. It's also crucial to remember that, even if this approach can give a reliable indicator of the quality of the water, potability should still be determined by traditional methods of water testing, not by it. This Algorithm is used to find out the prediction of water Quality.

Naive Bayes

Based on the Bayes theorem, the probabilistic classification method known as Naive Bayes relies on the "naive" assumption of feature independence. It is often used in many different applications, such as spam filtering, text categorization, and medical diagnosis. In the context of water quality prediction using IoT sensors, Naive Bayes can be applied to classify water samples as portable or non-portable based on measurements of parameters like temperature, pH, humidity, turbidity, and ultrasonic sensor readings.

The foundation of naive Bayes is the Bayes theorem, which expresses the likelihood of a hypothesis given available data. In terms of math, it is stated as:

P(A|B) = P(B/A).P(A)/P(B) Where

P(A/B) is the posterior probability of hypothesis A given evidence B.

P(B/A) is the likelihood of evidence B given hypothesis A.

P(A) is the prior probability of hypothesis A.

P(B) is the probability of evidence B.

SVM (Support Vector Machine)

A potent supervised learning method for regression and classification issues is the Support Vector Machine (SVM). SVM can be used to categorize water samples as portable or non-portable based on characteristics like temperature, pH, humidity, turbidity, and signals from ultrasonic sensors in the context of water quality prediction utilizing machine learning and IoT sensors.

The function of SVM is defined as:

$$f(x)=\text{sign}\ (\textstyle\sum n \text{ to } i=1\ \alpha i\ yi\ K(xi,x)+b)$$

where x is the input feature vector, K(xi,x) is the kernel function, yi is the class label of the ith training sample, b is the bias term, and αi are the Lagrange multipliers that were found during training.

Based on the supplied sensor measurements, SVM can reliably identify water samples as portable or non-portable in the context of predicting the quality of the water. To guarantee the safety and purity of the water, precise predictions can be made by utilizing the SVM algorithm's capacity to manage intricate relationships and high-dimensional data.

V. SYSTEM REQUIREMENTS & ARCHITECTURE

Hardware requirements

- Arduino Micro Controller
- Ultrasonic Sensor
- Ph Sensor
- Humidity Sensor
- Temperature Sensor
- Motor
- Relay
- Turbidity Sensor

Software Requirements

- Operating System: Windows10
- IDE: anaconda Navigator
- Coding Language: Python
- Embedded C
- HTML
- Porteous Simulation

Hardware Implementation

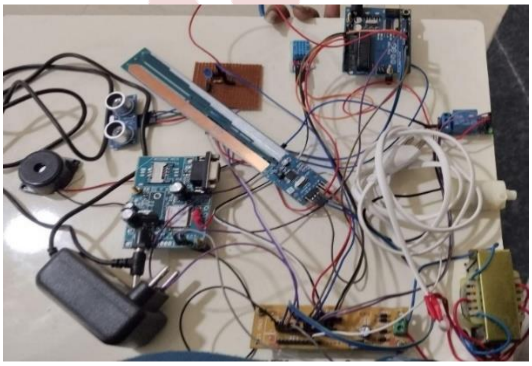

Fig 5.1 Hardware Implementation

This Hardware Consists of Arduino, Ultrasonic sensor, Ph Sensor, GSM, Alarm, Relay, Motor.

Arduino: Arduino has a wider spectrum of physical object perception and operation. Based on a simple microcontroller board, this physical computing platform is open-source and comes with a programming environment for creating board applications. Using a range of switches and sensors as inputs and various lights, motors, and other physical outputs as controllers, Arduino may be used to build interactive things.

