PROJECT REPORT

EFFICIENT WATER QUALITY ANALYSIS AND PREDICTION

By team – PNT2022TMID14644
Batch no – B7-1A3E

Kumaraguru College of Technology

Rashmi R (19BCS067)
Divyanjali S (19BCS097)
Anju A (19BCS109)
Raghav M (19BCS119)

Under the guidance of,

Ms. Bharathipriya C (Mentor)
Ms. Lalitha Gayathri (Industry Mentor)
Mr. K.R.Baskaran (Evaluator)

TABLE OF CONTENTS

1.INTRODUCTION

Water Quality can be defined as the chemical, physical and biological characteristics of water, usually in respect to its suitability for a designated use. Water quality analysis is to measure the required parameters of water, following standard methods, to check whether they are in accordance with the standard. Water quality analysis is required mainly for monitoring purpose. The parameters of water quality are selected entirely according to the need for a specific use of that water.

1.1 PROJECT OVERVIEW

Water is considered as a vital resource that affects various aspects of human health and lives. The quality of water is a major concern for people living in urban areas. The quality of water serves as a powerful environmental determinant and a foundation for the prevention and control of waterborne diseases. However predicting the urban water quality is a challenging task since the water quality varies in urban spaces non-linearly and depends on multiple factors, such as meteorology, water usage patterns, and land uses, so this project aims at building a Machine Learning (ML) model to Predict Water Quality by considering all water quality standard indicators.

1.2.PURPOSE

Water quality monitoring is the premise to ensure safe water use, the quality of drinking water is directly related to people's health, so in every production link of domestic water must be strictly checked, do a good quality sampling survey, only the water can be delivered to the water supply network.

Water quality testing is an important part of environmental monitoring. When water quality is poor, it affects not only aquatic life but the surrounding ecosystem as well. These sections detail all of the parameters that affect the quality of water in the environment.

Water quality's importance is the manner in which it assures that end-users will remain healthy and well-functioning if proper standards are maintained. The end users may be people drinking healthily, industries operating without impediments caused by off-spec water, or natural environments thriving thanks to lack of pollution. Each user has a concentration threshold for the different contaminants, beyond which poorer quality water will have adverse effects.

Poor quality of potable, domestic use, or even recreational water due to contamination can lead to human illness. Drinking water contaminated with microbial organisms contributes heavily to the global burden of disease in the form of diarrhea, cholera, dysentery, hepatitis A, typhoid, and polio. According to the WHO, cholera affects 1.4 to 4 million people and accounts for 21,000 to 143,000 deaths globally every year. This map from the WHO shows countries where cholera was reported from 2010 to 2015.

Contamination of water sources by chemicals such as solvents, heavy metals, and pesticides poses human risk. Chronic exposure to heavy metals such as arsenic, chromium, lead, mercury, and cadmium can increase the risk of cancers of the blood, lung, liver, urinary bladder, and kidney.

Massive increases in population, the industrial revolution, and the use of fertilizers and pesticides have led to serious effects on the WQ environments . Thus, having models for the prediction of the WQ is of great help for monitoring water contamination.

2. LITERATURE SURVERY

1.EFFICIENT WATER QUALITY PREDICTION USING SUPERVISED MACHINE LEARNING

Umair Ahmed, Rafia Mumtaz , Hirra Anwar, Asad A. Shah, Rabia Irfan and José García-Nieto (School of Electrical Engineering and Computer Science (SEECS), National University of Sciences and Technology (NUST), Islamabad 44000, Pakistan; Department of Languages and Computer Sciences, Ada Byron Research Building University of Málaga)

PUBLISHED: 24 OCTOBER 2019

This research explores the methodologies that have been employed to help solve problems related to water quality. Typically, conventional lab analysis and statistical analysis are used in research to aid in determining water quality, while some analyses employ machine learning methodologies to assist in finding an optimized solution for the water quality problem. Local research employing lab analysis helped us gain a greater insight into the water quality problem in Pakistan. In one such research study, Daud et al. gathered water samples from different areas of Pakistan and tested them against different parameters using a manual lab analysis and found a high presence of E. coli and fecal coliform due to industrial and sewerage waste. Alamgir et al. tested 46 different samples from Orangi town, Karachi, using manual lab analysis and found them to be high in sulphates and total fecal coliform count. After getting familiar with the water quality research concerning Pakistan, we explored research employing machine learning methodologies in the realm of water quality. When it comes to estimating water quality using machine learning, Shafi et al. estimated water quality using classical machine learning algorithms namely, Support Vector Machines (SVM), Neural Networks (NN), Deep Neural Networks (Deep NN) and k Nearest Neighbors (kNN), with the highest accuracy of 93% with Deep NN. The estimated water quality in their work is based on only three parameters: turbidity, temperature and pH, which are tested according to World Health Organization (WHO) standards.

2. MACHINE LEARNING ALGORITHMS FOR EFFICIENT WATER QUALITY PREDICTION

Mourade Azrour(Université Moulay Ismail), Ghizlane Fattah(Mohammadia School of Engineers), Azidine Guezzaz(Cadi Ayyad University Morocco), Jamal

Mabrouki(Mohammed V University of Rabat)

PUBLISHED : 26 AUGUST 2021

In this study, the advantages of machine learning algorithms are used to develop a model that is capable of predicting the water quality index and then the water quality class. The method we propose is based on four water parameters: temperature, pH, turbidity and coliforms. The use of the multiple regression algorithms has proven to be important and effective in predicting the water quality index. In addition, the adoption of the artificial neural network provides the most highly efficient way to classify the water quality.

3.WATER QUALITY PREDICTION BASED ON MACHINE LEARNING TECHNIQUES

Zhao Fu(Bachelor of Information Security China University of Geosciences (Wuhan))

PUBLISHED : 8 JANUARY 2020

In this dissertation, several methods have been proposed to improve the performance of ANFIS-based water quality prediction models. Stratified sampling is employed to cover different kinds of data distribution in the training and testing datasets. The wavelet denoising technique is used to remove the noise hidden in the dataset. A deep prediction performance comparison between MLR, ANN, and ANFIS model is presented after stratified sampling and wavelet denoising techniques are applied. Because water quality data can be thought as a time series dataset, a time series analysis method is integrated with the ANFIS model to improve prediction performance. Lastly, intelligence algorithms are used to optimize the parameters of membership functions in the ANFIS model to promote the prediction accuracy. Experiments based on water quality datasets collected from Las Vegas Wash since 2007 and Boulder Basin of Lake Mead, Nevada, between 2011 and 2016 are used to evaluate the proposed models. Various ANN models have been designed to predict water and wastewater discharge quality based on previous existing datasets. A two-layer ANN model has been applied to predict the DO concentration in the Mathura River, and the experimental result showed that the ANN model worked well. An ANFIS model with eight input parameters is used to predict total phosphorus and total nitrogen, the experiment result based on 120 water samples. Time series analysis is also proposed to address dissolve oxygen prediction, and the experimental results show that the proposed analysis method can find out valuable knowledge from water quality historical timeseries data In this , MLR, ANN, ANFIS, and FTS models are integrated with statistical analysis, wavelet denoising, and intelligence algorithm to explore the prediction of water quality.

4.GROUND WATER QUALITY PREDICTION USING MACHINE LEARNING ALGORITHMS IN R

S.Vijay & 2Dr.K.Kamaraj Assistant Professor,Assistant ProfessorDepartment of Computer Science, 2Department of Computer Science Vivekanandha College of Arts and Sciences for Women, Tiruchengode, Namakkal Dt, SSM College of Arts and Science, Komarapalayam, Namakkal.

