Pokhara University

Nepal Engineering College

A Report on

IOT BASED WATER QUALITY MEASUERMENT USING MEACHINE LEARNING

Under The Supervision of:

Asst.Prof.Deepesh Prakash Guragain

Assoc.Rupesh Dhai shrestha

SUBMITTED BY:

Aashish Timalsina(019-401)
BikeshBhatta(019-421)
Shambhu sha (019-414)

SUBMITTED TO:

Department of Electronics and Communication Engineering

Nepal Engineering College

Bhaktapur, Nepal

April, 2024

DEPARTMENTAL ACCEPTANCE

The project report entitled “Iot Base Water Quality Measurement Using Meachine Learning”, submitted by Aashish Timalsina, Bikesh Raj Bhatta and Shambhu Sha in partial fulfillment of the requirement for the Bachelor’s degree in Electronics and Communication Engineering has been accepted as a bonafide record of work independently carried out by the group in the department.

……..
Asst.Prof.Deepesh Prakash Guragain
Project Coordinator
Department of Electronics and Communication Engineering,
Nepal Engineering College,
Bhaktapur, Nepal.

ACKNOWLEDGEMENT

It gives us immense pleasure in presenting this project report on “IOT BASE WATER QULATIY MEASUREMENTS USING MEACHINE LEARNING”. We express our gratitude to ‘Nepal Engineering College’ and ‘Department of Electronics and Communication’ for providing us the opportunity to work on this project.

We would also like to thank our project supervisors Assoc.Rupesh Dhai shrestha, Asst. Prof. Deepesh Prakash Guragain for their critical advice and guidance which this project would make have been possible. Also, we want to thank Assoc.prof Sabitri Thripati and Department of water supply and sewerage management, National Academy of Science and Technology (NAAST), Kathmandu University, NAWAS, ENFOS for their valuable comments data and feedbacks

Last but not the least we place a deep sense of gratitude to our seniors and our colleague who have been constant source of inspiration during throughout this project work. We sincerely appreciate the inspiration; support and guidance of all those people who have helped us in making this project a success.

ABSTRACT

World Economic Forum ranked drinking water crisis as one of the global risk, due to which around 200 children are dying per day. Drinking unsafe water alone causes around 3.4 million deaths per year. Despite the advancements in technology, sufficient quality measures are not present to measure the quality of drinking water. By focusing on the above issue, this paper proposes a low cost water quality monitoring system using emerging technologies such as IOT(Internet of Thing) , Machine Learning studied classifiers included Support Vector Machine (SVM), Random Forest (RF), Logistic Regression (LR), Decision Tree (DT), CATBoost, XGBoost,and Cloud Computing which can replace traditional way of quality monitoring. This helps in saving people of rural areas from various dangerous diseases such as fluorosis, bone deformities etc.

The dataset used in the study included around 6000 samples and their meta-data collected over nine years. In addition, precision-recall curves and Receiver Operating Characteristic curves (ROC) were used to assess the performance of the various classifiers. The findings revealed that the CATBoost model offered the most accurate classifier with a percentage of 94.51.

The proposed model also has a capacity to control temperature of water and adjusts it so as to suit environment temperature. Water condition based on four physical parameters i.e., temperature, pH, electric conductivity and turbidity properties. Four sensors are connected with ESP-32in discrete way to detect the water parameters. Extracted data from the sensors are transmitted to a cloud via webpage/mobile app for user proper information.

[Keywords:—Water quality monitoring, Internet of Things, Machine learning, Cloud Computing,Web-basedapplication;Mobile-app]

Table of **Contents**

LIST OF FIGURES

LIST OF TABLES

Chapter 1: INTRODUCTION

1.1 Backgrounds and Statement of Problems

Water is one of the most valuable natural resources that humans have gifted. Water management becomes an important issue especially in industrial, agricultural and other sectors. Most of the people around the world lack behind drinkable water .Research by WHO (World Health Organization) shows that almost 1.4 million of child death can be prevented by providing drinkable water to them. The primary objective of this project is to introduce an intelligent water quality monitoring system in IoT (Internet of Things) platform which would help to monitoring different physical parameters of the drinkable water rather than relying on manual process. Moreover, We need a 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 Wi-Fi module, using which the data is passed over to cloud space and driftnet protocol give information to user at real time.

Monitoring water quality is essential for protecting human health and the environment and controlling water quality. Artificial Intelligence (AI)/Machine learning offers significant opportunities to help improve the classification and prediction of water quality (WQ). In this study, various AI algorithms are assessed to handle WQ data collected over an extended period and develop a dependable approach for forecasting water quality as accurately as possible. Specifically, various machine learning classifiers and their stacking ensemble models were used to classify the WQ data via the Water Quality Index (WQI). The studied classifiers included Support Vector Machine (SVM), Random Forest (RF), Logistic Regression (LR), Decision Tree (DT) and CAT Boost. The challenge lies in developing robust ML models capable of real-time analysis and prediction of water quality parameters, enabling timely intervention to prevent contamination and ensure safe water supply. By addressing this problem, the proposal aims to enhance the efficiency and effectiveness of water quality management systems, contributing to sustainable water resource utilization in the face of increasing environmental challenges.

1.1.1 Classification of water

Based on its source, water can be divided into ground water and surface water. Both types of water can be exposed to contamination risks from agricultural, industrial, and domestic activities, which may include many types of pollutants such as heavy metals, pesticides, fertilizers, hazardous chemicals, and oils.

Water quality can be classified into four types—potable water, palatable water, Contaminated (polluted) water, and infected water. The most common scientific Definitions of these types of water quality are as follows:

1. Potable water*:* It is safe to drink, pleasant to taste, and usable for domestic Purposes.

2. Palatable water*:* It is esthetically pleasing; it considers the presence of chemicals That do not cause a threat to human health.

3. Contaminated (polluted) water*:* It is that water containing unwanted physical, Chemical, biological, or radiological substances, and it is unfit for drinking or Domestic use.

4. Infected water*:* It is contaminated with pathogenic organism.

1.1.2 Water quality standard

Government of Nepal has issued this notice of implementation of National Drinking Water Quality Standards, 2062 under the provision of Water Resources Act, 2049, Clause 18 and Sub Clause 1

(A) National Drinking Water Quality Standard

Table 0-1.1.2 water quality standard

S.N.	Category	Parameters	Units	Concentration Limits
1	Turbidity	NTU	5 (10)	
2	pH			6.5-8.5*
3	Color	TCU	5 (15)	
4	Physical	Taste and Odor		Non-objectionable
5	TDS	mg/L		1000
6	Electrical conductivity (EC)	µS/cm		1500
7	Iron	mg/L		0.3 (3)
8	Manganese	mg/L		0.2
9	Arsenic	mg/L		0.05
10	Cadmium	mg/L		0.003
11	Chromium	mg/L		0.05
12	Cyanide	mg/L		0.07
13	Fluoride	mg/L		0.5-1.5*
14	Lead	mg/L		0.01
15	Ammonia	mg/L		1.5
16	Chloride	mg/L		250
17	Chemical	Sulphate	mg/L	250
18	Nitrate	mg/L		50
19	Copper	mg/L		1

S.N.	Category	Parameters	Units	Concentration Limits
20	Total Hardness	mg/L as CaCO3		500
21	Calcium	mg/L		200
22	Zinc	mg/L		3
23	Mercury	mg/L		0.001
24	Aluminum	mg/L		0.2
25	Residual Chlorine	mg/L		0.1-0.2*
26	E. Coli	MPN/100 ml		0
27	Microbiological	Total Coliform	MPN/100 ml	0 in 95% samples

[*Source: https://wepa-db.net/archive/policies/law/nepal/st01.html]*

1.2 Objective

- Enable continuous monitoring of water quality parameters in real-time to promptly detect any deviations or anomalies.
- Detect and identify contaminants or pollutants in water sources at an early stage to mitigate risks to public health and the environment.
- Develop a scalable and adaptable IoT infrastructure for water quality monitoring.
- Optimize water treatment processes based on real-time data to improve efficiency and effectiveness in maintaining water quality standards.
- Contribute to the protection and preservation of aquatic ecosystems by monitoring and managing water quality effectively.