Arduino projects may run alone or in tandem with other software applications like as Processing, Flash, and Max MSP. The boards may be purchased pre-assembled or can be put together manually using the free open-source IDE software. A microcontroller board based on the ATmega328 is called the Arduino Uno (datasheet).It contains six analogy inputs, a 16 MHz crystal oscillator, 14 digital input/output pins (six of which may be used as PWM outputs), a USB port, a power jack, an ICSP header, and a reset button. The only things needed to power the microcontroller are a battery, an AC-to-DC converter, or a USB connection to a computer. Everything needed to support the microcontroller is supplied. The FTDI USB-to-serial driver chip is not used by the Uno, in contrast to all previous boards. The Atmega8U2 is configured to function as a serial-to-USB converter instead. Ethernet There are two ways you can use an Arduino connection: 1) Server 2) Client

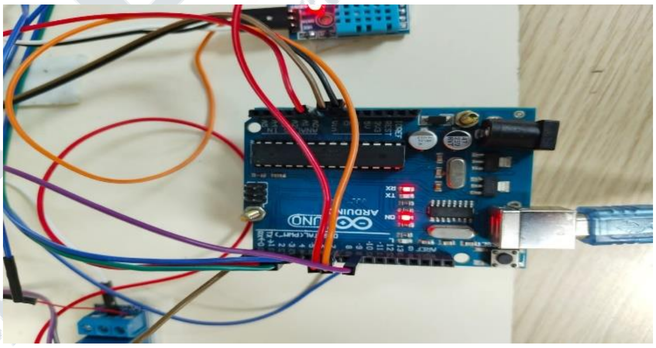

Fig 5.2 Arduino

Ph Sensor: Ph is an essential component to take into account when evaluating the acid-base balance of water. It also acts as a barometer to determine how acidic or alkaline the water is. The maximum pH range that the WHO has recommended is 6.5 to 8.5. The present experiment's ranges were 6.52–6.83, which match the WHO-recommended range.

Fig 5.3 Ph Sensor

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 11, Issue 8, August 2024

Relay: An electrically powered switch is called a relay. The magnetic field produced by the relay's coil pulls a lever, changing the switch contacts. Relays are double throw (changeover) switches because they have two switch positions since the coil current may be turned on or off. Two circuits may be switched with the use of relays, one of which can function completely independently of the other. For example, relays may be used to switch 230V AC mains circuits in low voltage battery circuits. The single internal electrical connection is made via the mechanical and magnetic link of the relay between the two circuits.

Fig 5.4 Relay

Temperature: Temperature is the environmental parameter that is most often monitored. This may be predicted as temperature affects most mechanical, chemical, electrical, and biological systems. Certain chemical reactions, biological activities, and electrical circuits perform best in temperature ranges that are restricted. Since temperature is one of the variables that is most often checked, it should not be surprising that there are several ways to measure it. It is possible to measure temperature using radiation either directly from the heating source or remotely without coming into touch with it. Among the many kinds of temperature sensors that are now on the market are thermocouples, semiconductor sensors, thermistors, infrared, and resistance temperature detectors (RTDs).

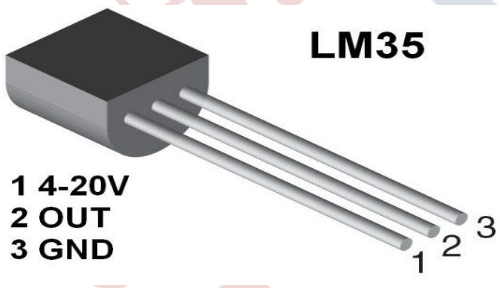

Fig 5.5 Temperature Sensor

Features of LM35 Temperature Sensor:

- Directly calibrated in ˚ Celsius (Centigrade)
- Rated for the whole range of l −55˚ to +150˚C.
- Well suited for distant applications
- Minimal expense because of wafer-level cutting.
- Functions between 4 and 30 volts.
- Minimal self-heating, with an average nonlinearity of ±1/4˚C

Like other integrated circuit temperature sensors, the LM35 is simple to attach. Its temperature will within around 0.01˚C of the surface temperature when it is adhered to or established on a surface.

