NALAIYA THIRAN

IBM

PROJECT REPORT

Efficient Water Quality Analysis & Prediction using Machine Learning

By team – PNT2022TMID52665

Presented by

DEEPTHY SHIVANI A	CITC1905074
DIVYA K	CITC1905079
SOUNDARYA M	CITC1905115
SUBIKSHA DEVI M S	CITC1905116

INDEX

INTRODUCTION

1.1 PROJECT OVERVIEW:

Water is the most important of sources, vital for sustaining all kinds of life; however, it is in constant threat of pollution by life itself. Water is one of the most communicable mediums with a far reach. Rapid industrialization has consequently led to deterioration of water quality at an alarming rate. Poor water quality results have been known to be one of the major factors of escalation of harrowing diseases. As reported, in developing countries, 80% of the diseases are water borne diseases, which have led to 5 million deaths and 2.5 billion illness. In truth, the repercussions of contaminated drinking water are quite harmful, posing a serious condition to human health, the environment, and infrastructure. According to a United Nations (UN) report, 1.5 million people expire each year as a result of illnesses that occurred in contaminated water. Water contamination is said to be the cause of 80% of health problems in impoverished countries. Every year, there are 5 million fatalities and 2.5 billion illnesses reported.

The most common of these diseases are diarrhea, typhoid, gastroenteritis, cryptosporidium infections, some forms of hepatitis and giardiasis intestinal worms. Water borne diseases cause a GDP loss of 0.6–1.44% every year. Water quality is currently estimated through expensive and time-consuming lab and statistical analyses, which require sample collection, transport to labs, and a considerable amount of time and Water, which is quite ineffective given water is quite a communicable medium and time is of the essence if water is polluted with disease-inducing waste. The horrific consequences of water pollution necessitate a quicker and cheaper alternative. In this regard, the main motivation in this study is to propose and evaluate an alternative method based on supervised machine learning for the efficient prediction of water quality in real-time. A representative set of supervised machine learning algorithms were employed on the said dataset for predicting the water quality index (WQI). The main contributions of this study are summarized as follows: A first analysis was conducted on the available data to clean, normalize and perform feature selection on the water quality measures, and therefore, to obtain the minimum relevant subset that allows high precision with low cost. A series of representative supervised prediction (classification and regression) algorithms were tested on the dataset worked here. Using different ANN models, several types of research have been conducted to simulate and predict water quality. The probability and usefulness of using

ANN applications to forecast the quality of drinking water have been confirmed in these investigations. Measurements of different parameters like chemical oxygen demand (COD), dissolved oxygen (DO), electrical conductivity (EC), biochemical oxygen demand (BOD), temperature, pH, K, Na, Mg, and other water quality components have been proposed. Indeed, the good quality of water resources significantly reduces the cost of water treatment for drinking and industrial purposes and improves agricultural production. However, monitoring all parameters involved in a river or groundwater is often insufficient in developing countries as it is laborious and expensive. However, in recent years, many water quality indexes (WQI) have been proposed and developed by several national and international organisations. Although there are many machine learning models developed for water quality prediction, there is little to no research that integrates the models with a graphical user interface (GUI). The user is able to input real-time data into a GUI that has been connected to a machine learning model, thus minimising time consumption by reducing the number of procedures required to conduct the analysis. The GUI-based app in this study also enables a simpler and more trouble-free workflow in predicting water quality parameters and can be accessible to more people, especially on-site operators and trainees.

1.2 PURPOSE:

Monitoring water quality in the 21st century is a growing challenge because of the large number of chemicals used in our everyday lives and in commerce that can make their way into our waters. Methods of chemical analysis and knowledge of chemical toxicity are available for only a few thousand of the more than 80,000 chemical compounds estimated by EPA to be in commercial use in the United States. Monitoring provides the objective evidence necessary to make sound decisions on managing water quality today and in the future. Water-quality monitoring is used to alert us to current, ongoing, and emerging problems; to determine compliance with drinking water standards, and to protect other beneficial uses of water.

2. LITERATURE SURVEY

Prediction of groundwater quality using efficient machine learning technique

AUTHORS: Sudhakar Singha, Srinivas Pasupuleti, Soumya S.Singha, Rambabu Singh, Suresh Kumar

DESCRIPTION:

To ensure safe drinking water sources in the future, it is imperative to understand the quality and pollution level of existing groundwater. The prediction of water quality with high accuracy is the key to control water pollution and the improvement of water management. In this study, a deep learning (DL) based model is proposed for predicting groundwater quality and compared with three other machine learning (ML) models, namely, random forest (RF), eXtreme gradient boosting (XGBoost), and artificial neural network (ANN). A total of 226 groundwater samples are collected from an agriculturally intensive area Arang of Raipur district, Chhattisgarh, India, and various physicochemical parameters are measured to compute entropy weight-based groundwater quality index (EWQI). Prediction performances of models are determined by introducing five error metrics. Results showed that DL model is the best prediction model with the highest accuracy in terms of R2, i.e., R2 = 0996 against the RF (R2 = 0.886), XGBoost (R2 = 0.0.927), and ANN (R2 = 0.917). The uncertainty of the DL model output is cross-verified by running the proposed algorithm with a newly randomised dataset for ten times, where minor deviations in the mean value of performance metrics are observed. Moreover, input variable importance computed by prediction models highlights that DL model is the most realistic and accurate approach in the prediction of groundwater quality.

ADVANTAGE:

Relatively higher prediction performance was observed in XGBoost model than RF and ANN models.

LIMITATION:

A single monsoon dataset is considered in the prediction models is the limitation.

Efficient Prediction of Water Quality Index (WQI) Using Machine Learning Algorithms

AUTHORS: Md.Mahedi Hassan, Laboni Akter, Mushfiqur Rahman, Nusrat Jahan

DESCRIPTION:

The quality of water has a direct influence on both human health and the environment. Water is utilised for a variety of purposes, including drinking, agriculture, and industrial use. The water quality index (WQI) is a critical indication for proper water management. The purpose of this work was to use machine learning techniques such as RF, NN, MLR, SVM, and BTM to categorise a dataset of water quality in various places across India. Water quality is dictated by features such as dissolved oxygen (DO), total coliform (TC), biological oxygen demand (BOD), Nitrate, pH, and electric conductivity (EC). These features are handled in five steps: data pre-processing using min-max normalisation and missing data management using RF, feature correlation, applied machine learning classification, and model's feature importance. The highest accuracy Kappa, Accuracy Lower, and Accuracy Upper findings in this research are 99.83, 99.17, 99.07, and 99.99, respectively. The finding showed that Nitrate, PH, conductivity, DO, TC, and BOD are the key qualities that contribute to the orderly classification of water quality, with Variable Importance values of 74.78, 36.805, 81.494, 105.770, 105.166, and 130.173, respectively.

ADVANTAGE:

Multinomial Logistic Regression has the highest accuracy at 99.83% Random forest has accuracy at 99.63%.

LIMITATION:

Only 7 parameters were considered to predict water quality.

Machine learning based marine water quality prediction for coastal hydro-environment management

AUTHORS: Tianan Deng, Kwok-Wing Chau, Huan-Feng Duan

DESCRIPTION:

During the past three decades, harmful algal blooms (HAB) events have been frequently observed in marine waters around many coastal cities in the world including Hong Kong. The increasing occurrence of HAB has caused acute influences and damages on water environment and marine aquaculture with millions of monetary losses. For example, the Tolo Harbour is one of the most affected areas in Hong Kong, where more than 30% HAB occurred. In order to forewarn the potential HAB incidents, the machine learning (ML) methods have been increasingly resorted to modelling and forecasting water quality issues. In this study, two different ML methods - artificial neural networks (ANN) and support vector machine (SVM) - are implemented and improved by introducing different hybrid learning algorithms for the simulations and comparative analysis of more than 30-year measured data, so as to accurately forecast algal growth and eutrophication in Tolo Harbour in Hong Kong. The application results show the good applicability and accuracy of these two ML methods for the predictions of both trend and magnitude of the algal growth. Specifically, the results reveal that ANN is preferable to achieve satisfactory results with quick response, while the SVM is suitable to accurately identify the optimal model but taking longer training time. Moreover, it is demonstrated that the used ML methods could ensure robustness to learn complicated relationships between algal dynamics and different coastal environmental variables and thereby to identify significant variables accurately. The results analysis and discussion of this study also indicate the potentials and advantages of the applied ML models to provide useful information and implications for understanding the mechanism and process of HAB outbreak and evolution that is helpful to improving the water quality prediction for coastal hydro-environment management.

ADVANTAGE:

ANN showed good predicting performances and there is no overfitting problem. SVM performance is better than all ANN models in terms of water quality prediction results.

LIMITATION:

It takes much longer to train the SVM than the ANN as the SVM takes a quadratic programming with time complexity of O(n3).

Water quality prediction using machine learning and flask

AUTHORS: S.Sharath, R.Harish, V.Aishwarya and Dr.M.Preetha

DESCRIPTION:

Generally, Water pollution refers to the release of pollutants into the water that is detrimental to human health and the planet as whole. The aim is to investigate machine learning-based techniques for water quality forecasting by predicting results with the best accuracy. The analysis of the data set by supervised machine Learning techniques (SMLT) to capture information like variable identification, uni-variate analysis, bi-variate and multivariate analysis, missing value treatments and analysis data validation, data cleaning/preparation, and data visualisation will be done on the entire given data set. Our analysis provides a comprehensive guide to sensitivity analysis of model parameters with regard to performance in the prediction of water quality pollution by accuracy calculation. To propose a machine learning-based method to accurately predict the Water Quality Index value by prediction results in the form of best accuracy from comparing supervised classification machine learning algorithms. . Additionally, to compare and discuss the performance of various machine learning algorithms from the given transport traffic department data set with evaluation classification report, identify the confusion matrix and categorising data from priority and the result shows that the effectiveness of the proposed machine learning algorithm technique can be compared with the best accuracy with precision, Recall and F1 Score.

ADVANTAGE:

Produces the appropriate results with the mentioned techniques.

LIMITATION:

Accuracy of the technique is not mentioned.

A novel machine learning application: Water quality resilience prediction Model

AUTHORS: **Maryam Imani, Md Mahmudul Hasan, Luiz Fernando Bittencourt, Kent McClymont and Zoran Kapelan**

DESCRIPTION:

Resilience-informed water quality management embraces the growing environmental challenges and provides greater accuracy by unpacking the systems' characteristics in response to failure conditions in order to identify more effective opportunities for intervention. Assessing the resilience of water quality requires complex analysis of influential parameters which can be challenging, time consuming and costly to compute. It may also require building detailed conceptual and/or physically process-based models that are difficult to build, calibrate and validate. This study utilises Artificial Neural Network (ANN) to develop a novel application to predict water quality resilience to simplify resilience evaluation. The Fuzzy Analytic Hierarchy Process method is used to rank water basins based on their level of resilience and to identify the ones that demand prompt restoration strategies. The commonly used 'magnitude * duration of being in failure state' quantification method has been used to formulate and evaluate resilience. A 17-years long water quality dataset from the 22 water basins in the State of São Paulo, Brazil, was used to train and test the ANN model. The overall agreement between the measured and simulated WQI resilience values is satisfactory and hence, can be used by planners and decision makers for improved water management. Moreover, comparative analyses show similarities and differences between the 'level of criticalities' reported in each zone by Environment Agency of the state of São Paulo (CETESB) and by the resilience model in this study.

ADVANTAGE:

Resilience is a fast-growing concept proved to be an effective approach in preparing engineering systems to tackle and cope with emerging challenges.

LIMITATION:

Model performance in validation dataset decreases abruptly in successive iterations.

Water Quality Prediction Method Based on LSTM-BP

Authors: Huimin Jia, Xiaofeng Zhou

DESCRIPTION:

Water quality prediction is of practical significance not only for the planning, evaluation, and management of the water environment, but also for the prevention and control of water pollution. In order to improve the accuracy of water quality prediction, an LSTM-BP combined model algorithm based on Long Short-term Memory Neural Network (LSTM NN) and BP neural network is proposed. Taking the water temperature data of No.6 large-scale integrated observation buoy on the Yangtze estuary as an example, a time series prediction model framework is established, and the data processing to model simulation is completed with the help of Python to realise the water quality prediction based on LAST MBP. The method is compared with the LSTM model and BP model, the experimental results show that the time series predicted by LSTM-BP is more accurate. This LSTM-BP model can be effectively applied to the prediction of water quality indicators and the early warning and prediction system of water quality trends.