PUBLISHED : MARCH 2019

Sundarambal Palani proposed ANN models to predict water quality parameters whereas salinity, temperature, dissolved oxygen and Chl-a concentrations using continuous weekly measurements at different locations.Changjun Zhu proposed fuzzy neural network(FNN) model to evaluate and classify outer water quality in suzhou. The FNN model is reliable and effective and can deal with the problem of solitary elements which reflects the water quality at current stage.Yafra Khan has developed a water quality forecast model using the support of water quality components applying Artificial Neural Network (ANN) and time-series analysis with ANN-NAR. The performance measures such as Regression, Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) indicated the best prediction accuracy results with ANN-NAR time series algorithm.

5.WATER QUALITY ANALYSIS USING MACHINE LEARNING ALGORITHMS

Aleksei Shkurin(Bachelor’s Thesis in Environmental Engineering)

PUBLISHED : 29 MARCH 2016

Random forests (RF) show the best performance and are advised for scientists andengineers working with environmental data. Artificial neural networks (ANN) are another alternative, though their performance is inferior and they are prone to overfitting. Support vector machines (SVM) are the good example for the cases where a baseline model is needed, being one of the basic algorithms. K-nearest neighbours (KNN) model was successfully used for data imputation and is also suggested for this task for other researchers. Clusterization techniques, such as k-means clustering, may assist data scientist with possible algorithms to classify given data, for example defining good, average and bad conditions of the water based on various chemical, biological and physical parameter models generated during this research may be used by IT students for producing software meant to help environmental specialists in analysing collected water quality data.

6.GROUND WATER QUALITY PREDICTION USING MACHINE LEARNING ALGORITHMS IN R

S.Vijay(AssistantProfessor, Department of Computer Science, Vivekanandha College of Arts and Sciences for Women, Tiruchengode, Namakkal Dt) and Dr.K.Kamaraj (Assistant Professor, Department of Computer Science, SSM College of Arts and Science, Komarapalayam, Namakkal

PUBLISHED : 01 MARCH 2019

Water plays a dominant role in the growth of the country’s economy and essential for all the activities. The present study deals with the physico-chemical characteristics of ground water quality in Ranipet, Arcot, Walljah pet, towns in vellore district. Such a water samples were collected from different identified bore wells for the purpose of studying the quality of groundwater . The bore wells from which the samples were collected are extensively used for drinking purpose. The water quality parameters such as TDS, EC, Chloride, Sulphate, Nitrate, Carbonate, Bicarbonate, metal ions, trace elements have been estimated. There are two major classifications like High , Low level of water contamination observed in Vellore district. This paper focus on predicting water quality by using Machine Learning classifier algorithm C5.0, Naïve Bayes and Random forest as leaner for water quality prediction with high accuracy anf efficiency.

7.EFFICIENT PREDICTION OF WATER QUALITY INDEX (WQI) USING MACHINE LEARNING ALGORITHMS

Md. Mehedi Hassan (Computer Science and Engineering, North Western

University, Khulna, Bangladesh), Md. Mahedi Hassan(Computer Science and Engineering, Bangladesh University of Business and Technology, Dhaka, Bangladesh), Laboni Akter(Biomedical Engineering, Khulna University of Engineering & Technology, Khulna, Bangladesh), Md. Mushfiqur Rahman, Sadika Zaman, Khan Md. Hasib, Nusrat Jahan, Raisun Nasa Smrity, Jerin Farhana, M. Raihan, Swarnali Mollick

PUBLISHED : 22 November 2021

The quality of water has a direct influence on both human health and the environment. Water is utilized 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 categorize 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 normalization 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.

8.WATER QUALITY FACTOR PREDICTION USING SUPERVISED MACHINE LEARNINGMACHINE LEARNING

Kathleen Joslyn(Portland State University)

PUBLISHED : 2018

The objective of this research is to explore prediction accuracy of water quality factors with techniques and algorithms in machine learning consisting of a variation of support vector machines - Support Vector Regression (SVR) and the gradient boosting algorithm Extreme Gradient Boosting (XGBoost). Both the XGBoost and SVR algorithms were used to predict nine different factors with success rates ranging from 79% to 99%. Parameters of these algorithms were also explored to test the prediction accuracy levels of individual water quality factors. These parameters included normalizing the data, filling missing data points, and training and testing on a large set of data.

9.WATER QUALITY PREDICTION AND CLASSIFICATION BASED ON PRINCIPAL COMPONENT REGRESSION AND GRADIENT BOOSTING CLASSIFIER APPROACH

Md. SaikatIslam Khan(Department of Computer Science and Engineering, Santosh, Tangail-1902, Bangladesh) , NazrulIslam(Department of Information and Communication and Technology, Santosh, Tangail-1902, Bangladesh) , SifatulIslam Department of Technology Studies, Endicott College, Woosong University, Daejeon, South Korea , Mostofa KamalNasir(Mawlana Bhashani Science and Technology University, Santosh, Tangail 1902, Bangladesh)

PUBLISHED : 3 JUNE 2021

Estimating water quality has been one of the significant challenges faced by the world in recent decades. This paper presents a water quality prediction model utilizing the principal component regression technique. Firstly, the water quality index (WQI) is calculated using the weighted arithmetic index method. Secondly, the principal component analysis (PCA) is applied to the dataset, and the most dominant WQI parameters have been extracted. Thirdly, to predict the WQI, different regression algorithms are used to the PCA output. Finally, the Gradient Boosting Classifier is utilized to classify the water quality status. The proposed system is experimentally evaluated on a Gulshan Lake-related dataset. The results demonstrate 95% prediction accuracy for the principal component regression method and 100% classification accuracy for the Gradient Boosting Classifier method, which show credible performance compared with the state-of-art models.

10.PERFORMANCE OF MACHINE LEARNING METHODS IN PREDICTING WATER QUALITY INDEX BASED ON IRREGULAR DATA SET: APPLICATION ON ILLIZI REGION (Algerian southeast)

Saber Kouadri, Ahmed Elbeltagi, Abu Reza Md. Towfiqul Islam & Samir Kateb

PUBLISHED : 06 NOVEMBER 2021

Groundwater quality appraisal is one of the most crucial tasks to ensure safe drinking water sources. Concurrently, a water quality index (WQI) requires some water quality parameters. Conventionally, WQI computation consumes time and is often found with various errors during subindex calculation. To this end, 8 artificial intelligence algorithms, e.g., multilinear regression (MLR), random forest (RF), M5P tree (M5P), random subspace (RSS), additive regression (AR), artificial neural network (ANN), support vector regression (SVR), and locally weighted linear regression (LWLR), were employed to generate WQI prediction in Illizi region, southeast Algeria. Using the best subset regression, 12 different input combinations were developed and the strategy of work was based on two scenarios. The first scenario aims to reduce the time consumption in WQI computation, where all parameters were used as inputs. The second scenario intends to show the water quality variation in the critical cases when the necessary analyses are unavailable, whereas all inputs were reduced based on sensitivity analysis. The models were appraised using several statistical metrics including correlation coefficient (R), mean absolute error (MAE), root mean square error (RMSE), relative absolute error (RAE), and root relative square error (RRSE). The results reveal that TDS and TH are the key drivers influencing WQI in the study area. The comparison of performance evaluation metric shows that the MLR model has the higher accuracy compared to other models in the first scenario in terms of 1, 1.4572*10–08, 2.1418*10–08, 1.2573*10–10%, and 3.1708*10–08% for R, MAE, RMSE, RAE, and RRSE, respectively. The second scenario was executed with less error rate by using the RF model with 0.9984, 1.9942, 3.2488, 4.693, and 5.9642 for R, MAE, RMSE, RAE, and RRSE, respectively. The outcomes of this paper would be of interest to water planners in terms of WQI for improving sustainable management plans of groundwater resources.