Ultrasonic Sensor: The difference in the transit periods of ultrasonic pulses propagating in and against the direction of flow is detected by ultrasonic transit time flow meters. The average fluid velocity along the ultrasonic beam's course is determined by this time difference. Both the averaged fluid velocity and the sound speed can be computed using the absolute transit times. The following equations can be written using the two transit periods, the distance between the transmitting and receiving transducers, and the inclination angle:

$$v = \frac{L}{2\sin(\alpha)} \frac{t_{up} - t_{down}}{t_{up}\, t_{down}} \quad \text{and} \quad c = \frac{L}{2} \frac{t_{up} + t_{down}}{t_{up}\, t_{down}}$$

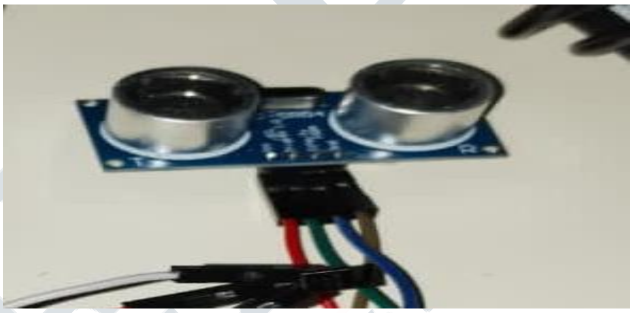

Fig 5.6 Ultrasonic Sensor

GSM (Global System for Mobile Communication): GSM technology plays a crucial role in enhancing the functionality and connectivity of IoT-based water quality monitoring systems. By enabling reliable data transmission, remote monitoring, alerting, and control features, GSM helps ensure timely detection of water quality issues and facilitates effective decision-making and intervention strategies. It is mainly used for:

1) Data Transmission
2) System Alters
3) System Notifications.

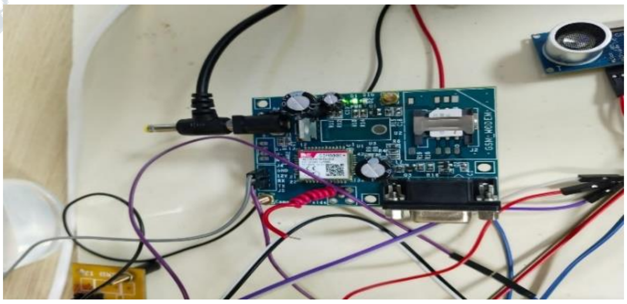

Fig 5.7 GSM

Turbidity Sensor: A turbidity sensor, sometimes referred to as a turbidimeter, is a tool used to gauge how hazy or cloudy a fluid is due to suspended solids or particles. In many different sectors and uses, such as water treatment, wastewater management, environmental monitoring, and beverage manufacturing, turbidity is a crucial metric. With the help of the turbidity sensor's quantitative turbidity measurements, water quality and other fluid properties can be

IFERP
Explore Your Research Journey...
ISSN (Online) 2394-2320
International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 11, Issue 8, August 2024

accurately assessed and controlled.

Fig 5.8 Turbidity Sensor

Architecture Diagram:

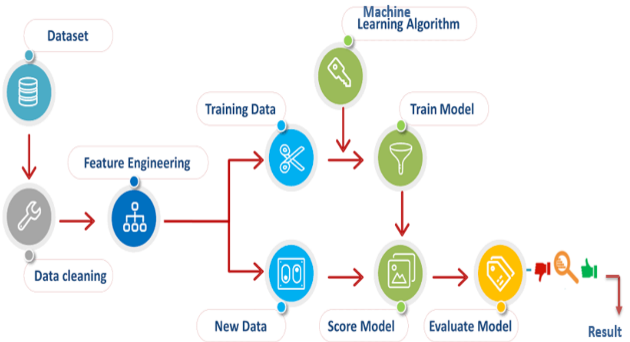

Fig 5.8: Architecture Diagram

Sequence Diagram

A software engineering tool called a sequence diagram, sometimes referred to as a system sequence diagram (SSD), is used to show the order of process interactions and messages that are sent back and forth. It is frequently connected to the realization of use cases in the 4+1 architectural view model of a system that is still in development. Sequence diagrams highlight external actor-generated events, their sequence, and potential inter-system events. They emphasize events that occur outside of system boundaries and treat all systems as opaque black boxes. For both the primary success scenario and frequent or complicated alternative scenarios, a system sequence diagram is necessary.