1.3 Application

- IoT sensors continuously monitor water quality parameters like pH, turbidity, dissolved oxygen, temperature, and chemicals. Machine learning detects anomalies in real-time.

- Machine learning analyzes past IoT sensor data to forecast maintenance needs in water treatment, reducing downtime and ensuring uninterrupted monitoring.

- Machine learning models trained on data collected from IoT sensors can detect the presence of contaminants in water sources, such as heavy metals, pesticides, or organic pollutants.

- IoT devices installed at consumer endpoints can provide real-time feedback on water quality to consumers.

1.4 Overview of proposal

The report is structured into five chapters, each focusing on different aspects of the project. Chapter 1: serves as brief overview of the importance of water quality management. Introduction to IoT and machine learning technologies. Rationale for combining IoT and machine learning for water quality monitoring. It outlines the objectives and scopes of the project while also discussing the applications of the undertaken research.

Chapter 2: provides a comprehensive literature review, exploring existing models and related works in the field, setting the stage for the project's methodology.

Chapter 3: details the methodology employed, presenting basic block diagrams, flow charts, and outlining the hardware and software used in the study, along with any underlying assumptions made.

Chapter 4: Expected Real-time monitoring of water quality parameters. Early detection of water quality issues and contamination events. Improved accuracy and efficiency of water quality management. Reduction in response time to water quality incidents.

Finally, Chapter 5. Recap of the proposed IoT-based water quality measurement system integrated with machine learning. Draws conclusions from the entire project, summarizing key findings and insights gained, Emphasis on the potential benefits and impact of the system.

Chapter 2: Literature Review

2.1 Essential Water quality parameters

Base on the different physical chemical and biological properties water quality will be classification with different parameter.

Table 0-1.1 Water Quality Standard in Nepal

No.	Physical Parameters	Chemical Parameters	Biological Parameters
1	Turbidity	pH	Bacteria
2	Temperature	Acidity	Algae
3	Color	Alkalinity	Viruses
4	Taste and odor	Chloride	Protozoa
5	Solids	Chlorine residual	
6	Electrical conductivity (EC)	Sulfate	
7		Nitrogen	
8		Fluoride	
9		Iron and manganese	
10		Copper and zinc	
11		Hardness	
12		Dissolved oxygen	
13		Biochemical oxygen demand (BOD)	
14		Chemical oxygen demand (COD)	
15		Toxic inorganic substances	
16		Toxic organic substances	
17		Radioactive substances	

2.2 WQI (Water Quality Index)

An index value is calculated for each of water quality parameters, temperature, biological oxygen demand (BOD), total suspended sediment (TSS), dissolved oxygen (DO), and

conductivity. A higher value of each index indicates better water quality.and the following relation was used to compute the WQI:.

$$WQI=\frac{\sum_{i=1}^{N} qi*wi}{\sum_{i=1}^{N} wi} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 1$$

$$Qi=100*\left(\frac{vi-videal}{si-videal}\right) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 2$$

$$Wi=\frac{k}{si} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 3$$

K is the proportionality constant, and the following equation can be used to compute it:

$$K=\frac{1}{\sum\frac{1}{si}}$$

Where *N* denotes the number of the total parameter, *qi* denotes the quality estimate scale for each parameter *i* calculated by Eq. (2)

Table 2.2 WQI and corresponding water quality status

S.No	WQI	Status	Possible usages
1	0 – 25	Excellent	Drinking, Irrigation and Industrial
2	25 – 50	Good	Domestic, Irrigation and Industrial
3	51 -75	Fair	Irrigation and Industrial
4	76 – 100	Poor	Irrigation
5	101 -150	Very Poor	Restricted use for Irrigation
6	Above 150	Unfit for Drinking	Proper treatment required before use.

2.3 Machine Learning Models

2.3.1 Logistic Regression.

- Logistic regression is one of the most popular Machine Learning algorithms, which comes under the Supervised Learning technique. It is used for predicting the categorical dependent variable using a given set of independent variables.

- Logistic regression predicts the output of a categorical dependent variable. Therefore the outcome must be a categorical or discrete value. It can be either Yes or No, 0 or 1, true or False, etc. but instead of giving the exact value as 0 and 1, it gives the probabilistic values which lie between 0 and 1.
- Logistic Regression is much similar to the Linear Regression except that how they are used. Linear Regression is used for solving Regression problems, whereas Logistic regression is used for solving the classification problems.
- Logistic Regression can be used to classify the observations using different types of data and can easily determine the most effective variables used for the classification. The below image is showing the logistic function:

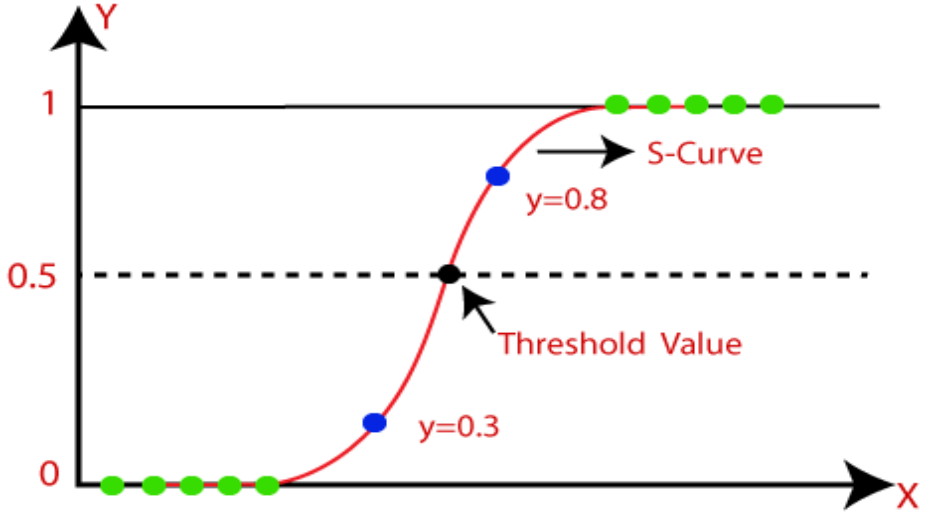


Figure 2.3.1 Logistic Regression analysis

2.3.2 Support Vector Regression (SVR)

Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems. However, primarily, it is used for Classification problems in Machine Learning.

The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future. This best decision boundary is called a hyper plane.