2.1 EXISITNG PROBLEM

the main problem lies here. For testing the water quality we have to conduct lab tests on the water which is costly and time-consuming as well. 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 WHO(World Health Organisation). The data taken in this paper is taken from the PCPB India which includes 3277 examples of the distinct wellspring. In this paper, WQI(Water Quality Index) is calculated using AI techniques. So in future work, we can integrate this with IoT based framework to study large datasets and to expand our study to a larger scale. By using that it can predict the water quality fast and more accurately than any other IoT framework. That IoT framework system uses some limits for the sensor to check the parameters like ph, Temperature, Turbidity, and so on. And further after reading this parameter pass these readings to the Arduino microcontroller and ZigBee handset for further prediction

2.2 REFERENCES

1.EFFICIENT WATER QUALITY PREDICTION USING SUPERVISED MACHINE LEARNING
Umair Ahmed, Rafia Mumtaz , Hirra Anwar, Asad A. Shah, Rabia Irfan and José García-Nieto (School of Electrical Engineering and Computer Science (SEECS), National University of Sciences and Technology (NUST), Islamabad 44000, Pakistan; Department of Languages and Computer Sciences, Ada
Byron Research Building University of Málaga)
PUBLISHED: 24 OCTOBER 2019

2.MACHINE LEARNING ALGORITHMS FOR EFFICIENT WATER QUALITY PREDICTIONMourade Azrour(Université Moulay Ismail), Ghizlane Fattah(Mohammadia School of Engineers), Azidine Guezzaz(Cadi Ayyad University Morocco), Jamal Mabrouki(Mohammed V University of Rabat)
PUBLISHED : 26 AUGUST 2021
3.WATER QUALITY PREDICTION BASED ON MACHINE LEARNING TECHNIQUES
Zhao Fu(Bachelor of Information Security China University of Geosciences (Wuhan)); PUBLISHED : 8 JANUARY 2020

4. GROUND WATER QUALITY PREDICTION USING MACHINE LEARNING ALGORITHMS IN R
S.Vijay & 2Dr.K.Kamaraj Assistant Professor,Assistant ProfessorDepartment of Computer Science, 2Department of Computer Science Vivekanandha College of Arts and Sciences for Women, Tiruchengode, Namakkal Dt, SSM College of Arts and Science, Komarapalayam, Namakkal. PUBLISHED : MARCH 2019

5. WATER QUALITY ANALYSIS USING MACHINE LEARNING ALGORITHMS
Aleksei Shkurin(Bachelor's Thesis in Environmental Engineering)
PUBLISHED : 29 MARCH 2016

6. GROUND WATER QUALITY PREDICTION USING MACHINE LEARNING ALGORITHMS IN R
S.Vijay(AssistantProfessor, Department of Computer Science, Vivekanandha College of Arts and Sciences for Women, Tiruchengode, Namakkal Dt) and

Dr.K.Kamaraj (Assistant Professor, Department of Computer Science, SSM College of Arts and Science, Komarapalayam, Namakkal
PUBLISHED : 01 MARCH 2019

7. EFFICIENT PREDICTION OF WATER QUALITY INDEX (WQI) USING MACHINE LEARNING ALGORITHMS
Md. Mehedi Hassan (Computer Science and Engineering, North Western University, Khulna, Bangladesh), Md. Mahedi Hassan(Computer Science and Engineering, Bangladesh University of Business and Technology, Dhaka, Bangladesh), Laboni Akter(Biomedical Engineering, Khulna University of Engineering & Technology, Khulna, Bangladesh), Md. Mushfiqur Rahman, Sadika Zaman, Khan Md. Hasib, Nusrat Jahan, Raisun Nasa Smrity, Jerin Farhana, M. Raihan, Swarnali Mollick
PUBLISHED : 22 November 2021

8. WATER QUALITY PREDICTION AND CLASSIFICATION BASED ON PRINCIPAL COMPONENT REGRESSION AND GRADIENT BOOSTING CLASSIFIER APPROACH
Md. SaikatIslam Khan(Department of Computer Science and Engineering, Santosh, Tangail-1902, Bangladesh) , NazrulIslam(Department of Information and Communication and Technology, Santosh, Tangail-1902, Bangladesh) , SifatulIslam Department of Technology Studies, Endicott College, Woosong University, Daejeon, South Korea , Mostofa KamalNasir(Mawlana Bhashani Science and Technology University, Santosh, Tangail 1902, Bangladesh). PUBLISHED : 3 JUNE 2021

9.WATER QUALITY FACTOR PREDICTION USING SUPERVISED MACHINE LEARNINGMACHINE LEARNING
Kathleen Joslyn(Portland State University)
PUBLISHED : 2018

10. PERFORMANCE OF MACHINE LEARNING METHODS IN PREDICTING WATER QUALITY INDEX BASED ON IRREGULAR DATA SET: APPLICATION ON ILLIZI REGION (Algerian southeast)
Saber Kouadri, Ahmed Elbeltagi, Abu Reza Md. Towfiqul Islam & Samir Kateb
PUBLISHED : 06 NOVEMBER 2021

2.3 PROBLEM STATEMENT DEFINITION

To predict the water safe or not for Access to safe drinking-water is essential to hea lth, a basic human right and a component of effective policy for health protection. This is important as a health and development issue at a national, regional and loca l level. In some regions, it has been shown that investments in water supply and sa nitation can yield a net economic benefit, since the reductions in adverse health eff ects and health care costs outweigh the costs of undertaking the interventions.