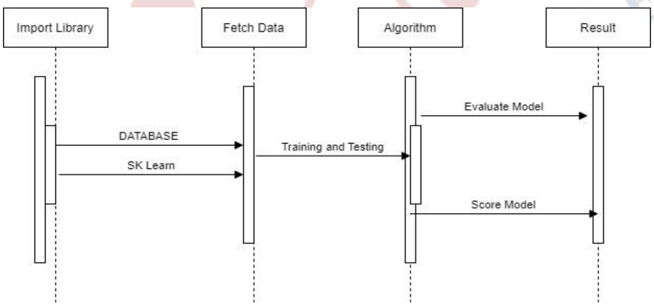

Fig 5.9: Sequence Diagram

Flow Diagram

A diagram that shows a flow or a collection of dynamic relationships inside a system is collectively referred to as a flow diagram. In addition to being a synonym for flowcharts, the phrase "flow diagram" is also occasionally used to refer to a flowchart's opposite.

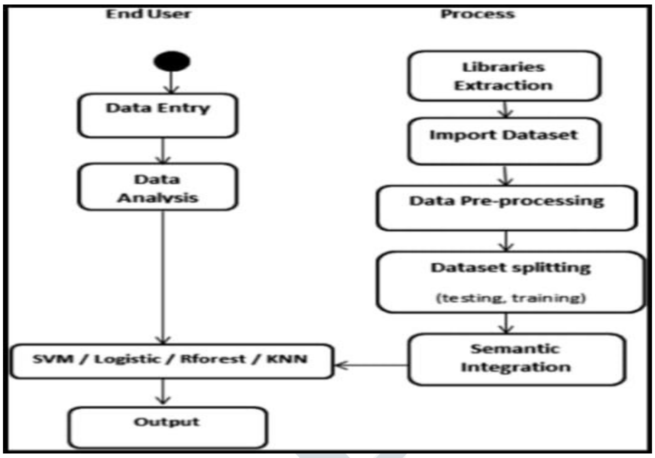

Fig 5.10 Flow Diagram

VI. RESULT AND DISCUSSION

This section assesses how well the monitoring system maintains water quality parameters within desired ranges, presenting the findings and analysis of the data collected by the IoT system. It also discusses any trends, correlations, or anomalies observed and their implications for aquaculture management. several water sources, and each sample's pH, temperature, electric conductivity, and turbidity were measured through testing. Three categories are used to classify these water sources: drinkable, impure, and natural. The retrieved values corresponding to the four physical parameters for every water sample are shown in Figs. 7 through 10.

A. Analysis of Physical Calculation

Parameters	Water Source	N	Min.	Med.	Max.	WHO SV	% not within SV
Temperature	Natural	23	24.11	25.11	27.2		-
	Impure	16	16	24.6	44.63		-
	Potable	21	23	25.76	27		-
	All	60	16	25.09	44.63		-
Conductivity	Natural	23	0.12	0.34	6.45	0.3-0.8 mS/cm	69.57
	Impure	16	0.12	0.57	20.44		81.25
	Potable	21	0.12	0.48	19.89		66.67
	All	60	0.12	0.44	20.44		71.67
PH	Natural	23	7.3	9.23	9.88	6.5-8.5	82.61
	Impure	16	7.66	8.9	10.26		68.75
	Potable	21	7.99	9.12	9.89		85.71
	All	60	7.3	9.15	10.26		80
Turbidity	Natural	23	0	0	2023	<5 NTU	13.04
	Impure	16	0	989.58	3000		56.25
	Potable	21	0	0	2902.89		33.33
	All	60	0	0	3000		31.67

Fig 6.1: Physical Calculation

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 11, Issue 8, August 2024

B. Prediction of Water Using PH

The pH range that the World Health Organization recommends for drinking water is between 6.5 and 8.5. Nearly 80% of the analyzed water samples are shown in the table to be alkaline and outside of the advised pH range, indicating the presence of carbonates and limestone in the water samples. As a result, an excess of alkanity in the human body can lead to metabolic alkalosis, gastrointestinal distress, and skin irritation. One important factor to consider while evaluating the water's microbiological quality is turbidity. The guidelines state that turbidity levels below 5 NTU are acceptable. When compared to natural water samples, the results indicate that unclean water has a higher turbidity value.