SVM chooses the extreme points/vectors that help in creating the hyper plane. These extreme cases are called as support vectors, and hence algorithm is termed as Support Vector Machine. Consider the below diagram in which there are two different categories that are classified using a decision boundary or hyper plane.

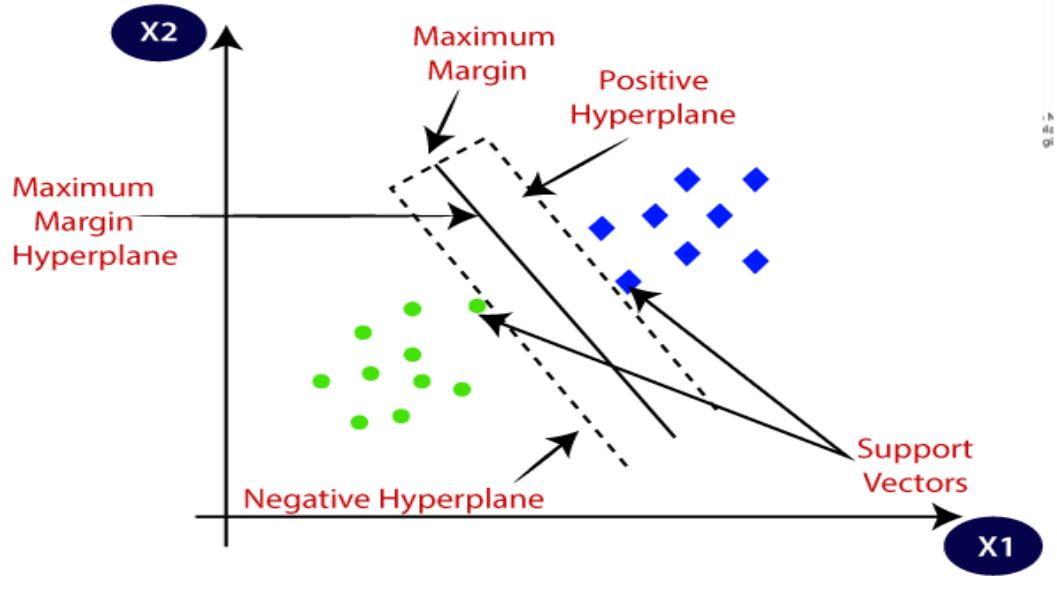


Figure 2.3.2 SVR Analysis

[Source:***https://www.javatpoint.com/machine-learning-support-vector-machine-algorithm]***

2.3.3 CAT Boost

Boost is a gradient boosted decision tree (GBDT) and category feature-based algorithm. Under the context of the GBDT algorithm, this method is better at implementation. The critical issue is dealing with categorical characteristics efficiently and reasonably. Boost is made up of two elements: category variables and boost. When the Boost algorithm analyzes categorical features,

it includes all sample data sets in the learning process. Then Boost organizes all these sample data sets at random and filters out samples from all characteristics with the same category.

Cat Boost overcomes a limitation of other decision tree-based methods in which, typically, the data must be pre-processed to convert categorical string variables to numerical values, one-hot-encodings, and so on. This method can directly consume a combination of categorical and non-categorical explanatory variables without preprocessing. It preprocesses as part of the algorithm. Cat Boost uses a method called ordered encoding to encode categorical features. Ordered encoding considers the target statistics from all the rows prior to a data point to calculate a value to replace the categorical feature.

Another unique characteristic of Cat Boost is that it uses symmetric trees. This means that at every depth level, all the decision nodes use the same split condition.

2.3.4 XGBosst

XGBoost is a gradient boosting algorithm that is widely used in data science. It is an implementation of gradient boosting that is designed to be highly efficient, flexible and portable.

XGBoost is based on the concept of boosting, which is an ensemble learning technique where multiple weak learners (typically decision trees) are combined to form a strong learner. Boosting works by sequentially training weak learners, each focusing on the mistakes made by the previous learners.

. The algorithm was designed with the following goals in mind:

- To be highly efficient
- To be flexible
- To be portable

2.4 Pros and cons of the used classifiers.

Algorithm	**Pros**	**Cons**
Support Vector Machine (SVM)	Even when there is insufficient data, it produces good outcomes.	When working with huge data sets, training takes a long time. In addition, it may be tough to perceive and comprehend due to issues created by personal circumstances and varied weights.
Logistic Regression	It has a low error Provides output probabilities It is simple to use and takes little time to train	When there are correlated attributes, it does not operate well.
XGBoost	Fast to interpret/capable of handling massive datasets If the data is clean, it can avoid over fitting.	Difficult to interpret/ If the data is noisy, the model may overfit.
Decision tree	Automatic Feature Selection/ Easy Visualization	Over fitting proclivity/data sensitivity Extremely slow
CATBoost	Implementation on the CPU is Efficient. Model appliers may be made Extremely quickly.	Features are less powerful in General

2.5. Related Work and Research gap

Koju, N. K., Prasad, T., Shrestha, S. M., &Raut, P[2] . (2014). Drinking water quality of Kathmandu Valley. *Nepal Journal of Science and Technology*, *15*(1), 115–120. doi:10.3126/njst.v15i1.12027Koju,

Pradeepkumar M, Monisha J, Pravenisha R, Praiselin V, Suganya Devi K[3]: entitled ”The Real Time Monitoring of Water Quality in IoT Environment”. This paper discusses not only sensor based system but also it introduces cloud computing architecture into IoT which makes the sensor data accessible worldwide.

During the research we can found that research done by 2009.08.001 Khatiwada, N. R.Takizawa, S., Tran, T. V. N[1]., & Inoue, M. (2002). Ground water contamination assessment for sustainable water supply in Kathmandu Valley, Nepal.*Water Science & Technology*, *46*(9), 147–154. doi:10.2166/wst.2002.0226

M. Valdivia, et.al [4], proposed a model to identify best predictors of THM levels in final potable water and distribution networks, and to decide the rate of change in future. The data between Jan 2011 and Jan 2013 from 93 full-scale Scottish water treatment plants were inspected to recognize the factors causing the advancement of THMs. Multilinear regression algorithms were used to build the models for individual THMs compounds. Pearson's correlation analysis was applied to measure data and concluded that ambient temperature, DOC, and chloride were important in the formation of THMs across Scottish WTPs.

Daigavane et.al.[5], the proposed system, used sensors with Wi-Fi module for conductivity, temperature, water level, pH and turbidity along with power supply were connected to the basic controller-Arduino UNO. The basic controller retrieves the values of the sensor to be assessed by placing the sensors in separate water samples and the data will be forwarded to the cloud using the WI-FI module. The recommended android application will be used to detect sensor values examined via cloud, and alerts will be provided to the user if the value exceeds the threshold value

Atif A, WasaiShaded, Mohammad Hassan, Shamim, Alelaiwi and Anwar Hossain[4]entitled ”A Survey on Sensor-Cloud: Architecture, Applications, and Approaches” discusses about the sensor-cloud infrastructure, approaches, and different layers of transferring generated data by connecting sensors with cloud services.

Nikhil Kedia[6] entitled ”Water Quality Monitoring for Rural Areas-A Sensor Cloud Based Economical Project” This paper not only highlights embedded sensor systems, but also discusses the challenges and economic viability of the system involving Mobile Network Operator and Government. This system directly contacts Government to take action based on the severity of quality issue.