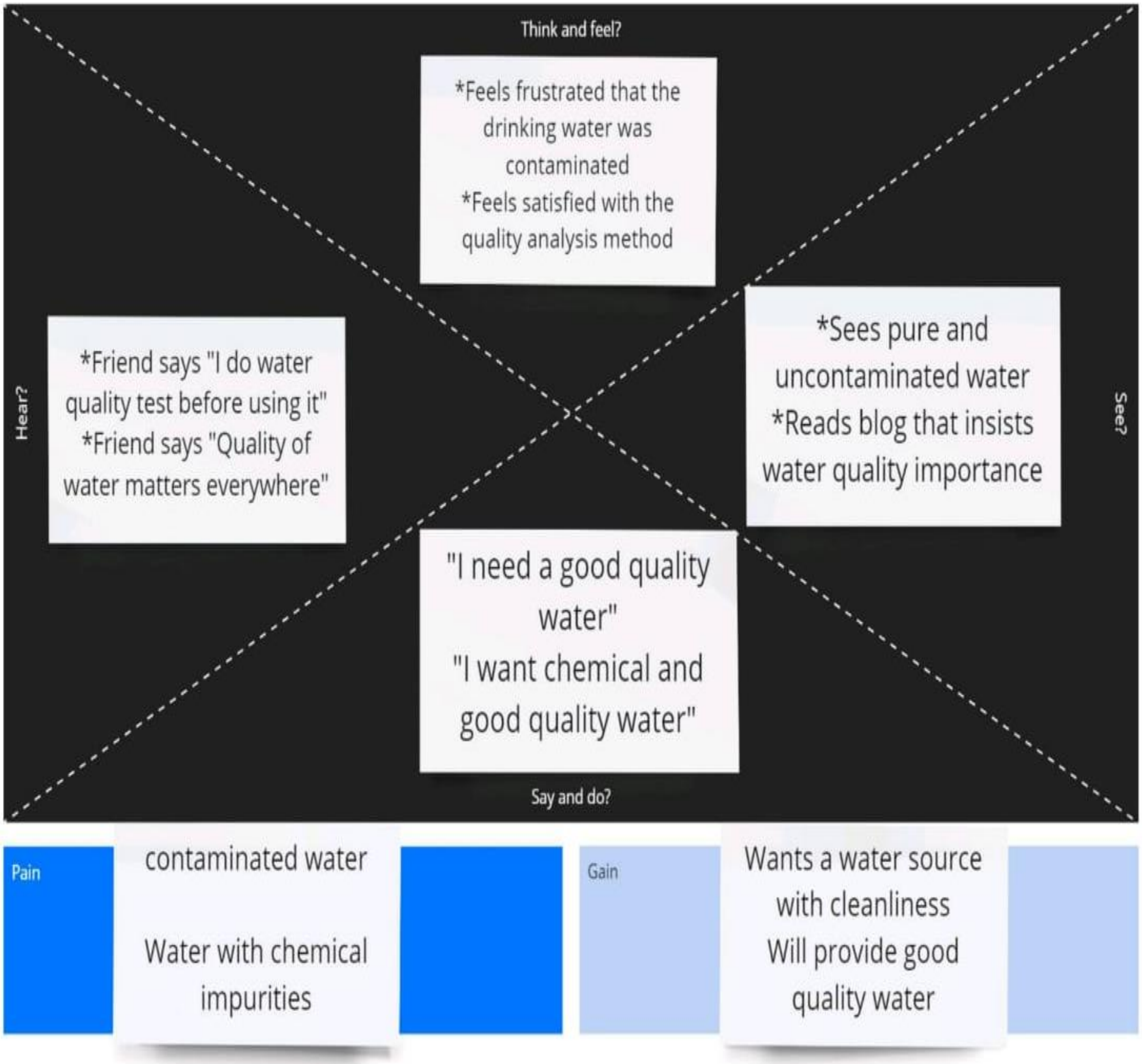
3.IDEATION AND PROPOSED SOLUTION

3.1 EMPATHY MAP CANVAS

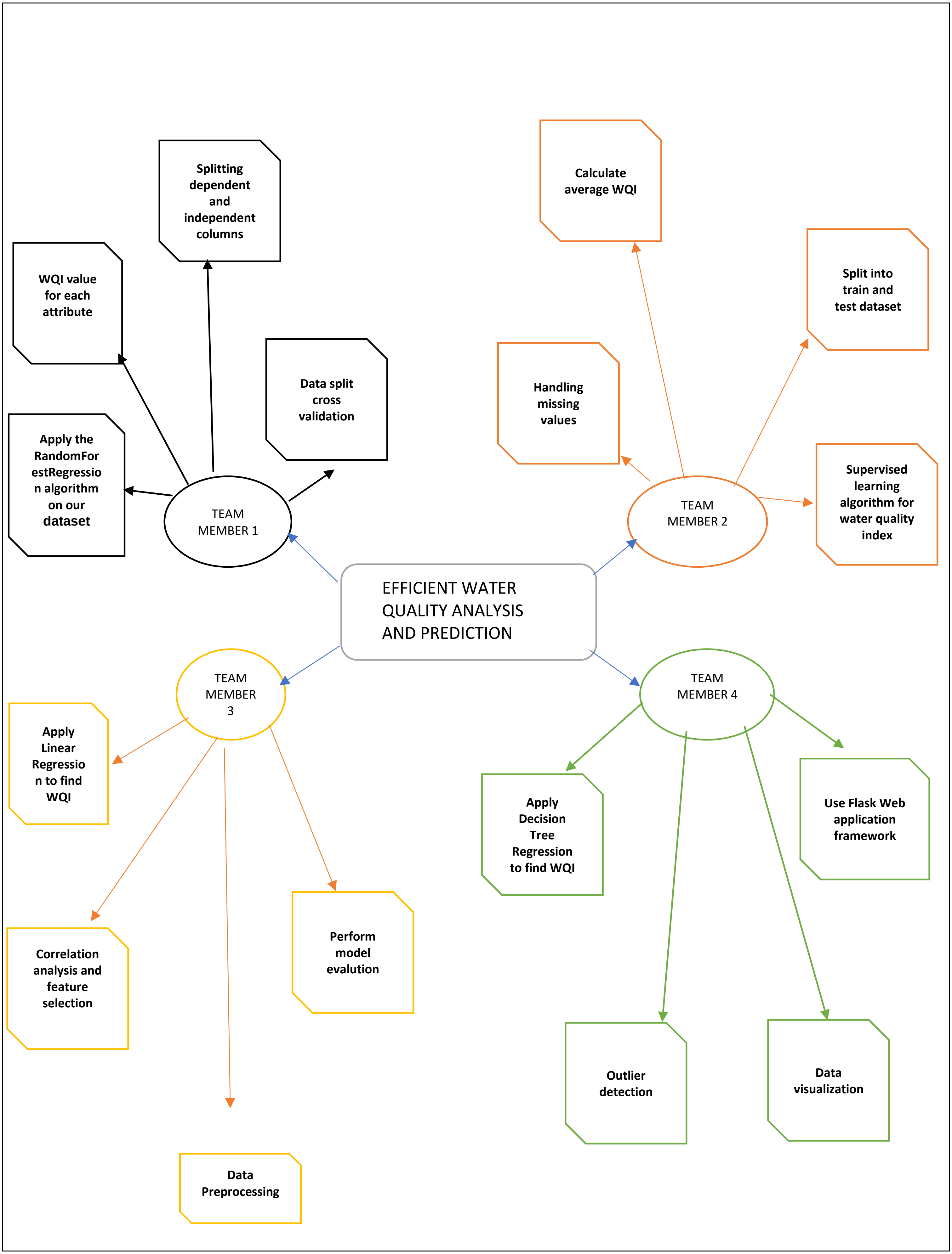
Empathy Map Canvas:
An empathy map is a simple, easy-to-digest visual that captures knowledge about a user's behaviours and attitudes.

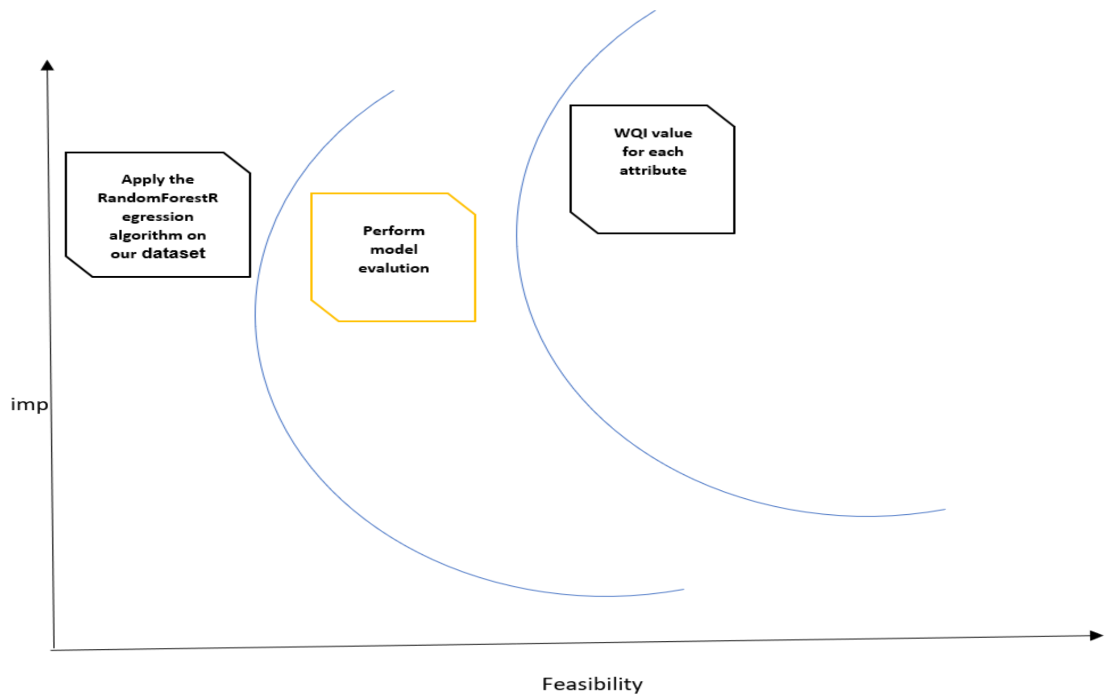
It is a useful tool to helps teams better understand their users.
Creating an effective solution requires understanding the true problem and the person who is experiencing it. The exercise of creating the map helps participants consider things from the user's perspective along with his or her goals and challenges.