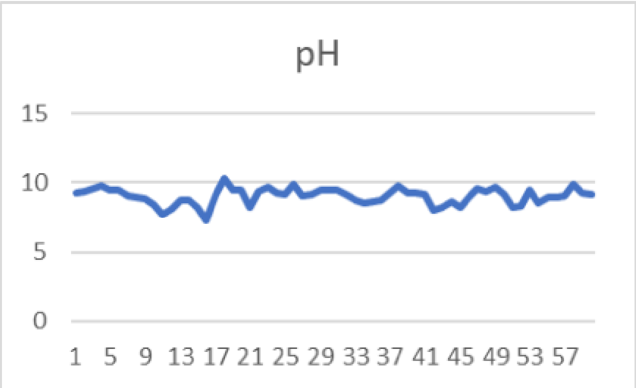

Fig 6.2: Ph Graph

Ph Values of Drinkable Water: - 6.5 to 8.5
Acidic: - <6.5
Basic: - >8.5
Formulae: - PH = -log[H_3O^+]

C. Prediction of Water Using Temperature

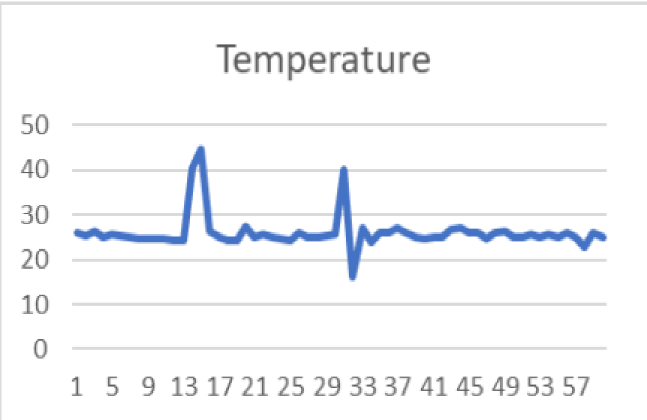

Fig 6.3 Temperature Graph

Temperature for Drinkable Water : 16 to 50$^{\cdot c}$

D. Accuracy Using Experimental Calculations

Water Sample	Temparature	EC	pH	Turbidity	System's prediction
1	25	0.14	8.2	0	Drinkable
2	25.67	0.19	8.3	0	Drinkable
3	25.03	0.89	9.5	0	Drinkable
4	25.76	3.89	8.5	1791	Not Drinkable
5	24.99	2.9	8.9	566	Not Drinkable
6	26	0.31	9	0	Drinkable
7	25	4.24	9.1	2023	Not Drinkable
8	23	4.39	9.89	0	Drinkable
9	25.99	6.45	9.29	1520	Not Drinkable
10	24.89	0.18	9.11	0	Drinkable

Fig 6.4 Experimental Calculations

E. Accuracy

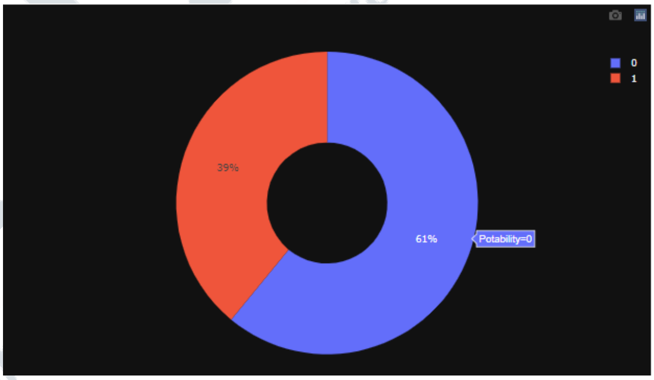

Fig 6.5 Accuracy Analysis

Accuracy of Drinkable Water: 61%
Accuracy of Non-Drinkable Water: 39%

F. Accuracy Using Algorithms

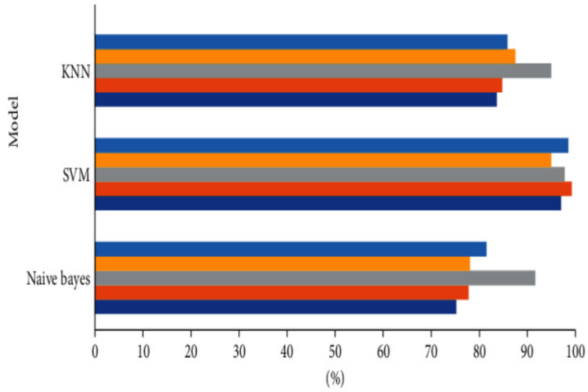

Fig 6.6 Accuracy Using Algorithms

IFERP
Explore Your Research Journey...