ADVANTAGE:

Integrated LSTM-BP has better prediction performance of time-series than the single LSTM and BP.

LIMITATIONS

The Back propagation method's convergence speed is slow

Implementation of ML models for monitoring and predicting water quality parameters

AUTHORS: Gasim Hayder, Osman Kurniawan

DESCRIPTION:

Since clean water is well known as one of the crucial sources that all living things need in their daily lives, the demand for clean freshwater nowadays has increased. However, water quality is slowly deteriorating due to anthropogenic and natural sources of pollution and contamination. Therefore, this study aims to develop artificial neural network (ANN) models to predict six different water quality parameters in the Langat River, Malaysia. Moreover, an application (app) equipped with a graphical user interface (GUI) was designed and developed to conduct real-time prediction of the water quality parameters by using real-time data as inputs together with the ANN models. As for the results, all of the ANN models achieved high coefficients of determination (R2), which were between 0.9906 and 0.9998, as well as between 0.8797 and 0.9972 for training and testing datasets, respectively. The developed app successfully predicted the outcome based on the run models. The implementation of a GUI-based app in this study enables a simpler and more trouble-free workflow in predicting water quality parameters. By eliminating sophisticated programming subroutines, the prediction process becomes accessible to more people, especially on-site operators and trainees.

ADVANTAGE:

As only one method is used, the results were accurate.

LIMITATION:

This method takes only 6 water quality parameters. Other parameters are not considered.

A water quality prediction method based on the multi-time scale bidirectional LSTM network

AUTHORS: Qinghong Zou, Qingyu Xiong, Hualing Yi

DESCRIPTION:

As an important factor affecting the mangrove wetland ecosystem, water quality has become the focus of attention in recent years. Therefore, many studies have focused on the prediction of water quality to help establish a regulatory framework for the assessment and management of water pollution and ecosystem health. To make a more accurate and comprehensive forecast analysis of water quality, we propose a method for water quality prediction based on the multi-timescale bidirectional LSTM network. In the method, we improve data integrity and data volume through data pre-processing. and the network processes input data forward and backward and considers the dependencies at multiple time scales. Besides, we use the Box–Behnken experimental design method to adjust hyper-parameters in the process of modelling. In this study, we apply this method to the water quality prediction research of Beilun Estuary, and the performance of our proposed model is evaluated and compared with other models. The experiment results show that this model has better performance in water quality prediction than that of using LSTM or bidirectional LSTM alone.

ADVANTAGE:

MT-BLSTM had the highest accuracy compared to LSTM or bidirectional LSTM.

LIMITATION:

The reusability of the model needs to be improved.

Prediction of irrigation water quality parameters using machine learning models in a semi-arid environment.

AUTHORS: Ali El Bilali, Abdeslam Taleb

DESCRIPTION:

Evaluation of the water suitability for irrigation purposes using conventional approaches is generally expensive because it requires several parameters, particularly in developing countries. Therefore, developing accurate and reliable models may be valuable to overcome this issue in the management of the water used in agriculture. To achieve this purpose, 8 Machine Learning (ML) models namely: Artificial Neural Network (ANN), Multiple Linear Regression (MLR), Decision Tree, Random Forest (RF), Support Vector Regression (SVR), k-Nearest Neighbour (kNN), Stochastic Gradient Descent (SGD) and Adaptive Boosting (AdaBoost) have been developed and validated for predicting of 10 Irrigation Water Quality (IWQ) parameters such as Sodium absorption ratio (SAR), adjusted SARa, Exchangeable Sodium Percentage (ESP), percentage of Sodium (%Na), Residual Sodium Carbonate (RSC), Permeability Index (PI), Kelly Ratio (KR), Chloride Cl^-, Magnesium Absorption Ratio (MAR), and TDS dissolved in water surface of Bouregreg watershed in Morocco using electrical conductivity (EC) and pH as input variables. 300 samples are analysed at 9 monitoring stations across four main rivers, processed and selected to train and validate the models. The results have revealed that, except for SVR and k-NN models and MAR and PI parameters, all other models are highly accurate in predicting the other parameters with coefficients of correlations (r) with ranges of [0.56, 0.99], and [0.64, 0.99] for training and validation processes sequentially.

ADVANTAGE:

Adaboost prediction accuracy is high for seven parameters compared to other models.

LIMITATION:

SVM and KNN models showed low performance for predicting contents in water. Different models showed different accuracy.

Designing Efficient and Sustainable Predictions of Water Quality Indexes at the Regional Scale Using Machine Learning Algorithms

AUTHORS: Abdessamed Derdour, Antonio Jodar-Abellan, Miguel Angel Pardo, Sherif S.M. Ghoneim, Enas E. Hussein.

DESCRIPTION:

In order to maintain the availability of resources for drinkable water and to monitor pollution, the prediction of water quality indexes is extremely important. Thus, planning and managing water resources can greatly benefit from precise groundwater level predictions. As a result, an effort is made in this work to create a forecasting model that is effective for predicting groundwater quality by using the water quality index (WQI) in the Wilaya of Naama, placed in the southwestern region of Algeria. Based on many characteristics and indexes, conventional approaches evaluate water suitability for drinking and domestic purposes. Although these techniques are reliable tools, they can be costly and time-consuming. Therefore, this study proposes an alternative machine learning method for predicting water quality using only a few simple water quality criteria. The data used to conduct the study were collected from 169 samples of groundwater from 12 municipalities in the Wilaya of Naâma. A set of representative supervised machine learning algorithms has been used to estimate the *WQI* indicator. Based on *WQI* results, four classes were fixed: excellent, good, poor, and very poor or unsafe water. A relevant percentage (62.7%) of the considered physicochemical parameters depicted good water quality results. Related to prediction tools, main results showed that Support Vector Machine (SVM) algorithms classify groundwater quality with high accuracy (95.4%) with standardized data and lower accuracy (88.88%) for raw data. Therefore, a great correlation between observed and predicted water quality data was obtained in the present manuscript. These results offer a useful performance assessment tool for decision-makers, and further investigation can be undertaken by integrating the findings of this research on a large scale in arid areas. In conclusion, the SVM model is a simple and effective empirical model to simulate water quality, and the method presented in this work is sufficiently general to be applied to a wide range of arid areas.

ADVANTAGE:

- It proposes an alternative machine learning method for predicting water quality using only a few simple water quality criteria.
- Support Vector Machine (SVM) algorithms classify groundwater quality with high accuracy (95.4%) with standardized data and lower accuracy (88.88%) for raw data.
- It has a great correlation between observed and predicted water quality data was obtained in the present manuscript.

LIMITATION:

Although these techniques are reliable tools, they can be costly and time-consuming.

2.1 EXISTING PROBLEM

The existing system uses labs by taking samples of water which can be costly and time consuming. When a user wants to get results quicker they can approach data analytics and get certain features from water and get prediction.It can also be done with the help of IoT with sensors.

2.2 REFERENCES

1. Water Quality Prediction Method Based on LSTM-BP Huimin Jia, Xiaofeng Zhou
2. A novel approach for water quality classification based on the integration of deep learning and feature extraction techniques Smail Dilmi , Mohamed Ladjal
3. Water Quality Prediction for Smart Aquaculture Using Hybrid Deep Learning Models
4. Efficient Prediction of Water Quality Index (WQI) Using Machine Learning Algorithms
5. Designing Efficient and Sustainable Predictions of Water Quality Indexes at the Regional Scale Using Machine Learning Algorithms
6. Water Quality Index Prediction for Improvement of Treatment Processes on Drinking Water Treatment Plant Goran Volf, Ivana Sušanj Čule, Elvis Žic and Sonja Zorko
7. Simple Prediction of an Ecosystem-SpecificWater Quality Index and the Water Quality Classification of a Highly Polluted River through Supervised Machine Learning Alberto Fernández del Castillo, Carlos Yebra-Montes, Marycarmen Verduzco Garibay, José de Anda, Alejandro Garcia-Gonzalez 4, and Misael Sebastián Gradilla-Hernández
8. Designing Efficient and Sustainable Predictions of Water Quality Indexes at the Regional Scale Using Machine Learning Algorithms Abdessamed Derdour, Antonio Jodar-Abellan, Miguel Ángel Pardo, Sherif S. M. Ghoneim and Enas E. Hussein
9. Prediction of the groundwater quality index through machine learning in Western Middle Cheliff plain in North Algeria Yamina Elmeddahi, Ragab Ragab
10. Robust machine learning algorithms for predicting coastal water quality index Md Galal Uddin, Stephen Nash, Mir Talas Mahammad Diganta, Azizur Rahman, Agnieszka I. Olbert

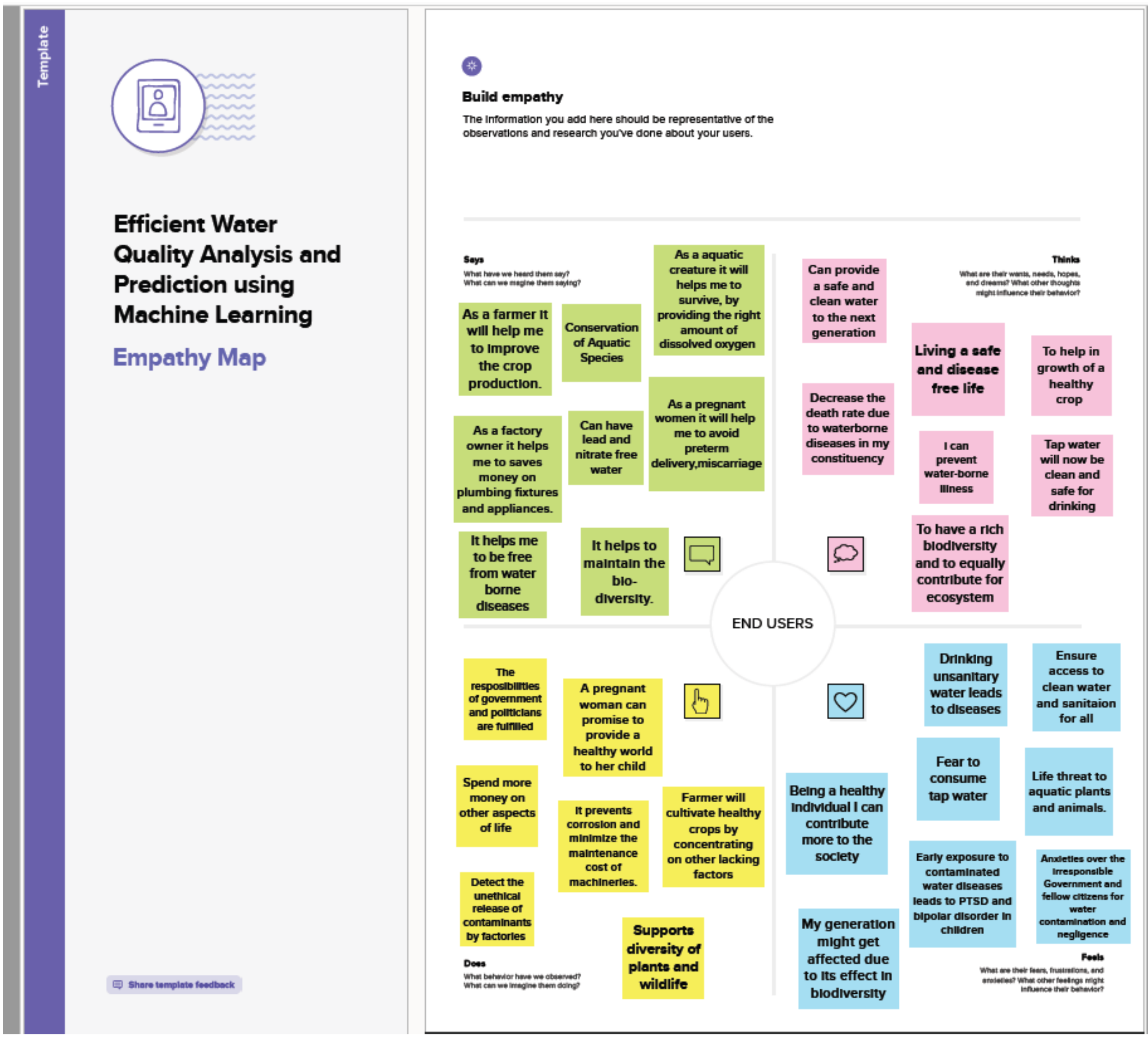
2.3 PROBLEM STATEMENT DEFINITION

To predict water quality in various regions with the help of certain features like temperature, conductivity, pH, coliform, biological oxygen demand.This will help in detecting whether the water quality is fit for human consumption or not.