EMPATHY MAP



3.2 IDEATION AND BRAINSTORMING

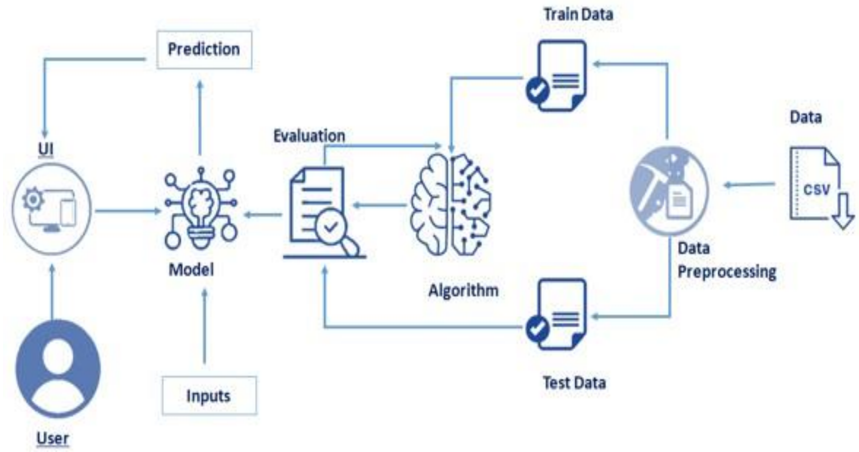

EFFICIENT WATER QUALITY ANALYSIS AND PREDICTION
TEAM MEMBER 1
Splitting dependent and independent columns
WQI value for each attribute
Data split cross validation
Apply the RandomForestRegression algorithm on our dataset
TEAM MEMBER 2
Calculate average WQI
Split into train and test dataset
Handling missing values
Supervised learning algorithm for water quality index
TEAM MEMBER 3
Apply Linear Regression to find WQI
Correlation analysis and feature selection
Perform model evalution
Data Preprocessing
TEAM MEMBER 4
Apply Decision Tree Regression to find WQI
Use Flask Web application framework
Outlier detection
Data visualization



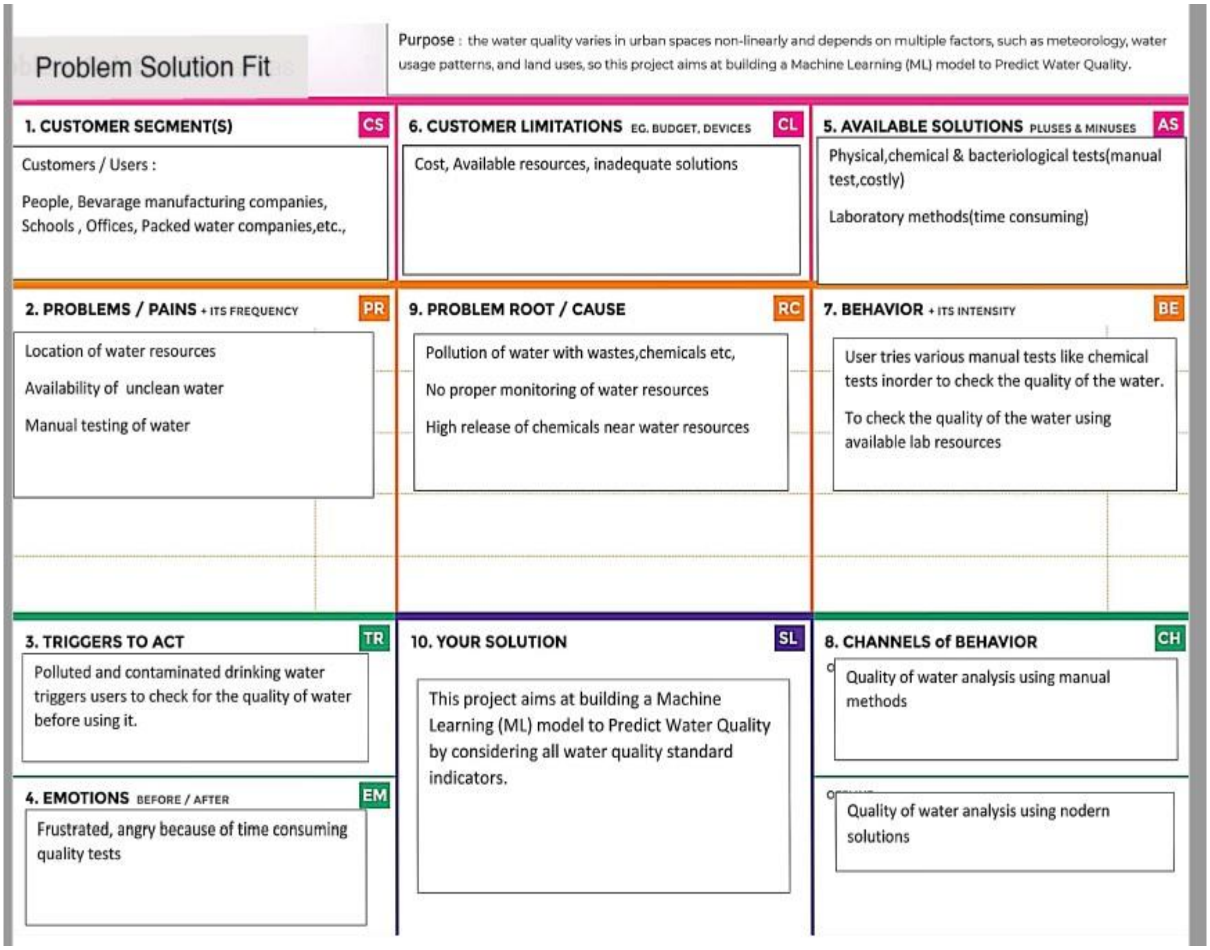
Apply the RandomForestRegression algorithm on our dataset
Perform model evalution
WQI value for each attribute
imp
Feasibility

3.3 PROPOSED SOLUTION

S.No.	Parameter	Description
1.	Problem Statement (Problem to be solved)	Efficient Water Quality Analysis & Prediction Using Machine Learning Water is considered as a vital resource that affects various aspects of human health and lives. The quality of water is a major concern for people living in urban areas. The quality of water serves as a powerful environmental determinant and a foundation for the prevention and control of waterborne diseases. However predicting the urban water quality is a challenging task since the water quality varies in urban spaces non-linearly and depends on multiple factors, such as meteorology, water usage patterns, and land uses, so this project aims at building a Machine Learning (ML) model to Predict Water Quality by considering all water quality standard indicators.
2.	Idea / Solution description	You will need to train the datasets to run smoothly and see an incremental improvement in the prediction rate. Different measures are used to assess the accuracy of the applied machine learning algorithms.
3.	Novelty / Uniqueness	Since the results from various algorithms are compared we can identify which method gives the highest accuracy.
4.	Social Impact / Customer Satisfaction	Assessing customer satisfaction is critical for companies looking to gain competitive advantage in the market. Therefore, customer satisfaction is tied to the standards that apply to different areas of activity, including water quality analysis. The present work aims to assess the satisfaction of the customers based on parameters that influence the management requirements and the technical requirements of the ISO/IEC 17025 standard.