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 11, Issue 8, August 2024

Models	Accuracy (%)	Sensitivity (%)	Specificity (%)	Precision (%)	F-score (%)
SVM	97.01	99.23	97.78	94.93	98.54
KNN	83.63	84.73	94.93	87.50	85.84
Naive Bayes	75.20	77.76	91.65	78.08	81.51

Fig 6.7 Accuracy Percentage

G. Dataset

Datasets used for predicting Output.

	Distance	Humidity	Temperature	PH value	label
0	196	79	27.96	6.8	Not-Portable
1	196	79	31.20	6.8	Portable
2	196	79	28.44	6.8	Not-Portable
3	195	79	31.20	6.8	Portable
4	196	79	29.04	6.8	Not-Portable
...	...	...	...	...	...
4492	52	89	28.32	7.8	Not-Portable
4493	106	89	30.84	7.8	Portable
4494	197	89	27.24	7.8	Not-Portable
4495	197	89	31.56	7.8	Portable
4496	50	89	26.76	7.8	Not-Portable

4497 rows × 5 columns

Fig 6.8 Datasets

H. Iot Output

```
COM13
|
Temperture :97.44*C
Overheat
message sent
Humidity value: 54.00
PH Value : 10.08
distance: 165cm
Temperture :10.56*C
Normal
Humidity value: 53.00
PH Value : 9.71
distance: 166cm
Temperture :12.84*C
Normal
```

Fig 6.9 Iot Output

I. Output

```
In [51]: df
Out[51]:
```

	Distance	Humidity	Temperature	PH value	label
0	196	79	27.96	6.8	Not-Portable
1	196	79	31.20	6.8	Portable
2	196	79	28.44	6.8	Not-Portable
3	195	79	31.20	6.8	Portable
4	196	79	29.04	6.8	Not-Portable
...	...	...	...	...	...
4492	52	89	28.32	7.8	Not-Portable
4493	106	89	30.84	7.8	Portable
4494	197	89	27.24	7.8	Not-Portable
4495	197	89	31.56	7.8	Portable
4496	50	89	26.76	7.8	Not-Portable

4497 rows × 5 columns

```
In [52]: result= gnb.predict(np.array([[196,79,27.96,6.8]]))
         result[0]
Out[52]: 'Not-Portable'

In [53]: result= gnb.predict(np.array([[106,89,30.84,7.8]]))
         result[0]
Out[53]: 'Portable'
```

REFERENCES

[1]. C. Sowmya, C. D. Naidu, R. P. Somineni and D. R. Reddy, "Implementation of Wireless Sensor Network for Real Time Overhead Tank Water Quality Monitoring," 2017 IEEE 7th International Advance Computing Conference (IACC), Hyderabad, India, 2017, pp. 546-551, doi: 10.1109/IACC. 2017.0118.

[2]. J. W. Arkwright, I. D. Underhill, S. A. Maunder, A. Jafari, N. Cartwright and C. Lemckert, "Fiber Optic Pressure Sensing Arrays for Monitoring Horizontal and Vertical Pressures Generated by Traveling Water Waves," in IEEE Sensors Journal, vol. 14, no. 8, pp. 2739-2742, Aug. 2014, doi: 10.1109/JSEN.2014.2311806.

[3]. B. P. Reddy and P. Eswaran, "Metro overhead tanks monitoring system using ZigBee based WSN," *2013 IEEE Conference on Information & Communication Technologies*, Thuckalay, India, 2013, pp. 367-371, doi: 10.1109/CICT. 2013.6558122.