Yafra Khan, et.al [7], proposed a prediction model for water quality using Artificial Neural Network and time-series analysis to support water quality factors. The water quality data from January to March 2014, were collected from an online re-source of the United States Geological Survey. The dataset includes chlorophyll, specific conductance, dissolved oxygen, and turbidity which affect and influence the quality of water. A feed-forward Neural Network with NAR time series mod-el had been used with the training algorithm of Scaled Conjugate Gradient and the activation function of Log Sigmoid. The performance evaluation of the ANN based predictive model were calculated using Regression, Mean Squared Error and Root Mean Squared Error. The ANN-NAR proposed model proves that the prediction accuracy indicating much improved values as compared to other algorithms.

Chapter 3: Methodology

3.1 Introduction to System Design

Design is the abstraction of a solution .It is general description of the solution to problem without the details. Design is a view pattern seen in the analysis phase to be a pattern in a design phase. After the design phase we can reduce the time required to create the implementation.

The design of the system is the most critical factor affecting the quality of the application .The system design aims to identify the modules that should be in the system, the specification of these module and how with each other to produce the desired result.

For a system like our needs some kind of dataset that includes multiple classes. Thus, to have proper classification, we will collect data from driffent sources like Department of water supply &sewerage management, ENFOS etc. and try to add more by doing field visits if needed. We need to go through some data preprocessing steps in case of noisy and messy data.

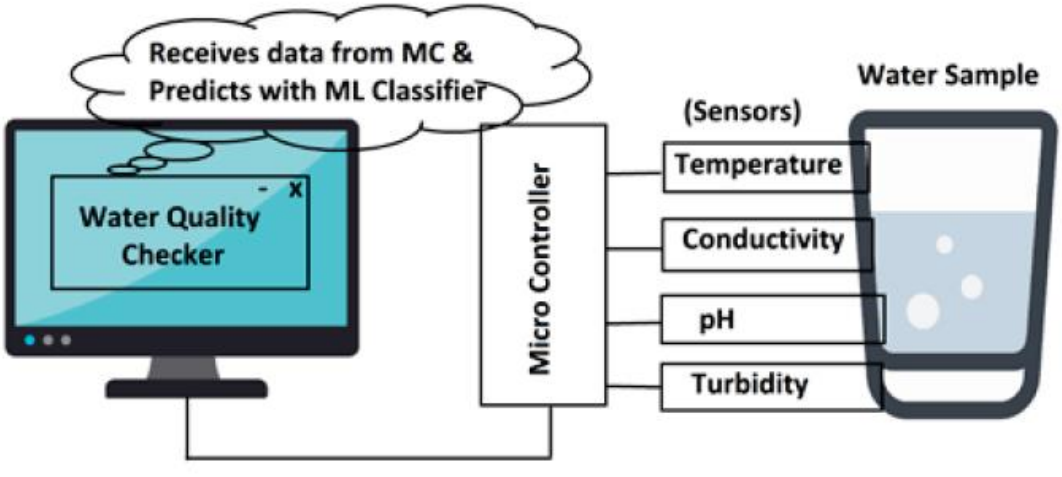


Figure 3.1 Block diagram of purposed system

The system will able to perform fallowing task:

- The collected data will be processed and augmented
- Training and validation of data
- Test set of data
- Predict water condition with driffent parameter
- Suggest their Remedies

3.1.1 Purposed System Circuit Diagram

Fig.3.1.1 shows the schematic circuit diagram of the hardware set-up of the proposed IWQM system. Except the temperature sensor, other three sensors are of analog type. Each sensor has three different color wires such as red, black and others. Here, red wires are for +5V power supply, black wires are for ground and others are used for data estimation. A breadboard is used for creating common points for ground and power supply separately. Then common node of ground is connected to the ground of ESP-32 and same process is repeated for power supply. The analog sensors are connected to the analog pins and digital sensor is connected to digital pin of the controller.

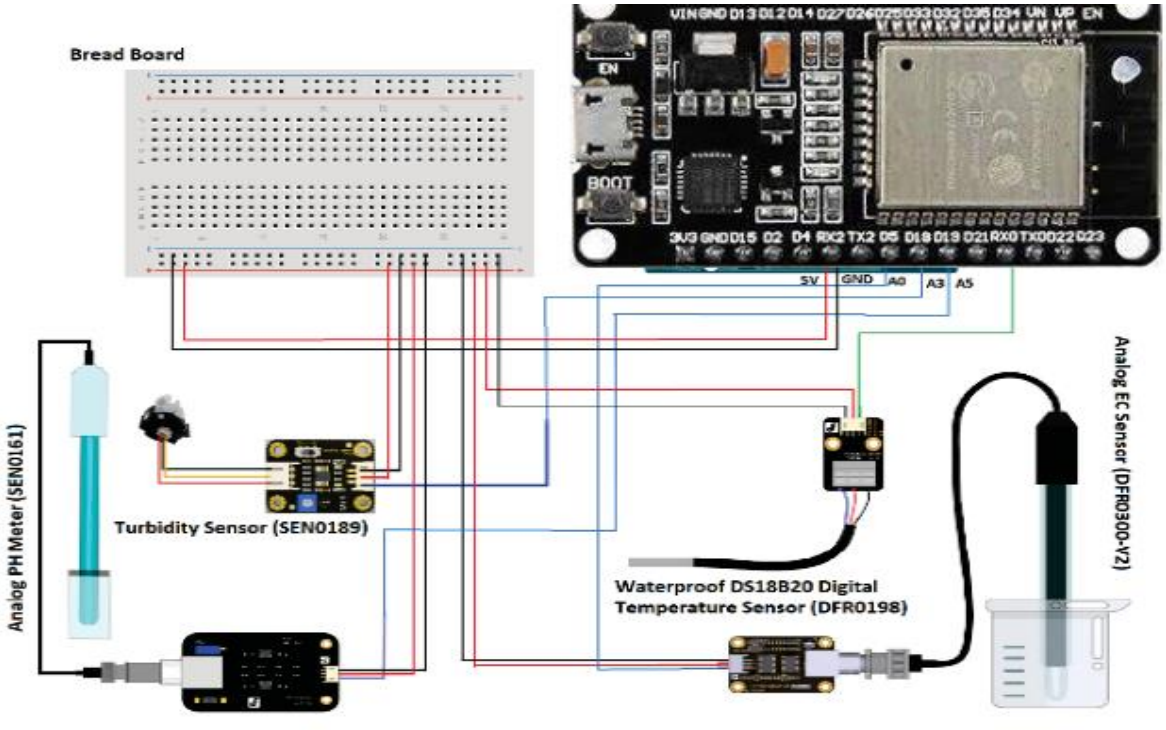


Figure 3.1.1 Purposed Circuit diagram

3.1.2 Data Modeling and analysis

Machine learning required a large amount of historical data. Data collection has a sufficient amount of historical and raw data. Raw data cannot be used directly prior to data pre-processing. It is then used to preprocess what kind of algorithm with the model. Training and testing this model to ensure that it predicts correctly and with minimal errors. A tuned model involves tuning from time to time to improve accuracy.