5.	Business Model (Revenue Model)	
6.	Scalability of the Solution	Reliable water quality prediction can improve environmental flow monitoring and the sustainability of the stream ecosystem.

3.3 PROBLEM SOLUTION FIT



Problem Solution Fit

Purpose : the water quality varies in urban spaces non-linearly and depends on multiple factors, such as meteorology, water usage patterns, and land uses, so this project aims at building a Machine Learning (ML) model to Predict Water Quality.

1. CUSTOMER SEGMENT(S) CS

Customers / Users :

People, Bevarage manufacturing companies, Schools , Offices, Packed water companies,etc.,

6. CUSTOMER LIMITATIONS EG. BUDGET, DEVICES CL

Cost, Available resources, inadequate solutions

5. AVAILABLE SOLUTIONS PLUSES & MINUSES AS

Physical,chemical & bacteriological tests(manual test,costly)

Laboratory methods(time consuming)

2. PROBLEMS / PAINS + ITS FREQUENCY PR

Location of water resources

Availability of unclean water

Manual testing of water

9. PROBLEM ROOT / CAUSE RC

Pollution of water with wastes,chemicals etc,

No proper monitoring of water resources

High release of chemicals near water resources

7. BEHAVIOR + ITS INTENSITY BE

User tries various manual tests like chemical tests inorder to check the quality of the water.

To check the quality of the water using available lab resources

3. TRIGGERS TO ACT TR

Polluted and contaminated drinking water triggers users to check for the quality of water before using it.

4. EMOTIONS BEFORE / AFTER EM

Frustrated, angry because of time consuming quality tests

10. YOUR SOLUTION SL

This project aims at building a Machine Learning (ML) model to Predict Water Quality by considering all water quality standard indicators.

8. CHANNELS of BEHAVIOR CH

Quality of water analysis using manual methods

Quality of water analysis using nodern solutions

4. REQUIREMENT ANALYSIS

4.1 FUNCTIONAL REQUIREMENT

Following are the functional requirements of the proposed solution.

FR No.	**Functional Requirement (Epic)**	**Sub Requirement (Story / Sub-Task)**
FR-1	Install Google colab and download libraries.	Install google collab and Import all the required libraries which are used to train the model or visualise the data.
FR-2	Dataset	Initial process 1.Download and import the dataset 2.Read the dataset
FR-3	WQI(water quality index)	The outcome to be found from the dataset, 1.calculate the quality index for each column. 2. Calculate the avg of WQI.
FR-4	Application Building	Use flask architecture which is used to create a user interface . 1.It accepts the individual inputs**(year,D.O, P.H,** **B.O.D, C.O, N,A, T.C)** and inturn produce the WQI as output
FR-5	Interface sensor	Confirmation via email and it is predicted by water level sensor

4.2 NON-FUNCTIONAL REQUIREMENT

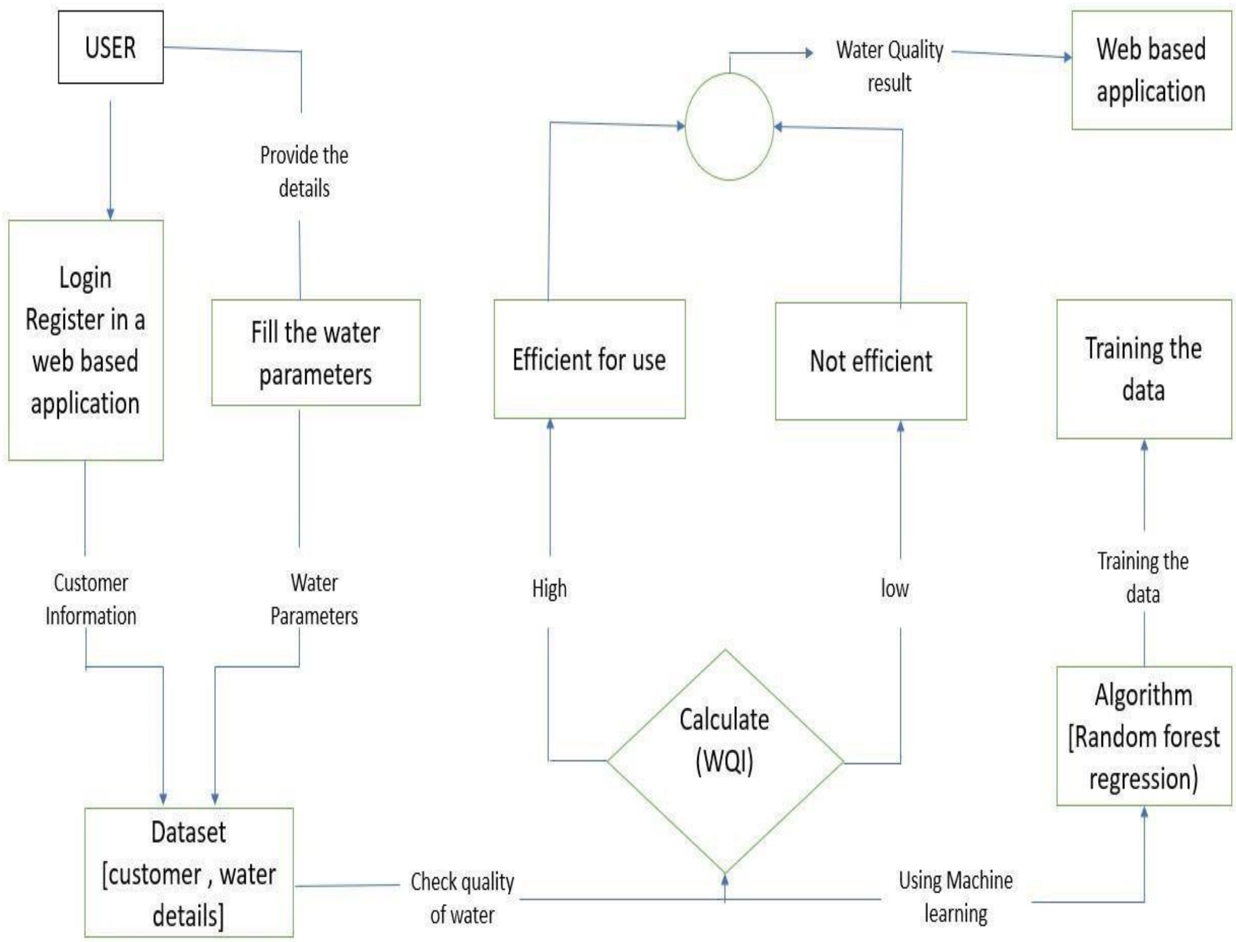
Following are the non-functional requirements of the proposed solution.