[4]. F. Andrén, T. Strasser, A. Zoitl and I. Hegny, "A reconfigurable communication gateway for distributed embedded control systems," *IECON 2012 - 38th Annual Conference on IEEE Industrial Electronics Society*, Montreal, QC, Canada, 2012, pp. 3720-3726, doi: 10.1109/IECON. 2012.6389299.

[5]. M. H. Jewel and A. Al Mamun, "Internet of Things (IoT) for Water Quality Monitoring and Consumption Management," *2022 4th International Conference on Sustainable Technologies for Industry 4.0 (STI)*, Dhaka, Bangladesh,2022, pp.1-5,doi: 10.1109/STI56238.2022.10103355.

[6]. A. H. Abdulwahid, "IoT Based Water Quality Monitoring System for Rural Areas," *2020 9th International Conference on Renewable Energy Research and Application (ICRERA)*,Glasgow,UK,2020,pp.279282,doi:10.1109/ICRER A49962.2020.9242798.

[7]. Pasika, Sathish, and Sai Teja Gandla.(2020), "Smart Water Quality Monitoring System with Cost-Effective Using IoT."Heliyon,vol.6,no. 7, doi:10.1016/j.heliyon.2020.e04096

[8]. M. Mukta, S. Islam, S. D. Barman (2019), A. W. Reza and M. S. Hossain Khan, "Iot based Smart Water Quality Monitoring System," 2019 IEEE 4th International Conference on Computer and Communication Systems (ICCCS), pp. 669-673,doi:10.1109/CCOMS.2019.8821742.

[9]. M. K. Amruta and M. T. Satish (2013), "Solar powered water quality monitoring system using wireless sensor network," International Mutli-Conference on Automation, Computing, Communication, Control and Compressed Sensing (iMac4s), pp. 281-285, doi: 10.1109/iMac4s.2013.6526423

[10]. Tha. Sugapriyaa, S. Rakshaya, K. Ramyadevi, M. Ramya, P.G. Rashmi (2018), Smart Water Quality Monitoring System for Real Time Applications, International Journal of Pure and Applied Mathematics, Volume 118, No. pp 1363-1369

[11]. Huang J., Liu N., Wang M., Yan K. Application WASP model on validation of reservoir-drinking water source protection areas delineation. 2010 3rd International Conference on Biomedical Engineering and Informatics; October 2010; Yantai, China. pp. 3031–3035.

[12]. Warren I. R., Bach H. K. MIKE 21: a modelling system for estuaries, coastal waters and seas. *Environmental Software.* 1992;7(4):229–240. doi: 10.1016/0266-9838(92)90006-P.

[13]. Hayes D. F., Labadie J. W., Sanders T. G., Brown J. K. Enhancing water quality in hydropower system operations. *Water Resources Research.* 1998;34(3):471–483. doi: 10.1029/97WR03038.

[14]. Tang G., Li J., Zhu Z., Li Z., Nerry F. Two-dimensional water environment numerical simulation research based on EFDC in Mudan River, Northeast China. 2015 IEEE European Modelling Symposium (EMS); October 2015; Madrid, Spain. pp. 238–243.

[15]. Hu L., Zhang C., Hu C., Jiang G. Use of grey system for assessment of drinking water quality: a case S study of Jiaozuo city, China. 2009 IEEE International Conference on Grey Systems and Intelligent Services (GSIS 2009); November 2009; Nanjing, China. pp. 803–808.

[16]. Batur E., Maktav D. Assessment of surface water quality by using satellite images fusion based on PCA method in the Lake Gala, Turkey. *IEEE Transactions on Geoscience andRemoteSensing.* 2019;57(5):2983–2989. doi: 10.1109/ TGRS.2018.2879024.

[17]. Jaloree S., Rajput A., Sanjeev G. Decision tree approach to build a model for water quality. *Binary Journal of Data Mining & Networking.* 2014;4:25–28.

[18]. Liu J., Yu C., Hu Z., et al. Accurate prediction scheme of water quality in smart mariculture with deep Bi-S-SRU learning network. *IEEE Access.* 2020;8:24784–24798. doi: 10.1109/ACCESS.2020.2971253