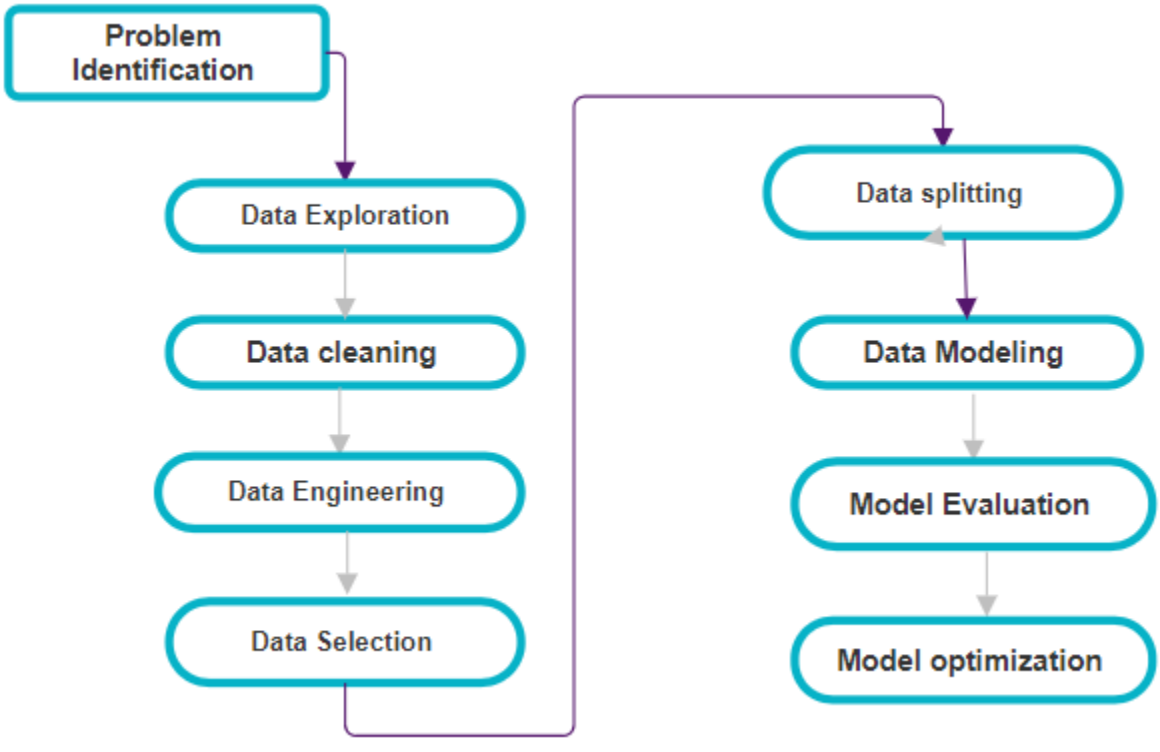


Fig 3.1.2: Process Flow of the Model

There are basically fallowing steps for making our model predict the water quality of the water samples. Those steps are:-

A. Problem Identification

In this step, we identify the problem which is solved by our model. So the problem to be solved by our model is water quality prediction using a dataset.

B. Data Extraction:-

In this, we extract the data from the internet to train our data and predict the water quality. So for that, we take the Department of water supply and Sewerage Management dataset which contains almost 2200 instances of different places which are collected between up to 2023.

C. Data Exploration:-

In this step, we analyze the data visually by comparing some parameters of water with the standards of water provided by NWAS. It gives a slight overview of the data.

D. Data Cleaning

In this step, we clean that data like if there are some missing values in it so we replace them with mean and remove noise from the data.

E. Data Engineering

In this step, we ensure that the data is quality data so that the prediction accuracy increases.

F. Data Selection

In this step, we select the data types and source of the data. The essential goal of data selection is deciding fitting data type, source, and instrument that permit agents to respond to explore questions sufficiently

G. Data Splitting

In this step, we divide the dataset into smaller subsets for easing the complexity. Normally, with a two-section split, one section is utilized to assess or test the information and the other to prepare the model.

H. Data Modeling

In this step, we create a graph of the dataset for visual representation of data for better understanding. A Data Model is this theoretical model that permits the further structure of conceptual models and to set connections between data.

3.2 Purposed Classification models for Water quality prediction

Water quality prediction is a critical aspect of environmental monitoring and management. Traditional methods often face challenges in providing real-time insights. This literature review explores recent advancements in using machine learning (ML) techniques for water quality prediction, focusing on key methodologies, models, and applications.

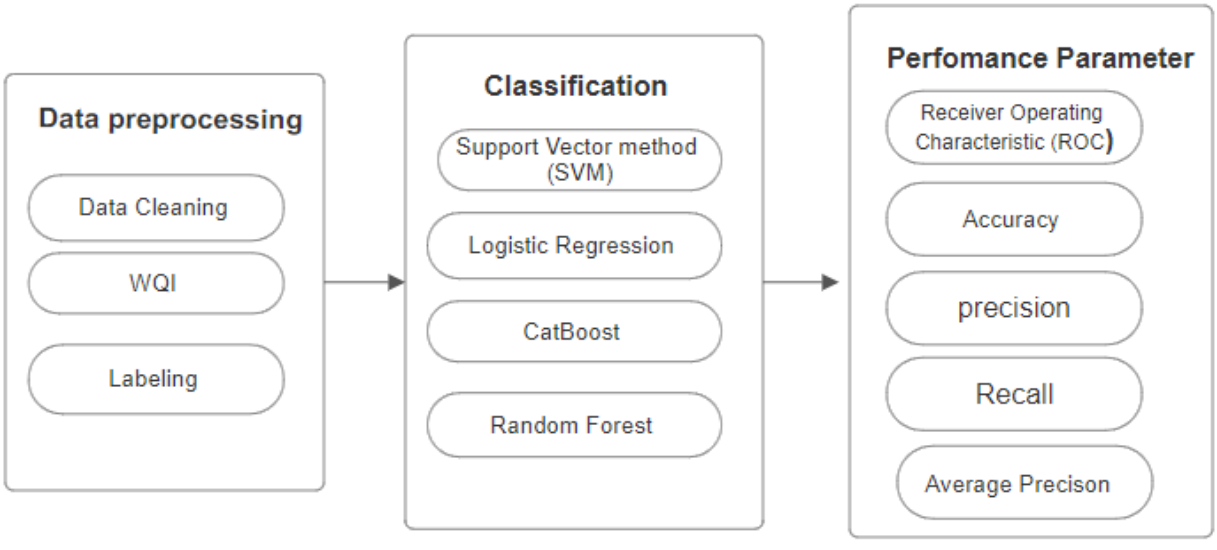


Fig 3.2: Methodology of proposal system

:.

3.4 Hardware and software

3.4.1 ESP 32 Microcontroller

ESP32 is a series of low-cost, low-power system on a chip microcontroller with integrated Wi-Fi and dual-mode Bluetooth.

- The ESP32 is dual core, this means it has 2 processors.
- It has Wi-Fi and Bluetooth built-in.
- The clock frequency can go up to 240MHz and it has a 512 kB RAM.
- This particular board has 30 or 36 pins, 15 in each row

Figure 3.4.1 ESP 32

3.4.2PH Sensor

A pH sensor is a device used to measure the acidity or alkalinity of a solution. PH is a measure of the hydrogen ion concentration in a solution and is typically expressed on a scale from 0 to 14, with 7 being neutral. A pH below 7 indicates acidity, while a pH above 7 indicates alkalinity.