NFR-1	**Usability**	The aim of this model is to predict the WQI(water quality index) based on some factors like(PH, B.O.D,Conductivity etc..). WQI helps in determining overall water quality status. Accurate water quality prediction is the basis of water environment management and is of great significance for water environment protection.
NFR-2	**Security**	It provides secured feel for the people while drinking water
NFR-3	**Reliability**	This project helps in protecting people health and the environment.
NFR-4	**Performance**	PH, Turbidity, temperature etc are calculated by sensors and recorded; the data is pre processed and WQI is calculated
NFR-5	**Availability**	By developing and deploying the software we can analyze the drinking water
NFR-6	**Scalability**	The project helps in providing a purified water and pollution free water

5.PROJECT DESIGN

5.1 DATA FLOW DIAGRAM

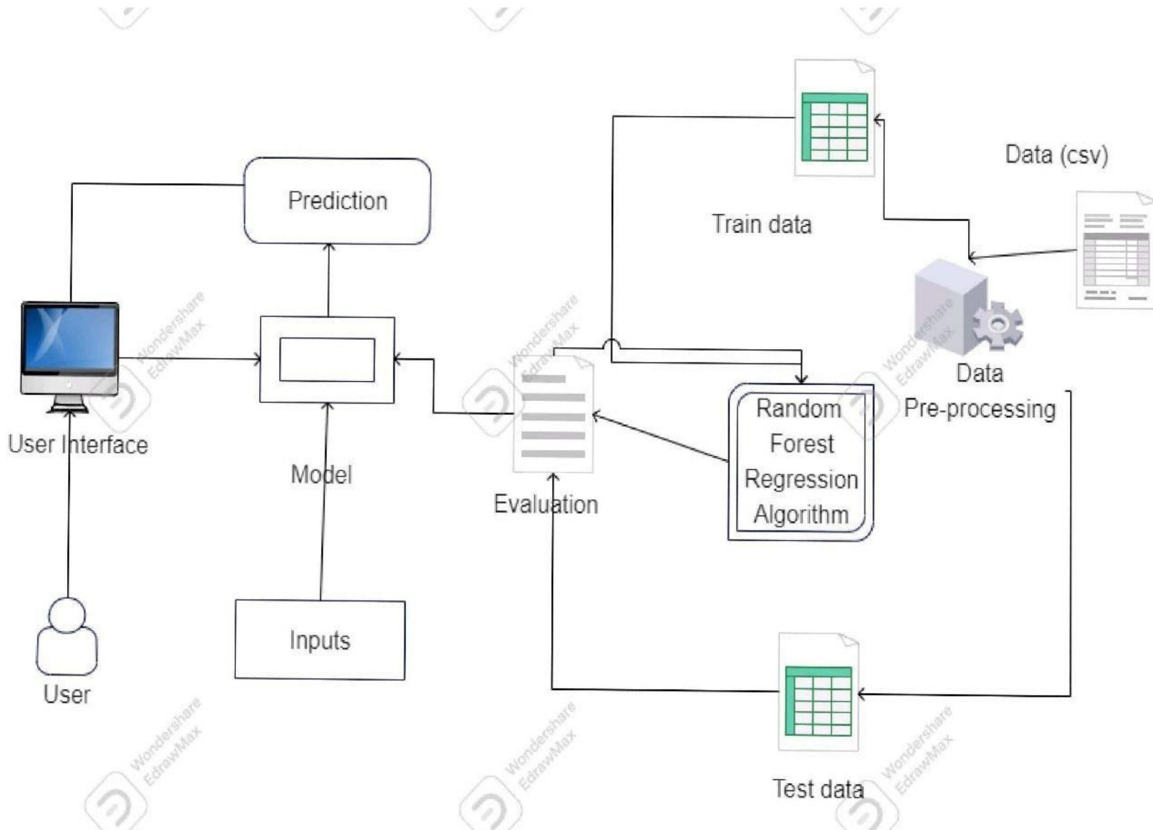
A Data Flow Diagram (DFD) is a traditional visual representation of the information flows within a system. A neat and clear DFD can depict the right amount of the system requirement graphically. It shows how data enters and leaves the system, what changes the information, and where data is stored.



5.2 SOLUTION AND TECHNINCAL ARCHITECTURE

Solution architecture is a complex process – with many sub-processes – that bridges the gap between business problems and technology solutions. Its goals are to:

- Find the best tech solution to solve existing business problems.
- Describe the structure, characteristics, behavior, and other aspects of the software to project stakeholders.
- Define features, development phases, and solution requirements.
- Provide specifications according to which the solution is defined, managed, and delivered.



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 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		Medium	Sprint-1

			Gmail			
	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 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 correct coefficient of water.		Medium	Sprint-3
	Data preprocessing	USN-10	As a user, I can see the loading information.		Medium	Sprint-3
	User Input Evaluation	USN-11	I can see the evaluation		High	Sprint-4

			quickly.			
	Prediction	USN-12	As a user, I can see the result of the water efficient.	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 webpage.	It results in user interaction.	Medium	Sprint-5
Administrator		USN-14	I can manage the application.		Medium	Sprint-5

6. PROJECT PLANNING AND SCDULING

6.1 SPRINT PLANNING AND ESTIMATION

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint 1	Dataset collection	USN-1	Collect the required data for the	3	High	Anju A, Raghav M

			water quality prediction			
Sprint 1	Data preprocessing	USN-2	Perform data cleaning to optimize the dataset	7	Medium	Anju A, Raghav M
Sprint 2	Training & Building Model , Model evaluation	USN-3	Build the model using regression algorithms to classify the data	10	High	Rashmi R,Divyanjali S
Sprint 3	Application Building	USN-5	Build the html and python code. Run flak app.	7	High	Rashmi R,Divyanjali S
		USN-6	Run Flask app.	3	Medium	Rashmi R,Divyanjali S
Sprint 4	Implementation of the Application	USN-7	Deploy the model on IBM cloud	10	Medium	Anju A, Raghav M

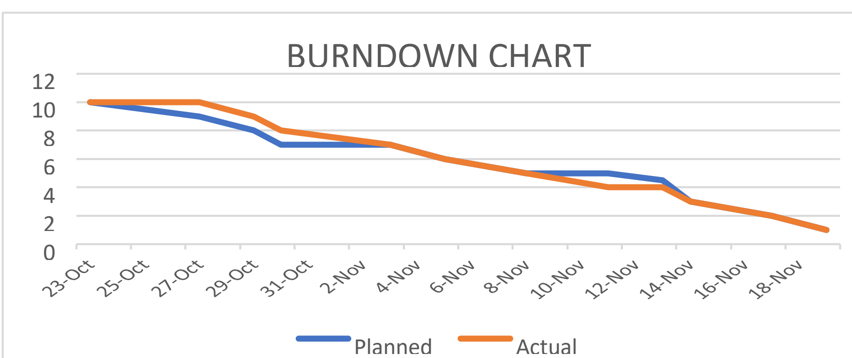
6.2 Sprint Delivery Schedule

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	10	7 Days	23 Oct 2022	29 Oct 2022	8	29 Oct 2022
Sprint-2	10	7 Days	30 Oct 2022	05 Nov 2022	8	06 Nov 2022
Sprint-3	10	7 Days	07 Nov 2022	13Nov 2022	7	13 Nov 2022
Sprint-4	10	7Days	13 Nov 2022	19 Nov 2022	10	19 Nov 2022

Imagine we have a 7 -day sprint duration, and the velocity of the team is 10 (points per sprint). Let's calculate the team's average velocity (AV) per iteration unit (story points per day).