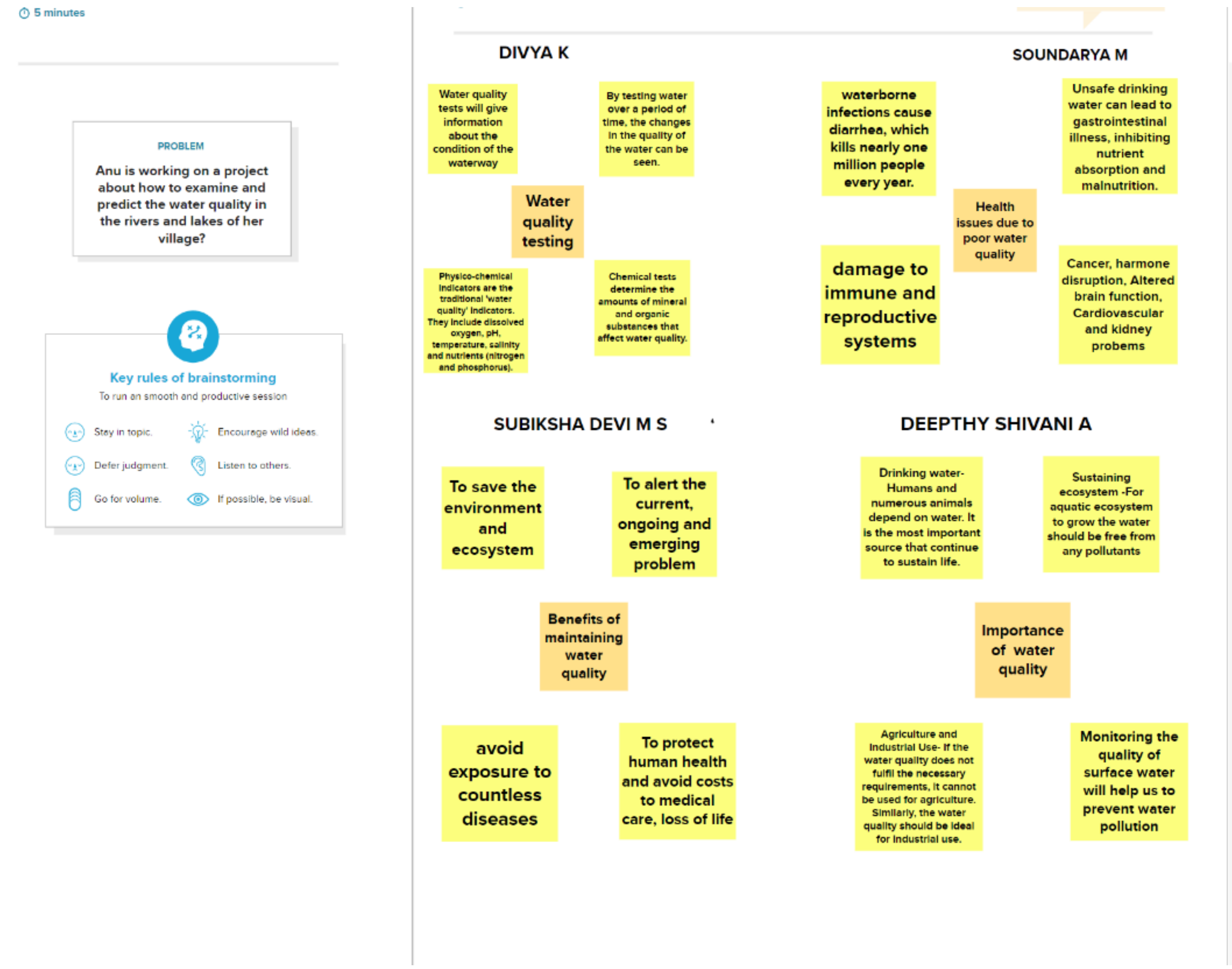
This is more important in our world today because of the increase in population and decrease in resources.

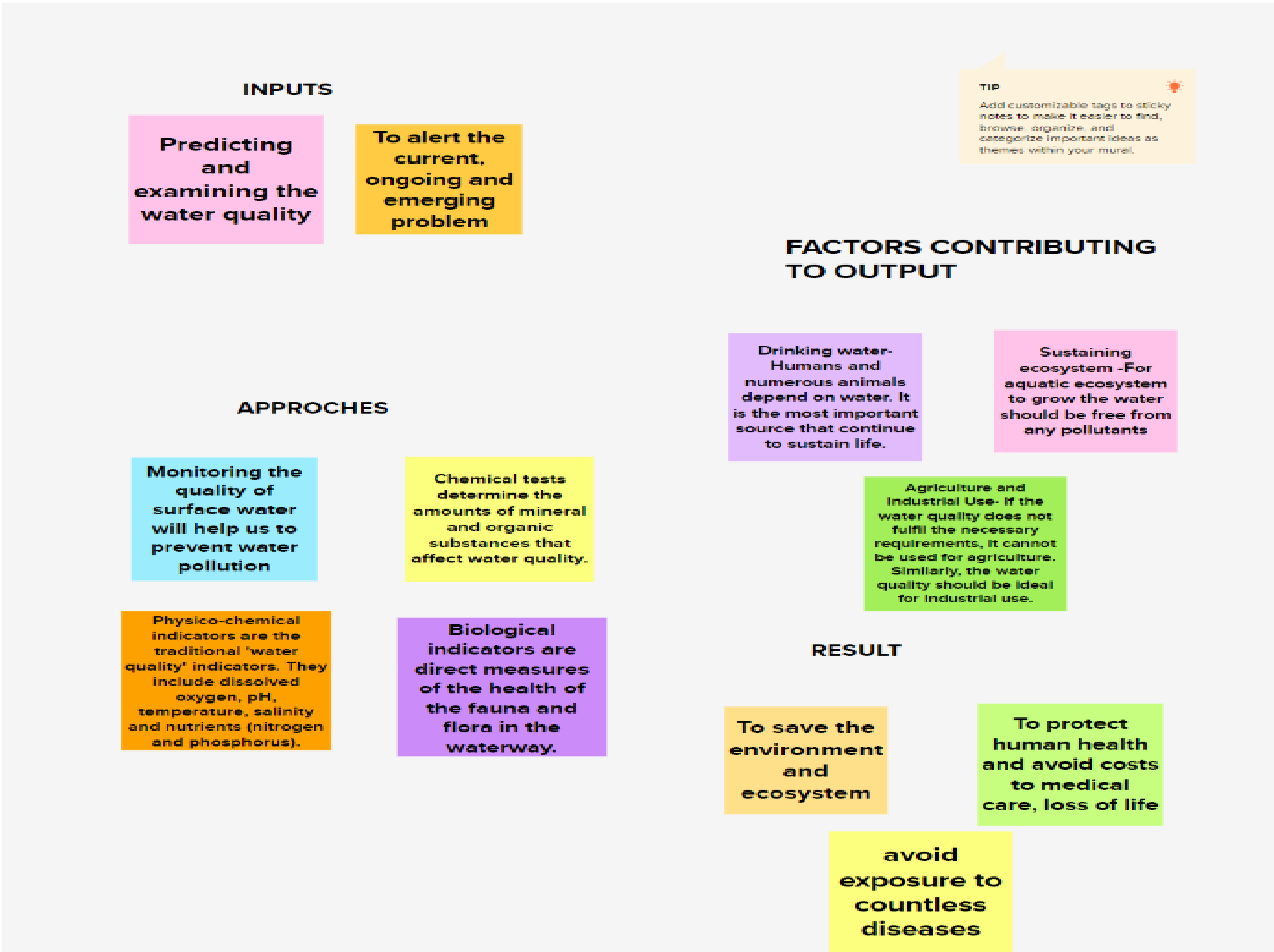
3.IDEATION & PROPOSED SOLUTION

3.1 EMPATHY MAP CANVAS



3.2 IDEATION & BRAINSTORMING




INPUTS
Predicting and examining the water quality
To alert the current, ongoing and emerging problem
TIP
Add customizable tags to sticky notes to make it easier to find, browse, organize, and categorize important ideas as themes within your mural.
FACTORS CONTRIBUTING TO OUTPUT
Drinking water- Humans and numerous animals depend on water. It is the most important source that continue to sustain life.
Sustaining ecosystem -For aquatic ecosystem to grow the water should be free from any pollutants
APPROCHES
Monitoring the quality of surface water will help us to prevent water pollution
Chemical tests determine the amounts of mineral and organic substances that affect water quality.
Agriculture and Industrial Use- If the water quality does not fulfil the necessary requirements, it cannot be used for agriculture. Similarly, the water quality should be ideal for industrial use.
Physico-chemical indicators are the traditional 'water quality' indicators. They include dissolved oxygen, pH, temperature, salinity and nutrients (nitrogen and phosphorus).
Biological indicators are direct measures of the health of the fauna and flora in the waterway.
RESULT
To save the environment and ecosystem
To protect human health and avoid costs to medical care, loss of life
avoid exposure to countless diseases

Sustaining ecosystem -For aquatic ecosystem to grow the water should be free from any pollutants
Several approaches are used to maintain an effective water quality.
Predicting and examing the water quality
Agriculture and Industrial Use- If the water quality does not fulfil the necessary requirements, it cannot be used for agriculture. Similarly, the water quality should be ideal for industrial use.
Upgrading the equipment used to measure the water quality
By measuring the dissolved oxygen level
Conductivity,Salinity ,TDS Monitoring are used to measure the level of organic substance in water.
Importance
If each of these tasks could get done without any difficulty or cost, which would have the most positive impact?
Drinking water- Humans and numerous animals depend on water. It is the most important source that continue to sustain life.
TIP
Participants can use their cursors to point at where sticky notes should go on the grid. The facilitator can confirm the spot by using the laser pointer holding the H key on the keyboard.
Chemical tests determine the amounts of mineral and organic substances that affect water quality.
Monitoring the quality of surface water will help us to prevent water pollution
Feasibility
Regardless of their importance, which tasks are more feasible then others? (Cost, time, effort, complexity, etc.)

3.3 PROPOSED SOLUTION

3.3.1 PROBLEM STATEMENT

Recently water contamination has increased as a result of human activities and environmental conditions. So, drinking water is unsafe to consume. The Software was designed to predict and analyze water quality to get better results.

3.3.2 SOLUTION DESCRIPTION

The project's major objective is to develop a machine-learning model for analyzing and predicting water quality. Gather the data and enter it using the user interface. The integrated model analyses the data that has been entered. After input analysis by the model, the prediction is displayed on the user interface.

3.3.3. NOVELTY

It makes use of past historical data, the user receives information from the UI rapidly, and it also sends information on recycling methods used.

3.3.4 SOCIAL IMPACT

Improved water quality predictions can promote the idea of a healthy country. It can able to determine the precise water level. Intake of clean water promotes a healthy lifestyle. Understand how to efficiently utilize water by reusing it.

3.3.5 BUSINESS MODEL

Can collaborate with governments and companies to measure the water quality and enhance it. We can promote our project to an NGO to enhance the upgrading of urban water quality.

3.3.6 SCALABILITY OF THE SOLUTION

In addition to learning more about the water's temperature, pH level, and salinity, we may also learn more about its physical, chemical, and biological features.

3.4 PROBLEM SOLUTION FIT

Project Title: Efficient water Quality analysis and prediction using Machine learning.
Project Design Phase-I - Solution Fit
Team ID: PNT2022TMID52665

Define CS, fit into CC

1. CUSTOMER SEGMENT(S) CS
Who is your customer?

1) Farmers
2) Elderly people, infants (individuals)
3) School, College, Hospital etc..

6. CUSTOMER CONSTRAINTS CC
What constraints prevent your customers from taking action or limiting their choices?

1)Requires high-quality and efficient water quality analysis at a low price.

2)Unawareness of the new advanced water quality methods.

5. AVAILABLE SOLUTIONS AS
Which solutions are available to the customers when they face a problem or need to get the job done? What have they tried in the past? What pros & cons do these solutions have?

1)Ozone treatment
2)Reverse Osmosis
3) Chlorine Treatment

Pros:
Provides water without harmful microorganisms and unwanted minerals.
Cons:
It is not cost-effective and requires regular maintenance.

Explore AS, differentiate

Focus on J&P, tap into BE, understand RC

2. JOBS-TO-BE-DONE/COMPLETED J&P
Which jobs-to-be-done (or problems) do you address for your customers? There could be more than one; explore different sides.