The sensor generates a voltage proportional to the hydrogen ion concentration, and this voltage is then converted into a pH value.

Fig 3.4.2 PH Sensor

3.4.3 Tempeture sensor

Waterproof temperature sensors are designed to operate in wet or submerged environments without being damaged by water. These sensors are commonly used in applications where accurate temperature measurements are required in conditions where water exposure is a concern.

Fig 3.4.3 DS18B20 Water Proof Sensor

3.4.3 Turbidity Sensor

Turbidity is a measure of water quality that reflects the amount of suspended particles in a water sample by observing the amount of light scattered through it**.**Water with high turbidity often requires purification processes before it can be used in industrial and domestic applications. This is because a decrease in turbidity often implies a reduction in harmful substances, bacteria, and viruses in the water.

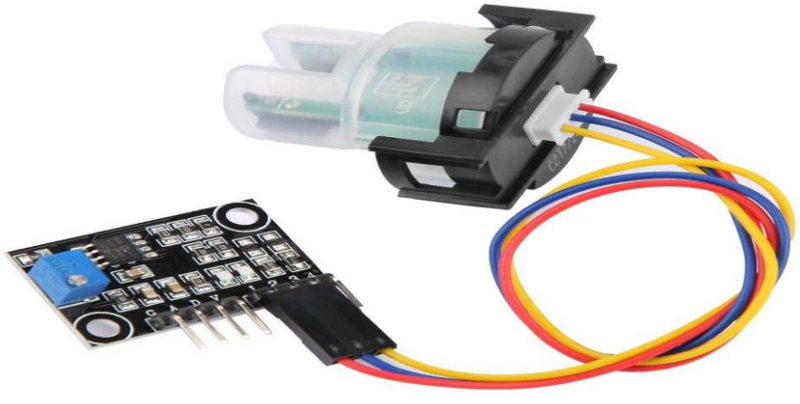
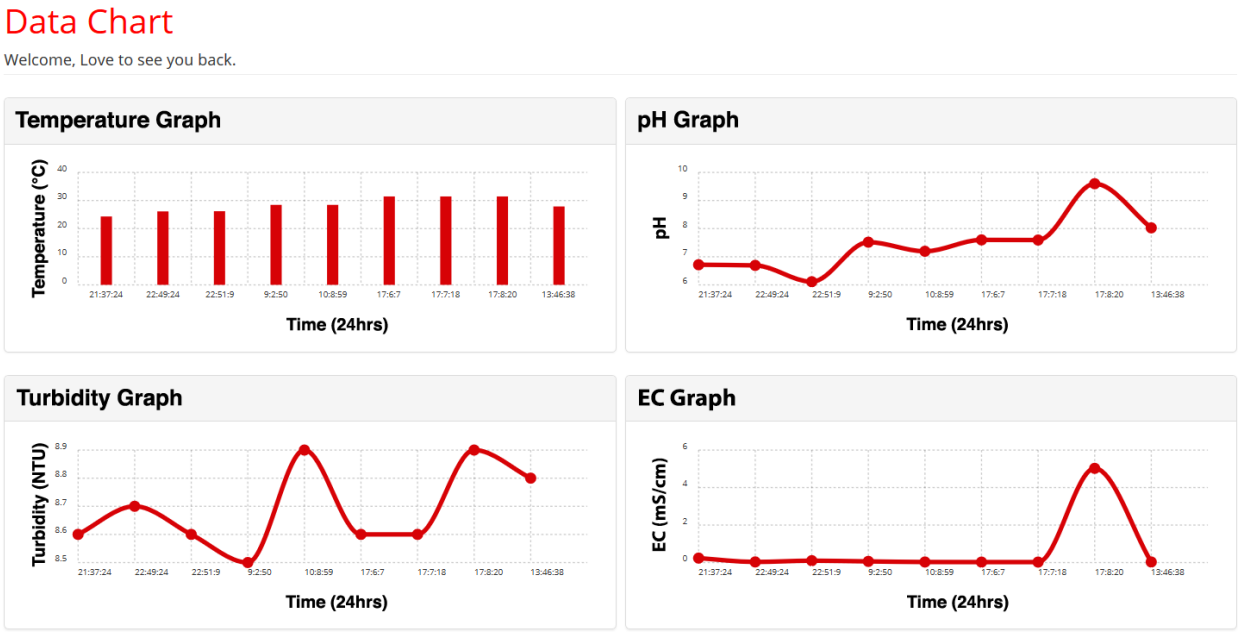


Fig 3.4.3 Turbidity Sensor

3.4.4 Thing Speak

Thing Speak enables sensors, instruments, and websites to send data to the cloud where it is stored in either a private or a public channel. Thing Speak stores data in private channels by default, but public channels can be used to share data with others.