1)To analyze and predict the Water Quality.
2)The dataset is updated regularly.

9. PROBLEM ROOT CAUSE RC
What is the real reason that this problem exists? What is the back story behind the need to do this job?

1)Individuals' lack of awareness
2) Requires that the machine and water storage devices be properly maintained.

7. BEHAVIOUR BE
What does your customer do to address the problem and get the job done?

1) Convenience, flexibility, and service
2) Consider the project's budget.
3) Determine the precision of the water quality

Focus on J&P, tap into BE, understand RC

3. TRIGGERS TR
What triggers customers to act?
Advertising and educating the people about the importance of water quality for good health.

4. EMOTIONS: BEFORE / AFTER EM
How do customers feel when they face a problem or a job afterward?

Before:
1) **Because the consumer consumes impure water, he is concerned about his health.**
After:
2. **Customers feel that by drinking our project's high-quality water, they are protecting their health.**

10. YOUR SOLUTION SL

Gathering data from many water bodies for the analysis.Our solution incorporates machine learning to find the water quality analysis and give the user more precise findings and some analysis to predict the outcome and generate the outcome.

8. CHANNELS OF BEHAVIOUR CH
8.1 ONLINE
What kind of actions do customers take online?

The consumer carefully reads the information and descriptions when making purchases, and they figure out the overall cost, taxes, services, and other costs.

8.2 OFFLINE
What kind of actions do customers take offline?

1)Consuming filtered water
2)Installing a reverse osmosis system

4. REQUIREMENT ANALYSIS

4.1 FUNCTIONAL REQUIREMENTS

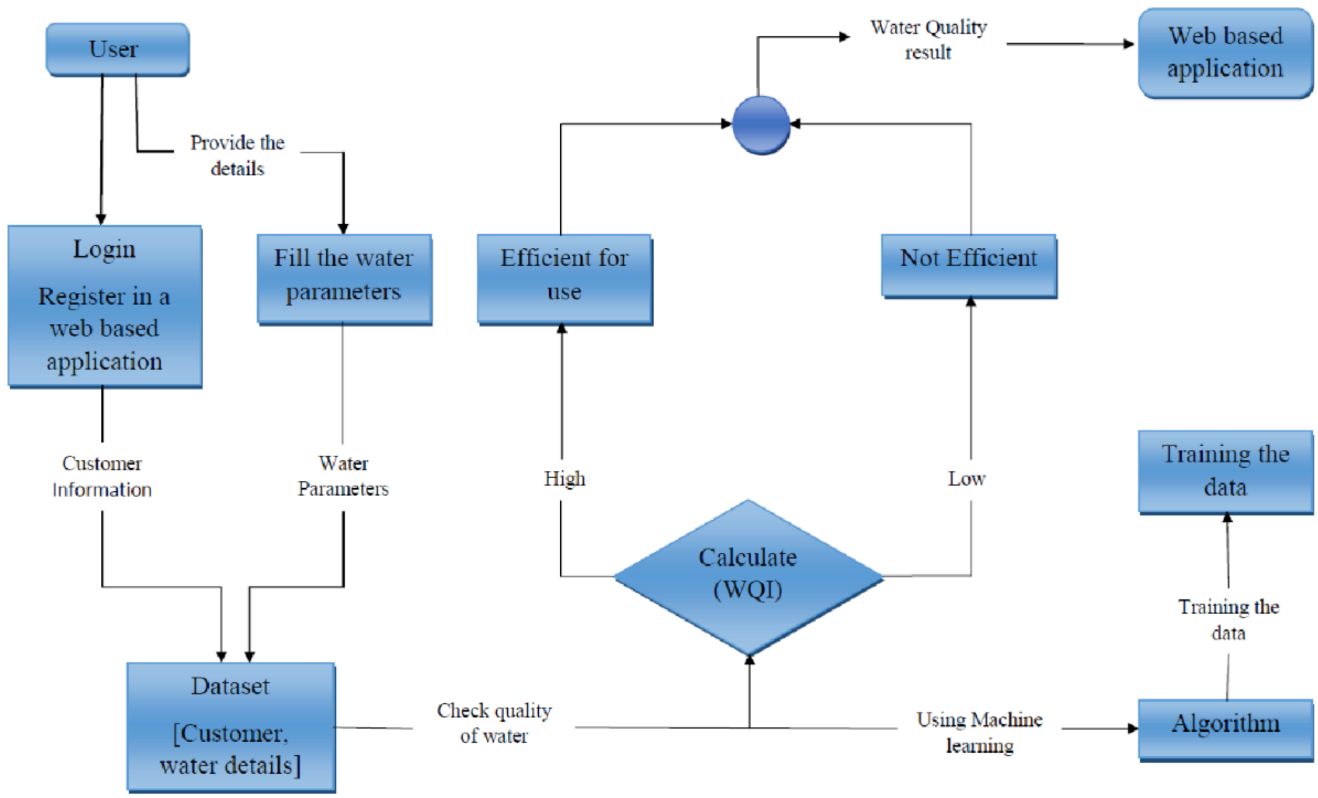
FR No.	Functional Requirement (Epic)	Sub Requirement (Story / Sub-Task)
FR-1	User Registration	Registration through Form Registration through Gmail Registration through LinkedIN
FR-2	User Confirmation	Confirmation via Email Confirmation via OTP
FR-3	Entering required parameters for prediction	With existing user profile they can enter the essential parameters for prediction
FR-4	Performance requirements	Support concurrent users
FR-5	Logical database requirements	The user details stored and encrypted

4.2 NON-FUNCTIONAL REQUIREMENTS

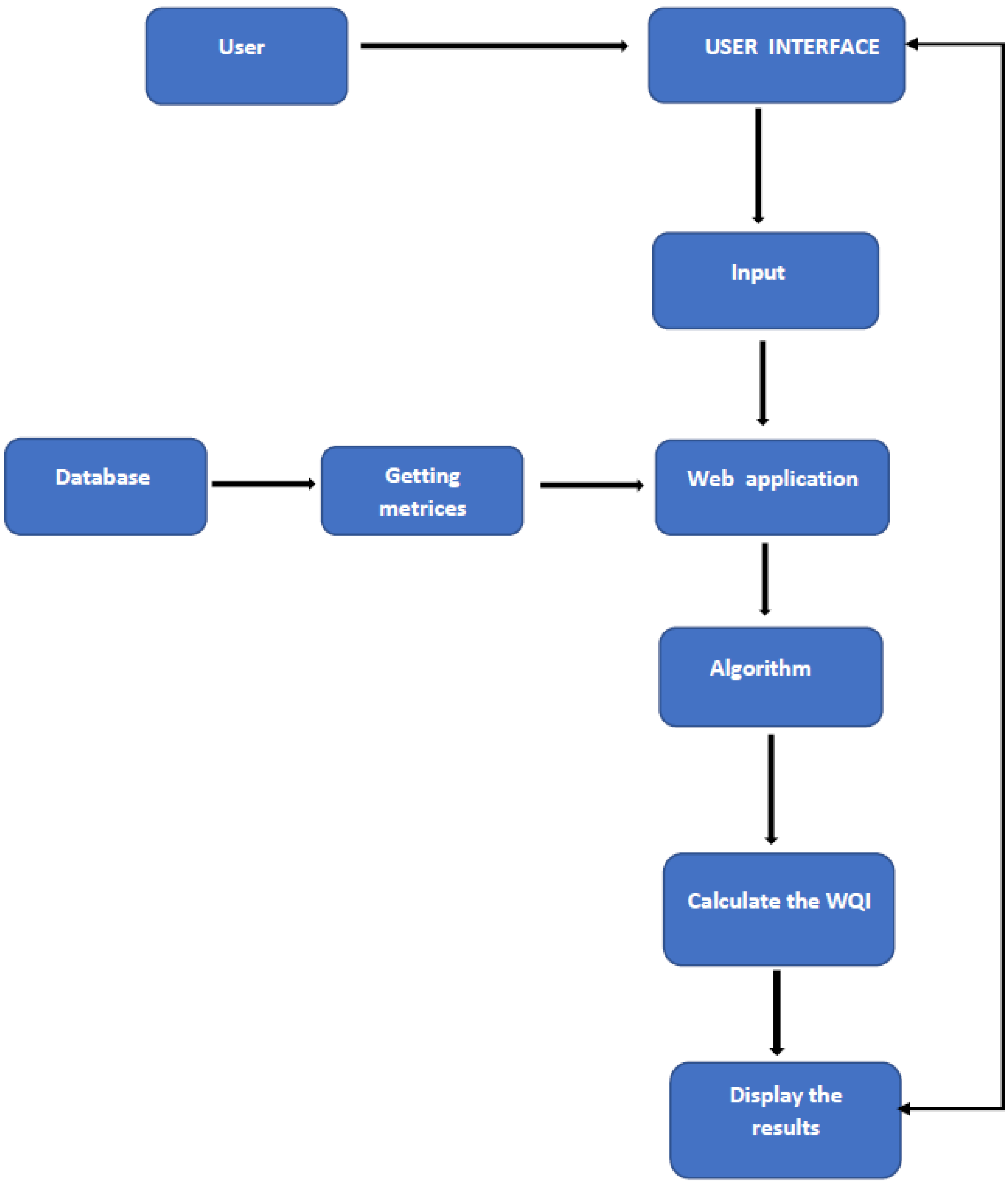
FR No.	Non-Functional Requirement	Description
NFR-1	Usability	This can be used by everyone with internet accessibility
NFR-2	Security	Passwords and user details will be encrypted. The user's IP will be logged
NFR-3	Reliability	The overall reliability depends on separate components
NFR-4	Performance	The application can support many users at a particular time
NFR-5	Availability	The user can access it using a web browser, only restricted by the down time of the server on which the system runs.
NFR-6	Scalability	The end-user part is fully scalable and any system using any web browser should be able to use the features of the application.

5.PROJECT DESIGN

5.1 DATA FLOW DIAGRAM



5.2 SOLUTION AND TECHNICAL ARCHITECTURE



5.3 USER STORIES

User Type	Functional Requirement	User Story Number	User Story / Task	Acceptance criteria	Priority	Release
Customer (Mobile user)	Registration	USN-1	As a user, I can register for the application by entering my email, password, and confirming my password.	I can access my account / dashboard.	High	Sprint-1
		USN-2	As a user, I will receive a confirmation email once I have registered for the application.	I can receive confirmation email & click confirm.	High	Sprint-1

		USN-3	As a user, I can register for the application through Facebook.	I can register & access the dashboard with Facebook Login.	Low	Sprint-2
		USN-4	As a user, I can register for the application through Gmail		Medium	Sprint-1
	Login	USN-5	As a user, I can log into the application by entering email & password		High	Sprint-1
	Dashboard	USN-6	As a user, I can check my login details and work details		High	Sprint-1

Customer (Web user)	Web Access	USN-7	As a user, I can enter the values about the water.	I can access the webpage through the internet.	High	Sprint-1
		USN-8	As a user, I can submit the values into the webpage.	I can click the submit button.	High	Sprint-2
		USN-9	As a user, I expect the correct coefficient of water.		Medium	Sprint-3
	Data pre-processing	USN-10	As a user, I can see the loading information.		Medium	Sprint-3
	User Input Evaluation	USN-11	I can see the evaluation quickly.		High	Sprint-4
	Prediction	USN-12	As a user, I can see the result of the water efficiency.	The results are visible on webpage.	High	Sprint-4

Customer Care Executive	Solving Customer issues.	USN-13	As a customer care executive, I solve the customer issues in using the application and web page.	It re su lts in us er int er ac tio n	Medium	Sprint-5
Administrator		USN-14	I can manage the application.		Medium	Sprint-5

6. PROJECT PLANNING AND SCHEDULING

6.1 SPRINT PLANNING AND ESTIMATION

Sprint	Functional Requirement	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint-1	Data Collection	USN-1	Collecting dataset pre-processing	10	High	Deepthy Shivani A
Sprint-1		USN-2	Data pre -processing. Used to r transform the data into useful format.	10	Medium	Deepthy Shivani A
Sprint-2	Model Building	USN-3	Calculate the Water Quality Index (WQI) using regression algorithm of machine learning	10	High	Divya K
Sprint-2		USN-4	Splitting the data into	10	Medium	Divya K

			training and testing from the entire dataset.			
Sprint-3	Training and Testing	USN-5	Training the model using regression algorithm and testing the performance of the model.	20	Medium	Soundarya M
Sprint-4	Implementation of Web Page	USN-6	Implementing the web page for collecting the data from the user	10	High	Subiksha Devi MS
Sprint-4		USN-6	Deploying the model using IBM cloud and IBM Watson studio	10	Medium	Subiksha Devi MS

6.2 SPRINT DELIVERY SCHEDULE

Sprint : Sprint 1

Functional Requirements: Registration & Login

User Story Number 1: As a user, I can register for the application by entering my email, password, and confirming

User Story Number 2: As a user, I will receive confirmation email
Once I have registered for the application.