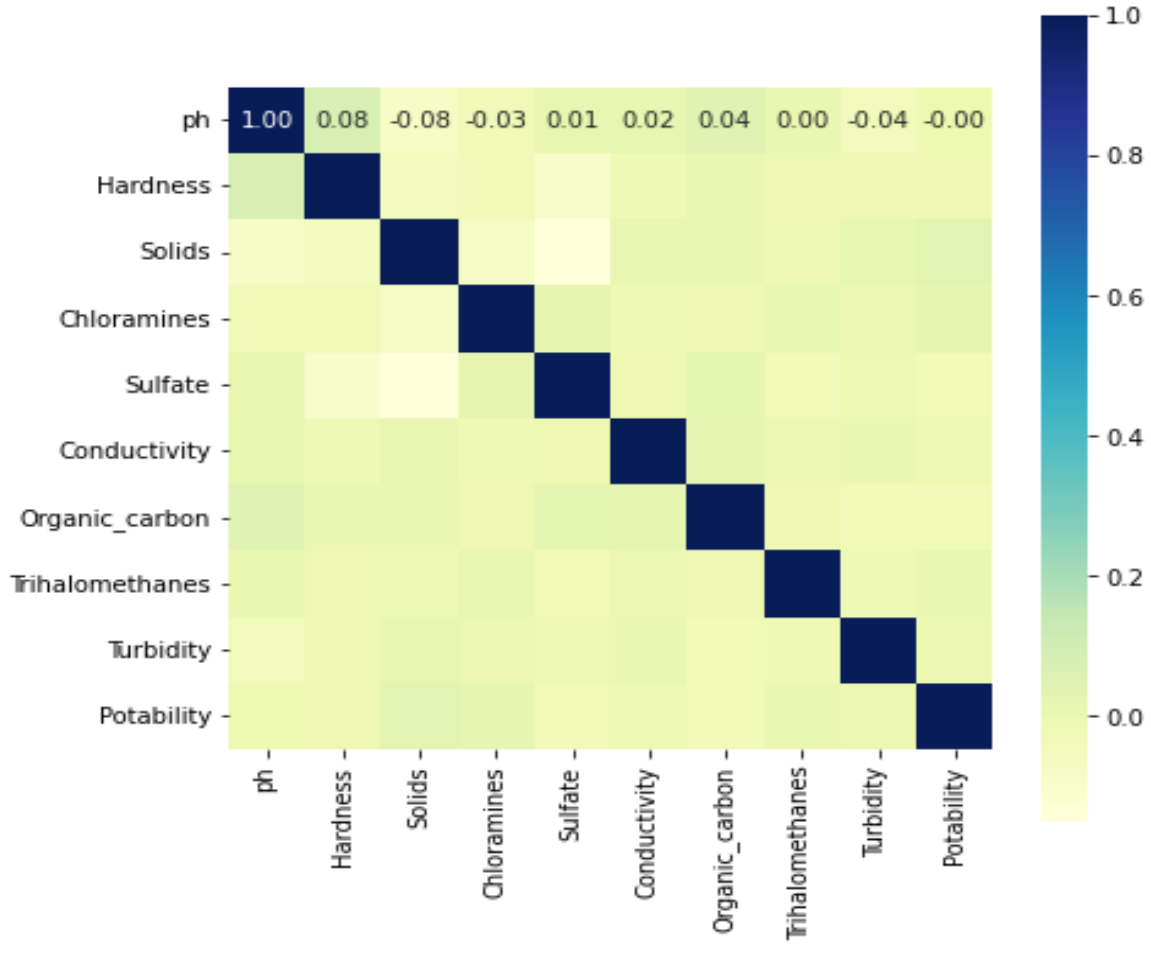


3.4.5 Web base application /Mobile app

Figure 3.4.5: prediction of water quality system

Chapter4 Result& Discussion

Results of the proposed system classifiers:



Correlations between the water parameter

4.1 Findings Description

4.1.1 ANN

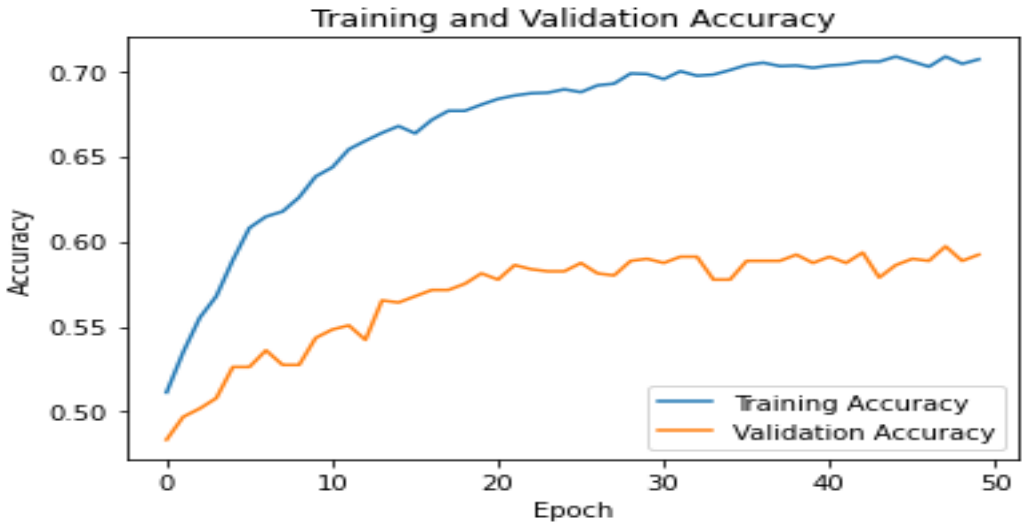


Figure4-1: Training and validation accuracy graph of ANN

Figure4-2: Training and validation loss graph of ANN

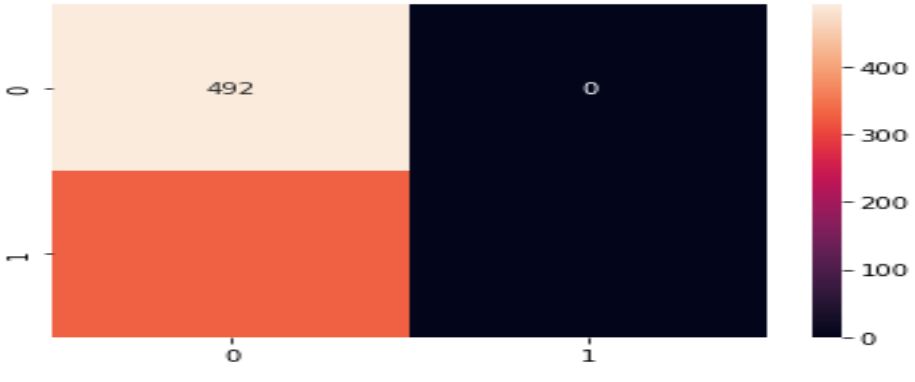


Fig4-3: Confusion matrixes of ANN

Table 4-1 ANN Performance Report

Water portability	Precision	Recall	F1-score
0	0.60	1.00	0.75
1	0.00	0.00	0.00
Accuracy			**0.98**
macro avg	0.30	0.50	0.38
weighted avg	0.36	0.60	0.45

4.1.2 Logistic Regression

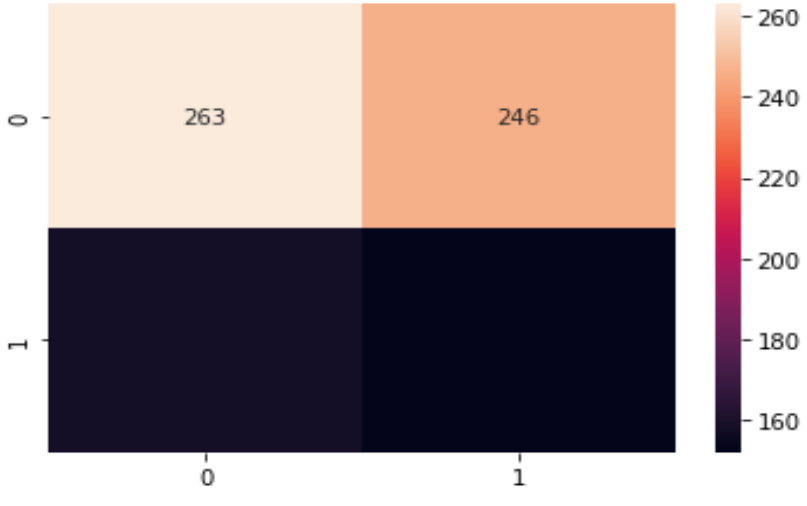


Fig4-3 confusion matrix of logistic regression

Table 4-2 Performance Report of Logistic Regression

Water portability	Precision	Recall	F1-score
0	0.57	0.51	0.54
1	0.40	0.46	0.43
Accuracy			**0.98**
macro avg	0.49	0.48	0.38
weighted avg	0.50	0.49	0.49

4.1.3 CATBOOST

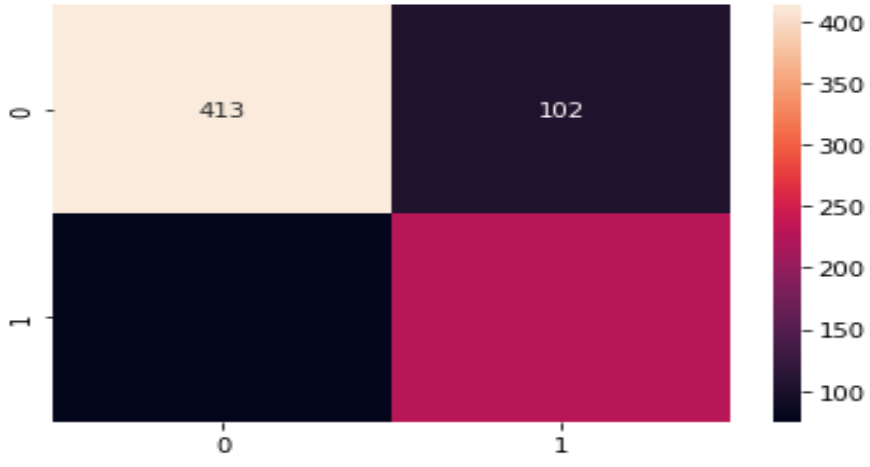


Figure4-4: confusion matrix of CATBOOST

Table 4 -3 Performance Report of CATBOOST

Water portability	**Precision**	**Recall**	**F1-score**
0	0.85	0.80	0.82
1	0.69	0.75	0.72
Accuracy			**0.78**
macro avg	0.77	0.78	0.77
weighted avg	0.79	0.78	0.79