User Story Number 3: As a user, I can register for the application through Facebook.

User Story Number 4: As a user, I can register for the application through Gmail

User Story Number 5: As a user, I can log into the application by entering email & password

User Story Number 6: As a user, I can check my login details and work details

User Story Number 7: As a user, I can enter the values about the water.

Sprint : Sprint 2

Functional Requirements: Dashboard & Web Access

User Story Number 8: As a user, I can submit the values into the webpage.

Sprint: Sprint 3

Functional Requirements: Data Pre-processing

User Story Number 9: As a user, I expect correct coefficient of water.

User Story Number 10: As a user, I can see the loading information.

Sprint: Sprint 4

Functional Requirements: User Input Evaluation & Prediction

User Story Number 11: I can see the evaluation quickly.

User Story Number 12: As a user, I can see the result of the water efficient.

Sprint: Sprint 5

Functional Requirements: Solving Customer Issues

User Story Number 13: As a customer care executive, I solve the customer issues in using the application and web page.

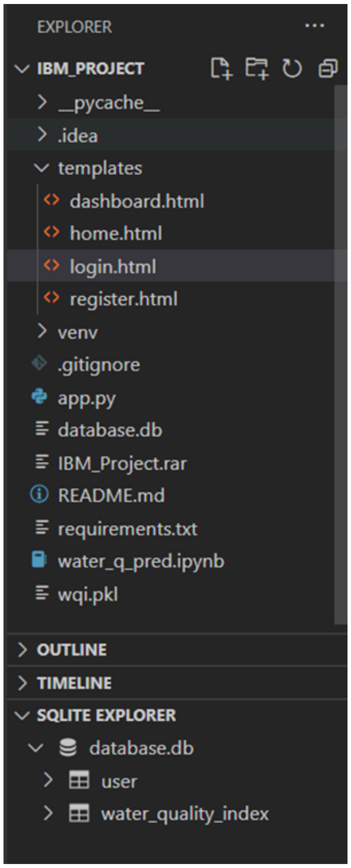
User Story Number 14: I can manage the application.

6.3 REPORTS FROM JIRA

Sprint	Total Story Points	Duration	Sprint Start Date	Sprint End Date (Planned)	Story Points Completed (as on Planned End Date)	Sprint Release Date (Actual)
Sprint-1	20	6 Days	24 Oct 2022	29 Oct 2022	20	29 Oct 2022
Sprint-2	20	6 Days	31 Oct 2022	05 Nov 2022	20	05 Nov 2022
Sprint-3	20	6 Days	07 Nov 2022	12 Nov 2022	20	12 Nov 2022
Sprint-4	20	6 Days	14 Nov 2022	19 Nov 2022	20	19 Nov 2022

7.CODING & SOLUTIONING

Folder Structure:



app.py:

```
from flask import Flask, render_template, url_for, redirect, request
from flask_sqlalchemy import SQLAlchemy
from flask_login import UserMixin, login_user, LoginManager, login_required, logout_user, current_user
from flask_wtf import FlaskForm
from wtforms import StringField, PasswordField, SubmitField, IntegerField, FloatField
from wtforms.validators import InputRequired, Length, ValidationError
from flask_bcrypt import Bcrypt
import pickle
import numpy as np
import sklearn
```

```python
model = pickle.load(open("wqi.pkl", "rb"))

app = Flask(__name__)
db = SQLAlchemy(app)
bcrypt = Bcrypt(app)
app.config['SQLALCHEMY_DATABASE_URI'] = 'sqlite:///database.db'
app.config['SECRET_KEY'] = 'thisisasecretkey'

login_manager = LoginManager()
login_manager.init_app(app)
login_manager.login_view = 'login'

@login_manager.user_loader
def load_user(user_id):
   return User.query.get(int(user_id))


class User(db.Model, UserMixin):
   id = db.Column(db.Integer, primary_key=True)
   username = db.Column(db.String(20), nullable=False, unique=True)
   password = db.Column(db.String(80), nullable=False)

class WaterQualityIndex(db.Model):
   id = db.Column(db.Integer, primary_key=True)
   StationCode = db.Column(db.Integer, nullable=False)
   State = db.Column(db.String(40), nullable=False)
   Temp = db.Column(db.Float, nullable=False)
   do = db.Column(db.Float, nullable=False)
   ph = db.Column(db.Float, nullable=False)
   co = db.Column(db.Integer, nullable=False)
   bod = db.Column(db.Float, nullable=False)
   na = db.Column(db.Float, nullable=False)
   tc = db.Column(db.Integer, nullable=False)
   Year = db.Column(db.Integer, nullable=False)
   WQI = db.Column(db.Float, nullable=False)
```

```
class RegisterForm(FlaskForm):
    username = StringField(validators=[
                                     InputRequired(), Length(min=4, max=20)], render_kw={"placeholder":
"Username"})
    password = PasswordField(validators=[
                                     InputRequired(), Length(min=8, max=20)], render_kw={"placeholder":
"Password"})
    submit = SubmitField('Register')

    def validate_username(self, username):
        existing_user_username = User.query.filter_by(
            username=username.data).first()
        if existing_user_username:
            raise ValidationError(
                'That username already exists. Please choose a different one.')

class WQIForm(FlaskForm):
    StationCode = IntegerField(validators=[
                        InputRequired()], render_kw={"placeholder": "Station Code"})
    State = StringField(validators=[
                         InputRequired(), Length(min=8, max=20)], render_kw={"placeholder": "State"})
    Temp = FloatField(validators=[
                         InputRequired()], render_kw={"placeholder": "Temp"})
    do = FloatField(validators=[
                         InputRequired()], render_kw={"placeholder": "D.O"})
    ph = FloatField(validators=[
                         InputRequired()], render_kw={"placeholder": "PH"})
    co = IntegerField(validators=[
                         InputRequired()], render_kw={"placeholder": "Conductivity"})
    bod = FloatField(validators=[
                         InputRequired()], render_kw={"placeholder": "B.O.D"})
    na = FloatField(validators=[
                         InputRequired()], render_kw={"placeholder": "Nitratenen"})
    tc = IntegerField(validators=[
                         InputRequired()], render_kw={"placeholder": "Coliform"})
    Year = IntegerField(validators=[
                         InputRequired()], render_kw={"placeholder": "Year"})
    submit = SubmitField('Predict')
```

```
class LoginForm(FlaskForm):
    username = StringField(validators=[
                                   InputRequired(), Length(min=4, max=20)], render_kw={"placeholder":
"Username"})
    password = PasswordField(validators=[
                                   InputRequired(), Length(min=8, max=20)], render_kw={"placeholder":
"Password"})
    submit = SubmitField('Login')


@app.route('/')
def home():
    return render_template('home.html')


@app.route('/login', methods=['GET', 'POST'])
def login():
    form = LoginForm()

     if form.validate_on_submit():
        user = User.query.filter_by(username=form.username.data).first()
        if user:
            if bcrypt.check_password_hash(user.password, form.password.data):
                login_user(user)
                return redirect(url_for('dashboard'))
    return render_template('login.html', form=form)


@app.route('/dashboard', methods=['GET', 'POST'])
@login_required
def dashboard():
    form = WQIForm()
    if form.validate_on_submit():
        feature_val = []
        feature_val.append(form.Temp.data)
        feature_val.append(form.do.data)
        feature_val.append(form.ph.data)
```

```
        feature_val.append(form.co.data)
        feature_val.append(form.bod.data)
        feature_val.append(form.na.data)
        feature_val.append(form.tc.data)
        float_features = [float(x) for x in feature_val]
        features = [np.array(float_features)]
        prediction = model.predict(features)
                        new_data = WaterQualityIndex(StationCode=form.StationCode.data, State
=form.State.data, Temp=form.Temp.data, do=form.do.data, ph=form.ph.data, co=form.co.data,
bod=form.bod.data, na=form.na.data, tc=form.tc.data, Year=form.Year.data, WQI=prediction )
        db.session.add(new_data)
        db.session.commit()
         return render_template('dashboard.html',form=form, prediction_text = "The water quality
index is {}".format(prediction))

    return render_template('dashboard.html',form=form)

@app.route('/logout', methods=['GET', 'POST'])
@login_required
def logout():
    logout_user()
    return redirect(url_for('login'))


@ app.route('/register', methods=['GET', 'POST'])
def register():
    form = RegisterForm()
    if form.validate_on_submit():
        hashed_password = bcrypt.generate_password_hash(form.password.data)
        new_user = User(username=form.username.data, password=hashed_password)
        db.session.add(new_user)
        db.session.commit()
        return redirect(url_for('login'))

    return render_template('register.html', form=form)

if __name__ == "__main__":
    app.run(debug=True)
```

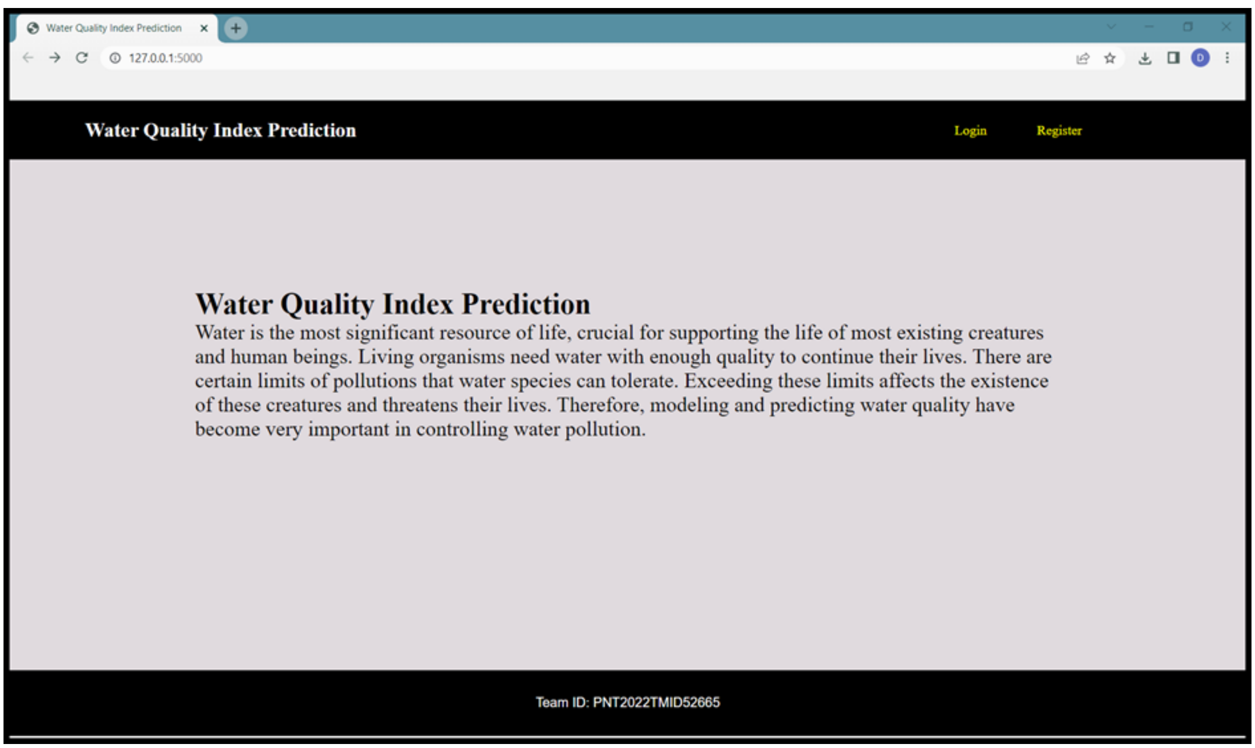
home.html

```
<!DOCTYPE html>
<html>
<head>
  <meta charset="UTF-8">
  <meta http-equiv="X-UA-Compatible" content="IE=edge">
  <meta name="viewport" content="width=device-width, initial-scale=1.0">
  <title>Water Quality Index Prediction</title>
  <style>
    * {
      margin: 0;
      padding: 0;
    }

    .navbar {
      display: flex;
      align-items: center;
      justify-content: center;
      position: sticky;
      top: 0;
      cursor: pointer;
    }

    .background {
      background: black;
      background-blend-mode: darken;
      background-size: cover;
    }

    .nav-list {
      width: 30%;
      display: flex;
      align-items: center;
    }

    .logo {
      padding-left: 90px;
      justify-content: center;
      align-items: center;
      width: 80%;
```

```
    color:white;
    font-weight:bold;
}


.nav-list li {
    list-style: none;
    padding: 26px 30px;
}

.nav-list li a {
    text-decoration: none;
    color: yellow;
}

.nav-list li a:hover {
    color: grey;
}

.firstsection {
    background-image: url("static/images/img1.jpeg");
    background-color: #E0DADE;
    height: 80vh;
}

.box-main {
    display: flex;

    justify-content: center;
    align-items: center;
    color: black;
    max-width: 70%;
    margin: auto;
    height: 80%;
}

.firsthalf {
    width: 80%;
    display: flex;
    flex-direction: column;
```

```
            justify-content: center;
        }

        .text-big {
            font-family: 'Piazzolla', serif;
            font-weight: bold;
            font-size: 35px;
        }

        .text-small {
            font-size: 25px;
        }

        .text-footer {
            text-align: center;
            padding: 30px 0;
            font-family: 'Ubuntu', sans-serif;
            display: flex;
            justify-content: center;
            color: white;
        }
    </style>
</head>

<body>
    <nav class="navbar background">
        <div class="logo">
                <h1>Water Quality Index Prediction</h1>
        </div>

        <ul class="nav-list">
            <li><a href="{{ url_for('login') }}">Login</a></li>
            <li><a href="{{ url_for('register') }}">Register</a></li>
        </ul>
    </nav>

    <section class="firstsection">
        <div class="box-main">
            <div class="firstHalf">
                <h1 class="text-big" id="web">Water Quality Index Prediction</h1>
```

```
        <p class="text-small">
            Water is the most significant resource of life, crucial for supporting the life of most existing creatures and human beings. Living organisms need water with enough quality to continue their lives. There are certain limits of pollutions that water species can tolerate. Exceeding these limits affects the existence of these creatures and threatens their lives. Therefore, modeling and predicting water quality have become very important in controlling water pollution.
        </p>
      </div>
    </div>
  </section>

  <footer class="background">
    <p class="text-footer">
      Team ID: PNT2022TMID52665
    </p>
  </footer>
</body>
</html>
```



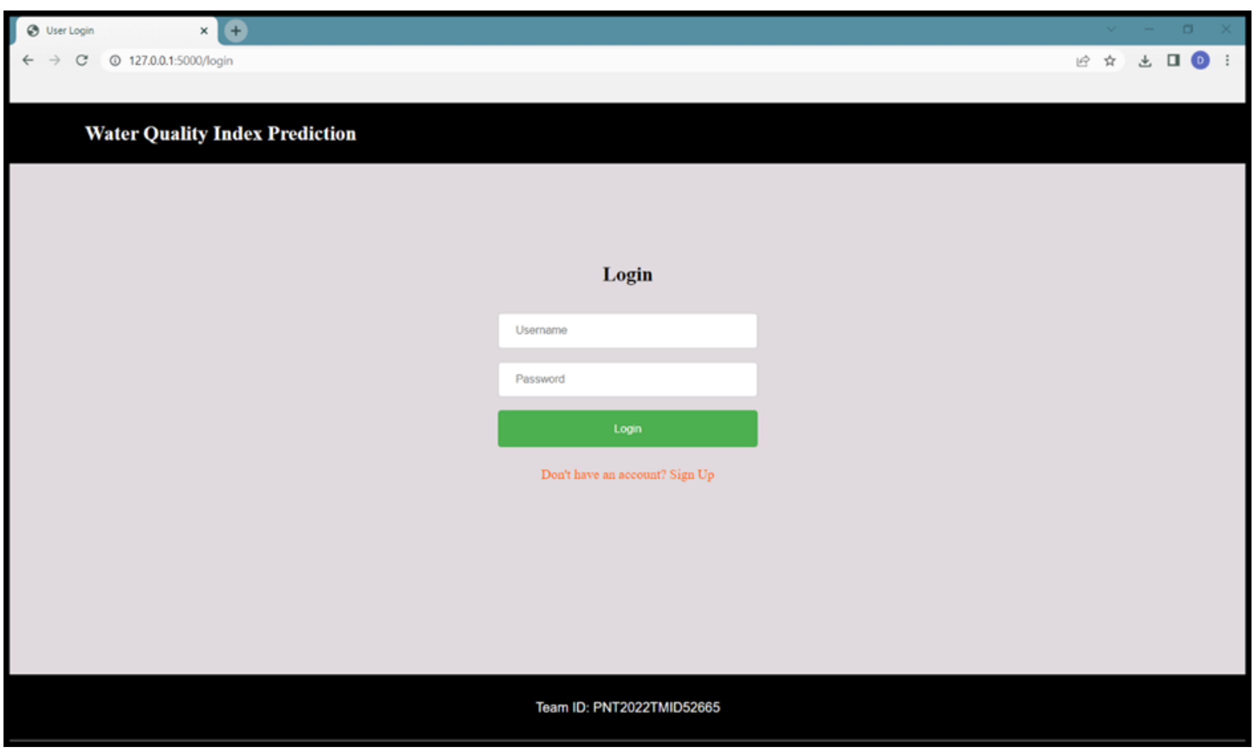
login.html:

```
<!DOCTYPE html>
<html lang="en">
<head>
   <meta charset="UTF-8">
   <meta http-equiv="X-UA-Compatible" content="IE=edge">
   <meta name="viewport" content="width=device-width, initial-scale=1.0">
   <title>User Login</title>
   <style>
            * {
         margin: 0;
         padding: 0;
      }

      .navbar {
         display: flex;
         align-items: center;
         justify-content: center;
         position: sticky;
         top: 0;
         height: 9.5vh;
         cursor: pointer;
      }

      .background {
         background: black;
         background-blend-mode: darken;
         background-size: cover;
      }
      .logo {
         padding-left: 90px;
         justify-content: center;
         align-items: center;
         width: 100%;
         color:white;
         font-weight:bold;
      }
      .firstsection {
         background-image: url("static/images/img1.jpeg");
         background-color: #E0DADE;
```

```
    height: 80vh;
  }

  .box-main {
    display: flex;
    flex-direction: column;
    justify-content: center;
    align-items: center;
    color: black;
    max-width: 70%;
    margin: auto;
    height: 80%;
  }

  .firstHalf {
    width: 30%;
    display: flex;
    flex-direction: column;
    justify-content: center;
  }

  .form-footer a{
    text-decoration: none;
    color: rgb(255, 94, 0);
    justify-content: center;
  }

  .form-footer a:hover{
    text-decoration: none;
    color: rgb(38, 4, 159);
    justify-content: center;
  }
  .text-footer {
    text-align: center;
    padding: 30px 0;
    font-family: 'Ubuntu', sans-serif;
    display: flex;
    justify-content: center;
    color: white;
}
```

```
        .form-header {
            padding: 25px;
        }

        .form-footer {
            padding: 15px;
        }

        input {
            width: 100%;
            padding: 12px 20px;
            margin: 8px 0;
            display: inline-block;
            border: 1px solid #ccc;
            border-radius: 4px;
            box-sizing: border-box;
        }

        input[type=submit] {
            width: 100%;
            background-color: #4CAF50;
            color: white;
            padding: 14px 20px;
            margin: 8px 0;
            border: none;
            border-radius: 4px;
            cursor: pointer;
        }
    </style>
</head>

<body>
    <nav class="navbar background">
        <div class="logo">
            <h1>Water Quality Index Prediction</h1>
        </div>
    </nav>
    <section class="firstsection">
        <div class="box-main">
            <div class="form-header">
```

```
            <h2>Login</h2>
        </div>
        <div class="firstHalf">
            <form method="POST" action="">
                <p>{{ form.hidden_tag() }}</p>
                <p>{{ form.username }}</p>
                <p>{{ form.password }}</p>
                <p>{{ form.submit }}</p>
            </form>
        </div>
        <div class="form-footer">
            <a href="{{ url_for('register') }}">Don't have an account? Sign Up</a>
        </div>
    </div>
  </section>
  <footer class="background">
    <p class="text-footer">
        Team ID: PNT2022TMID52665
    </p>
  </footer>
</body>
</html>
```



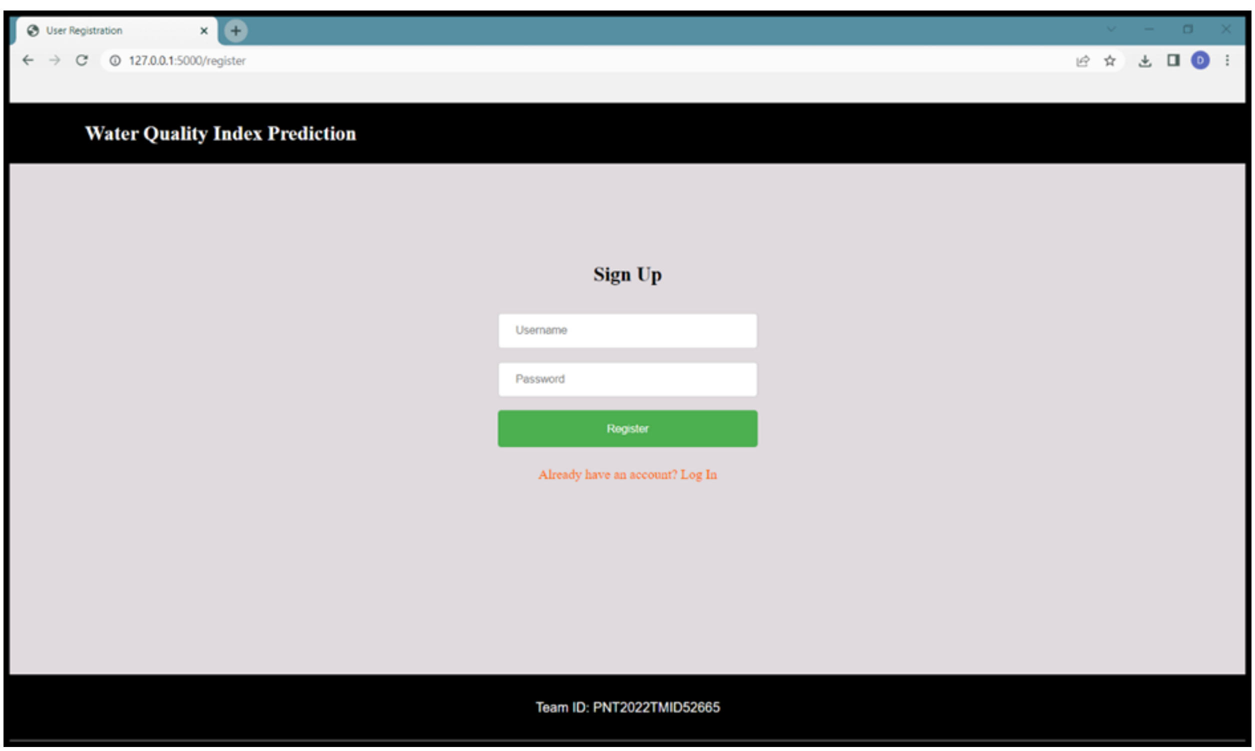
register.html

```
<!DOCTYPE html>
<html lang="en">
<head>
  <meta charset="UTF-8">
  <meta http-equiv="X-UA-Compatible" content="IE=edge">
  <meta name="viewport" content="width=device-width, initial-scale=1.0">
  <title>User Registration</title>
  <style>
    * {
    margin: 0;
    padding: 0;
  }
  .navbar {
    display: flex;
    align-items: center;
    justify-content: center;
    position: sticky;
    top: 0;
    height: 9.5vh;
    cursor: pointer;
  }
  .background {
    background: black;
    background-blend-mode: darken;
    background-size: cover;
  }
  .logo {
    padding-left: 90px;
    justify-content: center;
    align-items: center;
    width: 100%;
    color:white;
    font-weight:bold;
  }
  .firstsection {
    background-image: url("static/images/img1.jpeg");
    background-color: #E0DADE;
    height: 80vh;
 }
```

```
.box-main {
   display: flex;
   flex-direction: column;
   justify-content: center;
   align-items: center;
   color: black;
   max-width: 70%;
   margin: auto;
   height: 80%;
}

.firstHalf {
   width: 30%;
   display: flex;
   flex-direction: column;
   justify-content: center;
}

.form-footer a{
   text-decoration: none;
   color: rgb(255, 94, 0);
   justify-content: center;
}
.form-footer a:hover{
   text-decoration: none;
   color: rgb(38, 4, 159);
   justify-content: center;
}
.text-footer {
   text-align: center;
   padding: 30px 0;
   font-family: 'Ubuntu', sans-serif;
   display: flex;
   justify-content: center;
   color: white;
}

.form-header {
   padding: 25px;
}
```

```
    .form-footer {
        padding: 15px;
    }

    input {
        width: 100%;
        padding: 12px 20px;
        margin: 8px 0;
        display: inline-block;
        border: 1px solid #ccc;
        border-radius: 4px;
        box-sizing: border-box;
    }

    input[type=submit] {
        width: 100%;
        background-color: #4CAF50;
        color: white;
        padding: 14px 20px;
        margin: 8px 0;
        border: none;
        border-radius: 4px;
        cursor: pointer;
    }
</style>
</head>

<body>
    <nav class="navbar background">
        <div class="logo">
            <h1>Water Quality Index Prediction</h1>
        </div>
    </nav>

    <section class="firstsection">
        <div class="box-main">
            <div class="form-header">
                <h2>Sign Up</h2>
            </div>
            <div class="firstHalf">
```

```
            <form method="POST" action="">
                {{ form.hidden_tag() }}
                {{ form.username }}
                {{ form.password }}
                {{ form.submit }}
            </form>
        </div>
        <div class="form-footer">
            <a href="{{ url_for('login') }}">Already have an account? Log In</a>
        </div>
    </div>
</section>

<footer class="background">
    <p class="text-footer">
        Team ID: PNT2022TMID52665
    </p>
</footer>
</body>
</html>
```