4.1.4 SVM

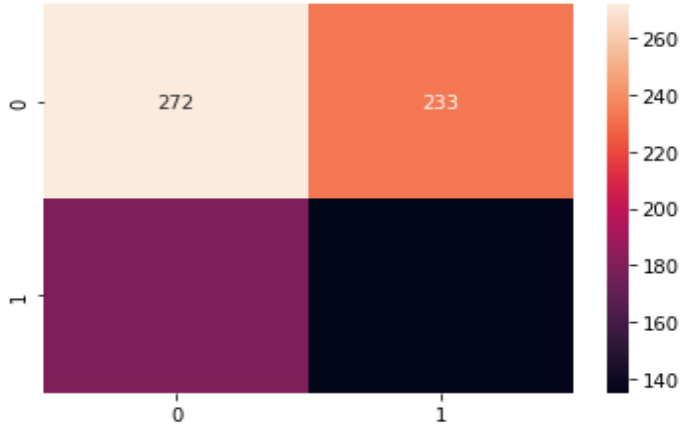


Figure4-5 confusion matrix of SVM

Table 4 -5 Performance Parameter of SVM

Water portability	Precision	Recall	F1-score
0	0.06	0.54	0.57
1	0.37	0.43	0.40
Accuracy			**0.50**
macro avg	0.48	0.48	0.48
weighted avg	0.51	0.50	0.50

4.1.5 XGBOOST

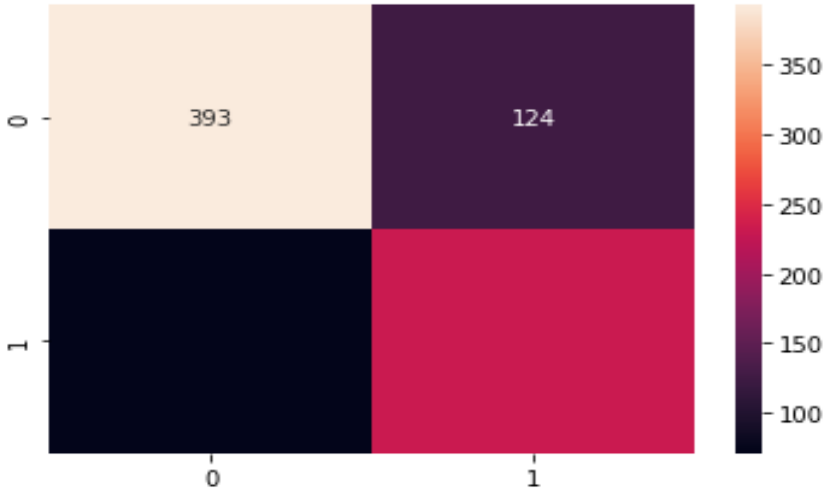


Figure4-6 Confusion matrix of XGBOOST

Table 4 -6 Performance of XGBOOST

Water portability	Precision	Recall	F1-score
0	0.85	0.76	0.80
1	0.65	0..76	0.70
Accuracy			**0.76**
macro avg	0.75	0.48	0.48
weighted avg	0.77	0.76	0.77

4.2 Comparison of different model

Parameters	ANN	Logistic regression	SVM	XGBoost	CATBoost
Precision	0.60	0.40	0.37	0.65	0.69
Recall	1.00	0.46	0.43	0.76	0.75
F1-score	0..986	0.43	0.40	0.80	0.086
Accuracy	0.984	0.98	0.50	0.76	0.78

Chapter 5 Conclusion

The evaluation of five different models used for water classification reveals varying levels of performance. Logistic regression demonstrate robust results, showcasing high precision, recall, F1-score, and accuracy. These models appear to be the most reliable for this classification task

The data has been split into 80 % for training and 20 % for testing. illustrates the 4 × 4 confusion matrix for each classifier with their color-coded values. All the performance metrics have been calculated using confusion matrices of each classifier, as mentioned. According to the estimates, loges tic regression and RF performed better than other.

This paper presented a practical and economical solution to monitor the quality of water especially in rural areas without any human intervention. To solve this problem this paper presented various contemporary technologies such as IoT, cloud computing and Machine learning. On combining these technologies we are able to solve one of the basic and emerging problem of human survival to certain extent.

So, in this paper, we propose an alternative approach using artificial intelligence to predict water quality. This method uses a significant and easily available water quality index which is set by the NWAS Nepal. The data taken from "Department of water supply and sewerage management"& ENFOS, National Academy of Science and Technology (NAAST), Kathmandu University which includes around 5000 above sample.

REFERENCES

[1] Pradeepkumar M, Monisha J. ”The Real Time Monitoring of Water Quality in IoT Environment” 2016 International Journal of Innovative Research in Science, Engineering and Technology, 2015 ISSN(Online): 2319-8753

[2] AtifAlamri, WasaiShadab Ansari, Mohammad Mehedi Hassan, M. ShamimHossain, AbdulhameedAlelaiwi, and M.AnwarHossain, ”A Survey on Sensor-Cloud: Architecture, Applications, and Approaches”, International Journal of Distributed Sensor Networks, Volume 2013

[3] Kedia, Nikhil. ”Water Quality Monitoring for Rural Areas- a Sensor Cloud Based Economical Project” 2015 1st International Conference on Next Generation Computing Technologies (NGCT), 2015, doi:10.1109/ngct.2015.7375081.

[4] Vijayakumar, N, and R Ramya. ”The Real Time Monitoring of Water Quality in IoT Environment” 2015 International Conference on Circuits, Power and Computing Technologies [ICCPCT-2015], 2015, doi:10.1109/iccpct.2015.7159459.

[5] R.Karthik Kumar, M.Chandra Mohan, S.Vengateshapandiyan, M.Mathan Kumar, R.Eswaran. ”Solar based advanced water quality monitoring system using wireless sensor network” 2014, International Journal of Science, Engineering and Technology Research, 2014

[6] Fiona Regan, Antoin, McCarthy. ”Smart Coast Projectˆa Smart Water Quality Monitoring Systemˆa 2006, Marine Institute/Environmental Protection Agency Partnership, 2006

[7] Vaishnavi V. Daigavane, Dr. M.A Gaikwad. ”Water Quality Monitoring System Based on IOT” 2017 Advances in Wireless and Mobile Communications, Nov 2017 ISSN 0973-6972

[8] Pradeepkumar M, Monisha J. ”The Real Time Monitoring of Water Quality in IoT Environment” 2016 International Journal of Innovative Research in Science, Engineering and Technology, 2015 ISSN(Online) : 2319-8753