dashboard.html

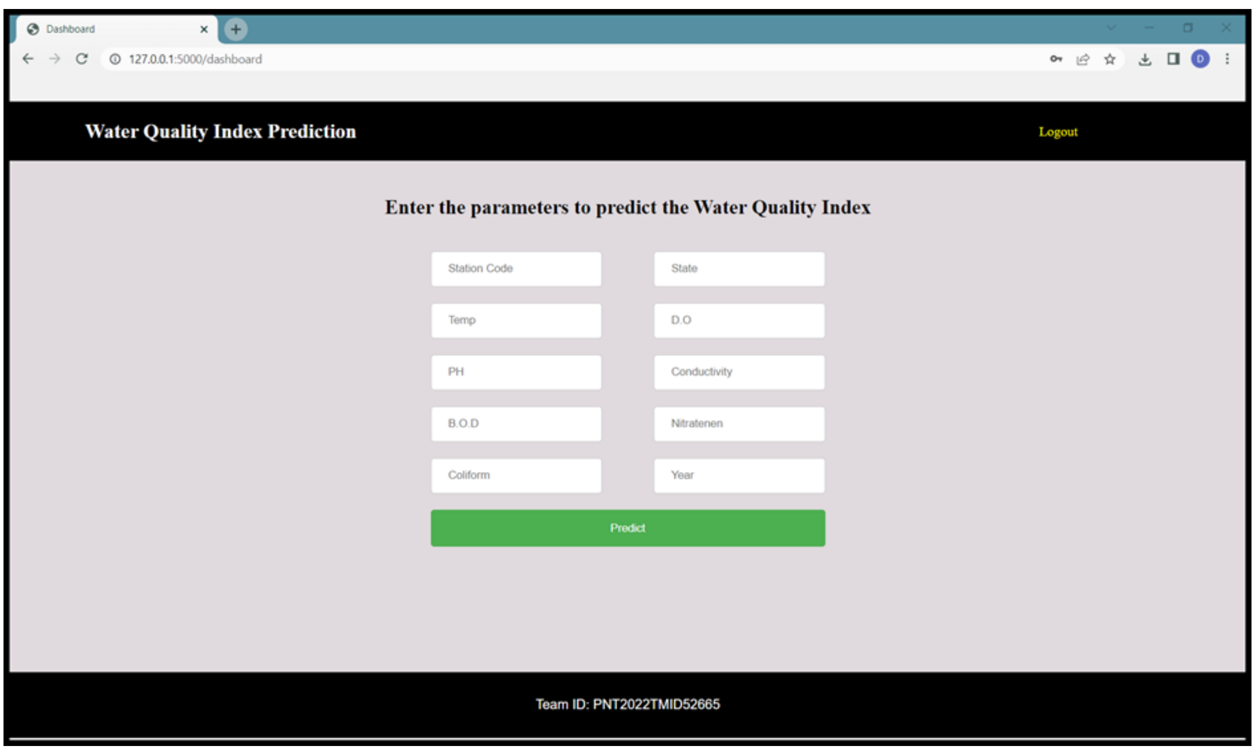
```
<!DOCTYPE html>
<html lang="en">
<head>
  <meta charset="UTF-8">
  <meta http-equiv="X-UA-Compatible" content="IE=edge">
  <meta name="viewport" content="width=device-width, initial-scale=1.0">
  <title>Dashboard</title>
  <style>
    * {
      margin: 0;
      padding: 0;
    }
    .navbar {
      display: flex;
      align-items: center;
      justify-content: center;
      position: sticky;
      top: 0;
      cursor: pointer;
    }
    .background {
      background: black;
      background-blend-mode: darken;
      background-size: cover;
    }

    .nav-list {
      width: 20%;
      display: flex;
      align-items: center;
    }
    .logo {
      padding-left: 90px;
      justify-content: center;
      align-items: center;
      width: 80%;
      color:white;
      font-weight:bold;
    }
```

```css
.nav-list li {
    list-style: none;
    padding: 26px 30px;
}

.nav-list li a {
    text-decoration: none;
    color: yellow;
}

.nav-list li a:hover {
    color: grey;
}

.firstsection {
    background-image: url("static/images/img1.jpeg");
    background-color: #E0DADE;
    height: 80vh;
}

.box-main {
    display: flex;
    flex-direction: column;
    justify-content: center;
    align-items: center;
    color: black;
    max-width: 70%;
    margin: auto;
    height: 80%;
}
.form-footer a{
    text-decoration: none;
    color: rgb(255, 94, 0);
    justify-content: center;
}
.form-footer a:hover{
    text-decoration: none;
    color: rgb(38, 4, 159);
    justify-content: center;
}
```

```
        .text-footer {
            text-align: center;
            padding: 30px 0;
            font-family: 'Ubuntu', sans-serif;
            display: flex;
            justify-content: center;
            color: white;
        }

        .form-header {
            padding: 25px;
        }

        .form-footer {
            padding: 15px;
        }
        input {
            width: 100%;
            padding: 12px 20px;
            margin: 8px 0;
            display: inline-block;
            border: 1px solid #ccc;
            border-radius: 4px;
            box-sizing: border-box;
        }
        input[type=submit] {
            width: 100%;
            background-color: #4CAF50;
            color: white;
            padding: 14px 20px;
            margin: 8px 0;
            border: none;
            border-radius: 4px;
            cursor: pointer;
        }
        td {
            padding: 1px 30px;
        }
    </style>
</head>
```

```
<body>
  <nav class="navbar background">
    <div class="logo">
        <h1>Water Quality Index Prediction</h1>
    </div>

    <ul class="nav-list">
      <li><a href="{{url_for('logout')}}">Logout</a></li>
    </ul>
  </nav>

  <section class="firstsection">
    <div class="box-main">
      <div class="form-header">
        <h2>Enter the parameters to predict the Water Quality Index</h2>
      </div>
      <table>
        <form action="" method="post">
          <tr>
            <td>{{ form.hidden_tag() }}</td>
          </tr>
          <tr>
            <td>{{ form.StationCode }}</td>
            <td>{{ form.State }}</td>
          </tr>
          <tr>
            <td>{{ form.Temp }}</td>
            <td>{{ form.do }}</td>
          </tr>
          <tr>
            <td>{{ form.ph }}</td>
            <td>{{ form.co }}</td>
          </tr>
          <tr>
            <td>{{ form.bod }}</td>
            <td>{{ form.na }}</td>
          </tr>
          <tr>
            <td>{{ form.tc }}</td>
            <td>{{ form.Year }}</td>
```

```
                </tr>
                <tr>
                    <td colspan = "2">{{ form.submit }}</td>
                </tr>
            </form>
        </table>
        <div>
            <h2>{{ prediction_text }}</h2>
        </div>
    </div>
</section>

<footer class="background">
    <p class="text-footer">
        Team ID: PNT2022TMID52665
    </p>
</footer>
</body>
</html>
```



DATABASE:

user table:

SQLite - IBM_Project - Visual Studio Code

SQLite

SQL ▾ 1 / 1 1 - 2 of 2

id	username	password
1	divya	$2b$12$4gontoUSl4/Cr1B3sSiy2ezKCpqOgVIt.8YoIsMoRz1WtYd2BtfYm
2	user1	$2b$12$4bitAN/YTTrqLP55.SBK6.jqazi4J.5UcremUDrgxD06M6UwWPuke

water_quality_index table:

SQLite - IBM_Project - Visual Studio Code

SQLite

SQL ▾ 1 / 1 1 - 4 of 4

id	StationCode	State	Temp	do	ph	co	bod	na	tc	Year	WQI
1	4521	TamilNadu	30.6	6.7	7.5	203	6.9	0.1	27	2018	45.726
2	6542	AndhraPradesh	32.8	7.1	8.1	155	5.0	0.2	20	2021	45.726
3	6541	AndhraPradesh	29.3	7.4	6.8	121	1.7	0.4	18125	2020	47.13
4	8452	TamilNadu	22.4	5.2	8.8	180	3.5	0.5	27	2022	45.726

Terminal:

```
PROBLEMS    OUTPUT    DEBUG CONSOLE    TERMINAL    JUPYTER

 * Serving Flask app 'app' (lazy loading)
 * Environment: production
   WARNING: This is a development server. Do not use it in a production deployment.
   Use a production WSGI server instead.
 * Debug mode: on
 * Running on http://127.0.0.1:5000 (Press CTRL+C to quit)
 * Restarting with stat
C:\Users\HP\Desktop\IBM_Project\venv\lib\site-packages\sklearn\base.py:329: UserWarning: Trying to unpickle estimator De
cisionTreeRegressor from version 0.23.2 when using version 1.1.3. This might lead to breaking code or invalid results. U
se at your own risk. For more info please refer to:
https://scikit-learn.org/stable/model_persistence.html#security-maintainability-limitations
  warnings.warn(
C:\Users\HP\Desktop\IBM_Project\venv\lib\site-packages\sklearn\base.py:329: UserWarning: Trying to unpickle estimator Ra
ndomForestRegressor from version 0.23.2 when using version 1.1.3. This might lead to breaking code or invalid results. U
se at your own risk. For more info please refer to:
https://scikit-learn.org/stable/model_persistence.html#security-maintainability-limitations
  warnings.warn(
C:\Users\HP\Desktop\IBM_Project\venv\lib\site-packages\flask_sqlalchemy\__init__.py:851: UserWarning: Neither SQLALCHEMY
_DATABASE_URI nor SQLALCHEMY_BINDS is set. Defaulting SQLALCHEMY_DATABASE_URI to "sqlite:///:memory:".
  warnings.warn(
C:\Users\HP\Desktop\IBM_Project\venv\lib\site-packages\flask_sqlalchemy\__init__.py:872: FSADeprecationWarning: SQLALCHE
MY_TRACK_MODIFICATIONS adds significant overhead and will be disabled by default in the future.  Set it to True or False
 to suppress this warning.
  warnings.warn(FSADeprecationWarning(
 * Debugger is active!
 * Debugger PIN: 119-411-840
```

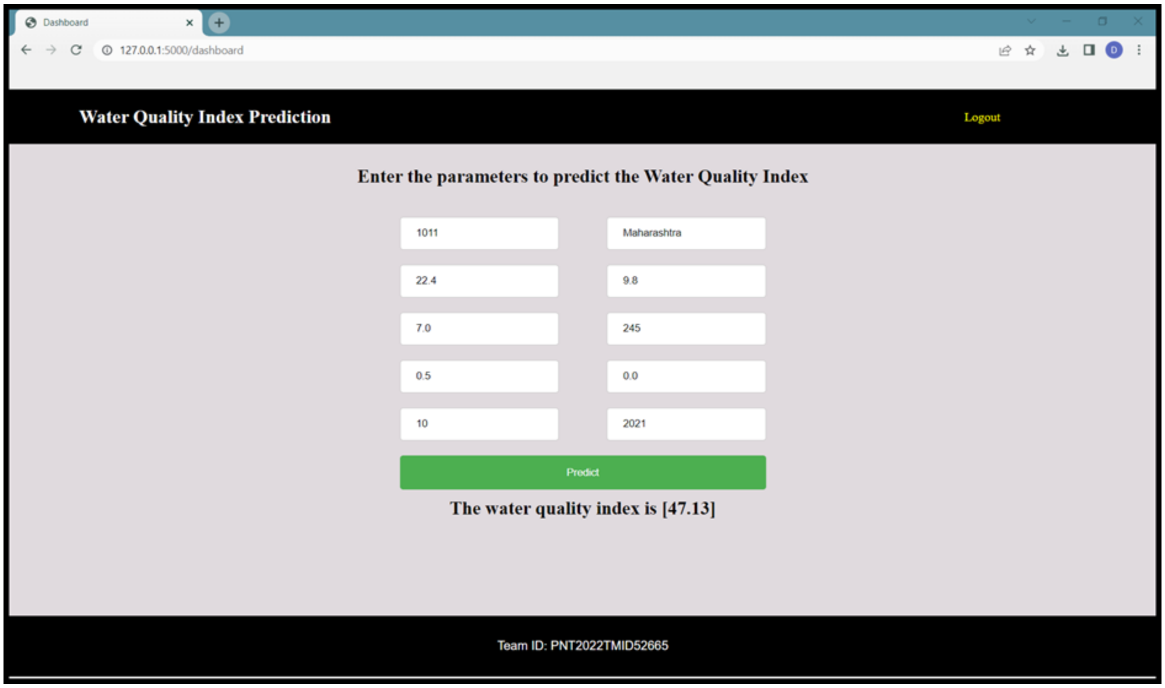
To run the application:

url: http://127.0.0.1:5000

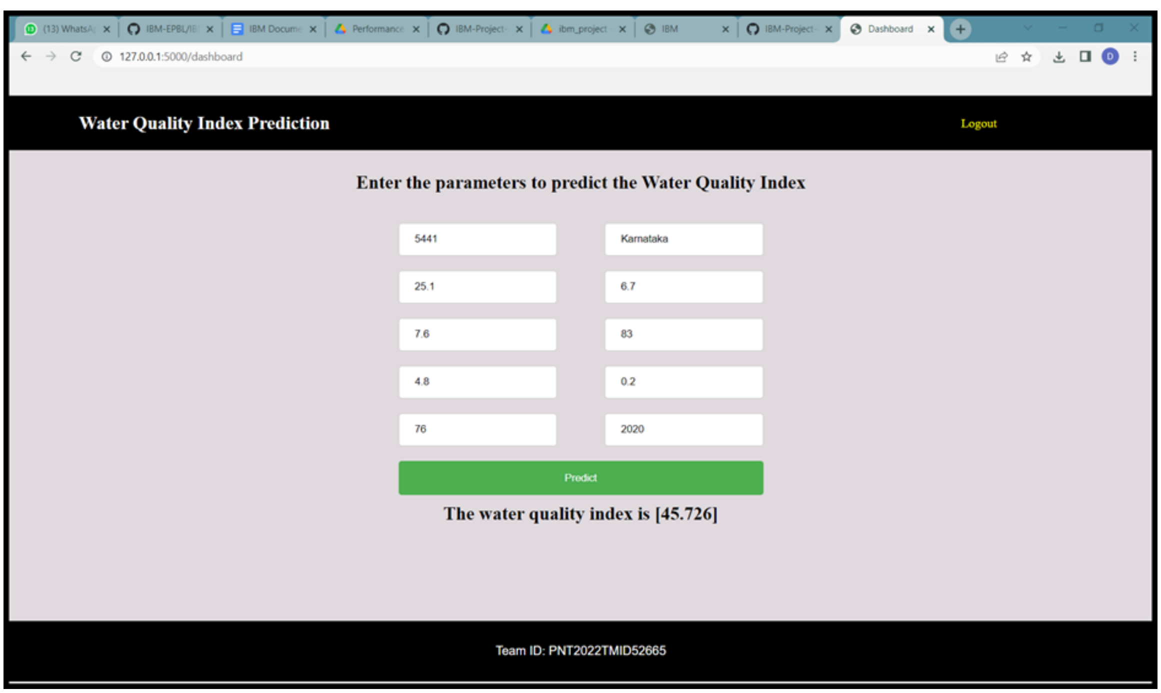
8.TESTING

8.1 TEST CASES

TEST CASE 1:



TEST CASE 2:



8.2 USER ACCEPTANCE TESTING

Purpose of Document

The purpose of this document is to briefly explain the test coverage and open issues of the r efficient water quality analysis and prediction project at the time of the release to User Acceptance Testing (UAT).

Defect Analysis

This report shows the number of resolved or closed bugs at each severity level, and how they were resolved

Resolution	Severity 1	Severity 2	Severity 3	Severity 4	Subtotal
By Design	**10**	**4**	**2**	**3**	**20**
Duplicate	**1**	**0**	**3**	**0**	**4**
External	**2**	**3**	**0**	**1**	**6**
Fixed	**11**	**2**	**4**	**20**	**37**
Not Reproduced	**0**	**0**	**1**	**0**	**1**
Skipped	**0**	**0**	**1**	**1**	**2**

Won't Fix	**0**	**5**	**2**	**1**	**8**
Totals	**24**	**14**	**13**	**26**	**7**

Test Case Analysis

This report shows the number of test cases that have passed, failed, and untested

Section	Total Cases	Not Tested	Fail	Pass
Print Engine	**7**	**0**	**0**	**7**
Client Application	**51**	**0**	**0**	**51**
Security	**2**	**0**	**0**	**2**
Outsource Shipping	**3**	**0**	**0**	**3**
Exception Reporting	**9**	**0**	**0**	**9**
Final Report Output	**4**	**0**	**0**	**4**
Version Control	**2**	**0**	**0**	**2**

9.RESULTS

9.1 PERFORMANCE METRICS

For validating the developed model, the dataset has been divided into 70% training and 30% testing subsets. While the ANN and LSTM models were used to predict the WQI, the SVM, KNN, and Naive Bayes were utilized for the water quality classification prediction Therefore, Random Forest Regressor is used. Performance Measures Results True Positives (TP) are when the model predicts the positive class properly. True Negatives (TN) is one of the components of a confusion matrix designed to demonstrate how classification algorithms work. Positive outcomes that the model predicted incorrectly are known as False Positives (FP). False Negatives (FN) are negative outcomes that the model predicts negative class. Accuracy is the most basic and intuitive performance metric, consisting of the ratio of successfully predicted observations to total observations.

Accuracy = TP+TN/ (TP+FP+FN+TN)

RANDOM FOREST REGRESSOR

Accuracy =96.8

Mean Absolute error = 1.00485

Mean Squared error = 5.84

10. ADVANTAGES AND DISADVANTAGES

10.1 ADVANTAGES

- Whether it be for groundwater, surface water or open water, there are a number of reasons why it is important for you to undertake regular water quality testing. This testing will also allow you to adhere to strict permit regulations and be in compliance with Australian laws.
- Identifying the health of your water will help you to discover where it may need some help. Ultimately, finding a source of pollution, or remaining proactive with your monitoring will enable you to save money in the long term. The more information that you can obtain will assist you with your decision on what product you may need to improve the condition of your water. Simply guessing and buying products based on a hunch or a general trend is ill-advised, as each body of water has unique properties that can only be discovered through testing.
- Measuring the amount of dissolved oxygen in your water is another important advantage of water quality testing, as typically the less oxygen, the higher the water temperature, resulting in a more harmful environment for aquatic life. These levels do fluctuate slightly across the seasons, but regular monitoring of your water quality will allow you to discover trends over time, and whether there are other factors that may be contributing to the results you discover.

10.2 DISADVANTAGES

- Training necessary and difficult to manage over time and with large data sets.
- Requires manual operation to submit data cannot respond to data queries from other nodes, and therefore cannot interact with the Exchange Network Technical expertise and network server required.
- Costly, usually only feasible under Exchange Network grants Technical expertise and network server required
- Requires manual operation to submit data, some configuration.

11.CONCLUSION

Water is one of the most essential resources for survival and its quality is determined through WQI. Traditionally, a costly and drawn-out lab analysis was needed to test the quality of water. This study investigated a different machine learning approach for forecasting water quality based on just a few basic water quality variables. A group of representative supervised machine learning methods were utilised to estimate. Before water was made available for consumption, it would find water of poor quality and alert the proper authorities. By decreasing the number of people who consume poor water, it should help prevent illnesses like typhoid and diarrhoea. The application of a prescriptive analysis based on expected values in this situation would lead to the development of future tools to support decision- and policy-makers.To this end most dataset related wellknown components, such as pH, SO4, Na, Ca, Cl, Mg, HCO3 etc., were collected. Results indicated that the applied models have suitable performance for predicting water quality.

12. FUTURE SCOPE

Environmental water quality monitoring aims to provide the data required for safeguarding the environment against adverse biological effects from multiple chemical contamination arising from anthropogenic diffuse emissions and point sources. strategies for identifying river basin specific pollutants, improvements in the diagnostics of ecological impacts and more powerful approaches for establishing causal links between chemical and ecological assessments are required.

13. APPENDIX

[1]. Ali El Bilali, Abdeslam Taleb, "Prediction of irrigation water quality parameters using machine learning models in a semi-arid environment",Journal of the Saudi Society of Agricultural Sciences ,Volume 19, Issue 7, October 2020, Pages 439-451

[2]. Qinghong Zou, Qingyu Xiong, Hualing Yi, " A water quality prediction method based on the multi-time scale bidirectional LSTM network",Environ Sci Pollut Res Int. 2020 May.

[3]. Sudhakar Singha, Srinivas Pasupuleti,Soumya S. Singha, Rambabu Singh, Suresh Kumar"Prediction of groundwater quality using efficient machine learning technique",Chemosphere,March 2021.

[4]. Md.Mahedi Hassan, Laboni Akter, Mushfiqur Rahman, Nusrat Jahan, "Efficient Prediction of Water Quality Index(WQI) Using Machine Learning Algorithms",Human-Centric Intelligent Systems,Volume 1, Issue 3-4, December 2021.

[5].Tianan Deng, Kwok-Wing Chau, Huan-Feng Duan, "Machine learning based marine water quality prediction for coastal hydro-environment management",J Environ Manage,Apr 15 2021.

[6]. Smail Dilmi, Mohamed Ladjal, “A novel approach for water quality classification based on the integration of deep learning and feature extraction techniques”,Chemometrics and Intelligent Laboratory Systems,July 2021.

[7]. S.Sharath, R.Harish, V.Aishwarya and Dr.M.Preetha, “Water quality prediction using machine learning and flask”,International Research Journal of Modernization in Engineering Technology and Science,Volume:0,/Issue:06,June-2022.

[8]. Maryam Imani, Md Mahmudul Hasan, Luiz Fernando Bittencourt, Kent McClymont and Zoran Kapelan, “A novel machine learning application: Water quality resilience prediction Model”,Science of The Total Environment,January 2021

[9]. Huimin Jia, Xiaofeng Zhou, “Water Quality Prediction Method Based on LSTM-BP”,2020 12th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC),in IEEE,23 August 2020.

[10]. Al-Akhir Nayan, Muhammad Golam Kibria, Md. Obaidur Rahman and Joyeta Saha, “River Water Quality Analysis and Prediction Using GBM”,2020 2nd International Conference on Advanced Information and Communication Technology (ICAICT),29 November 2020.

asgiref==3.5.0

bcrypt==3.2.0

cffi==1.15.0

click==8.1.2

dnspython==2.2.1

email-validator==1.1.3

Flask==2.1.1

Flask-Bcrypt==1.0.1

Flask-Login==0.6.0

Flask-SQLAlchemy==2.5.1

Flask-WTF==1.0.1

greenlet==1.1.2

idna==3.3

importlib-metadata==4.11.3

itsdangerous==2.1.2

Jinja2==3.1.1

MarkupSafe==2.1.1

pycparser==2.21

six==1.16.0

SQLAlchemy==1.4.35

sqlparse==0.4.2

Werkzeug==2.1.1

WTForms==3.0.1

zipp==3.8.

GITHUB LINK

https://github.com/IBM-EPBL/IBM-Project-48474-1660807603

VIDEO DEMO LINK

https://drive.google.com/file/d/1s6qJ0eMtPlCnB9IjIOkelkLZ1rfsDsg3/view?usp=sharing