$$AV = \frac{sprint\ duration}{velocity} \quad \frac{7}{10} = 0.7$$

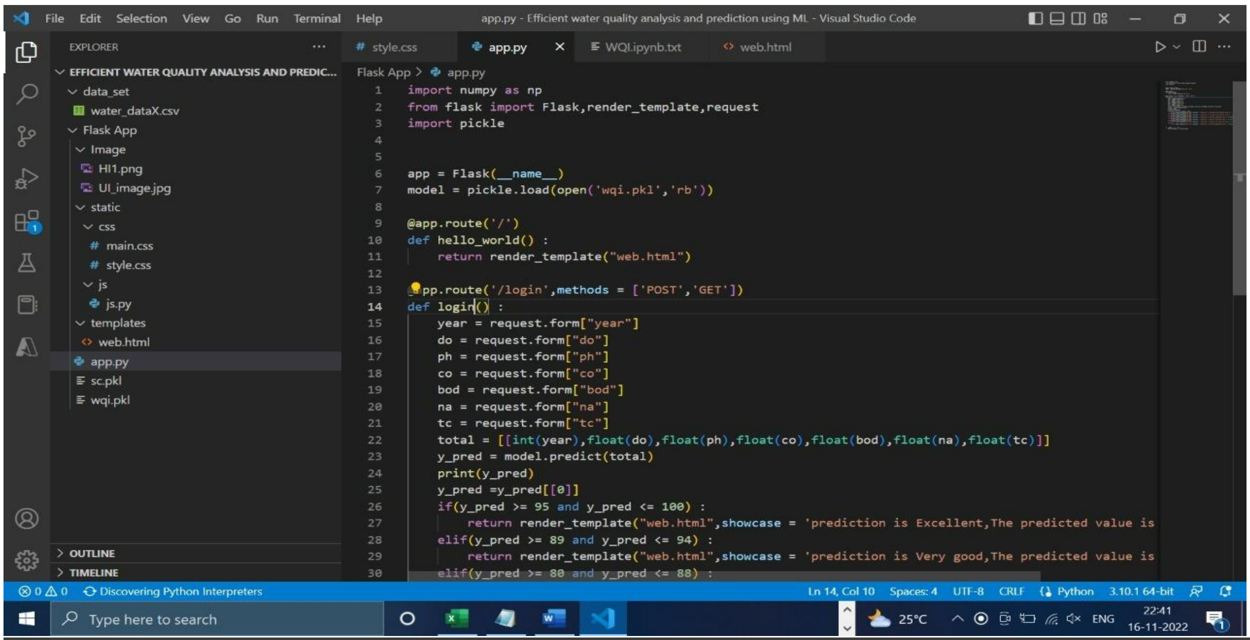
Burndown Chart:

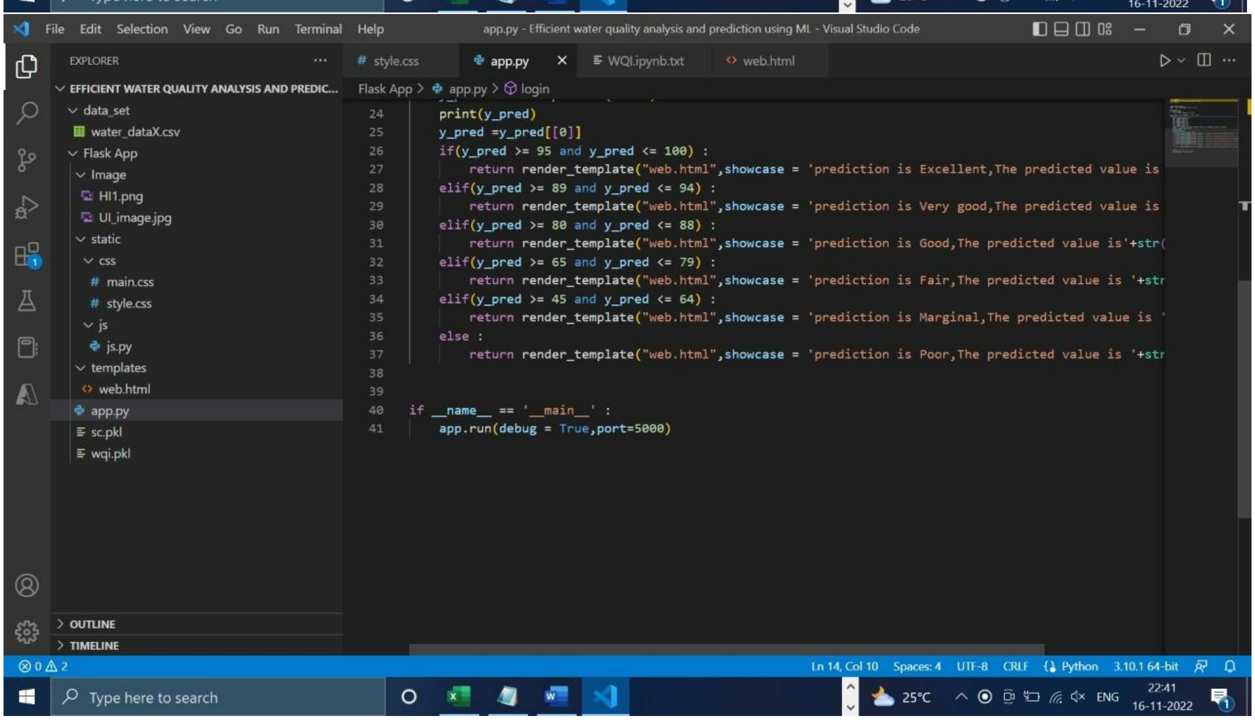


7. CODING AND SOLUTIONS

7.1 FEATURE 1

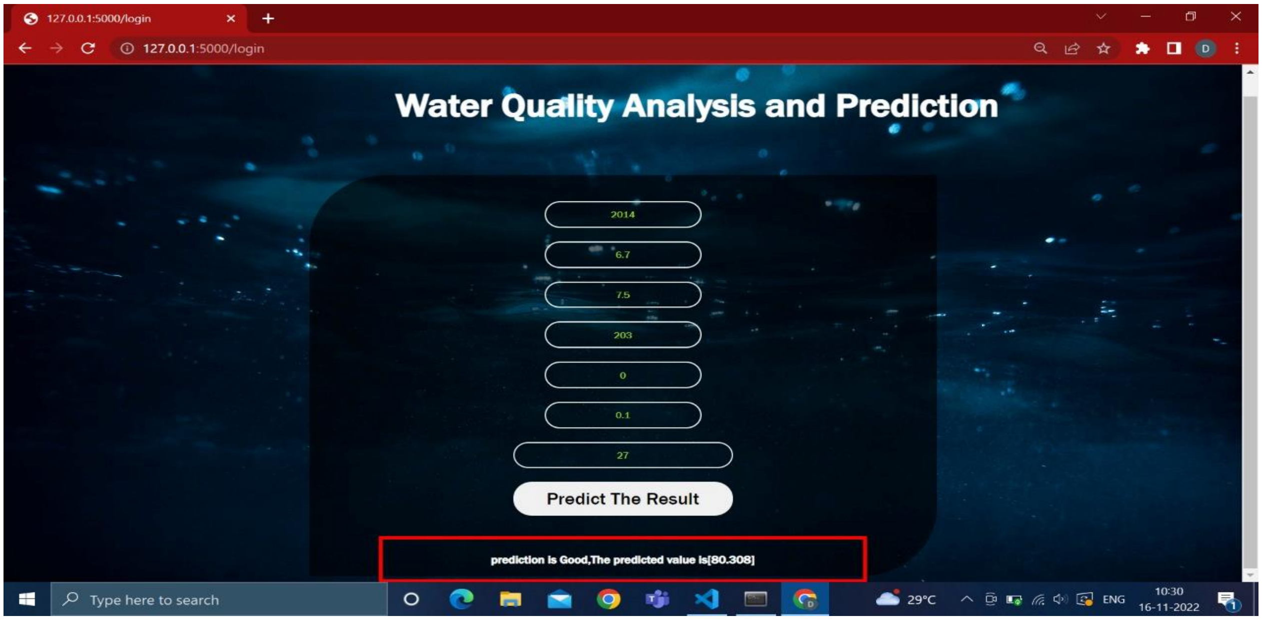
Flask is a web framework. This means flask provides you with tools, libraries and technologies that allow you to build a web application. This web application can be some web pages, a blog, a wiki or go as big as a web-based calendar application or a commercial website.



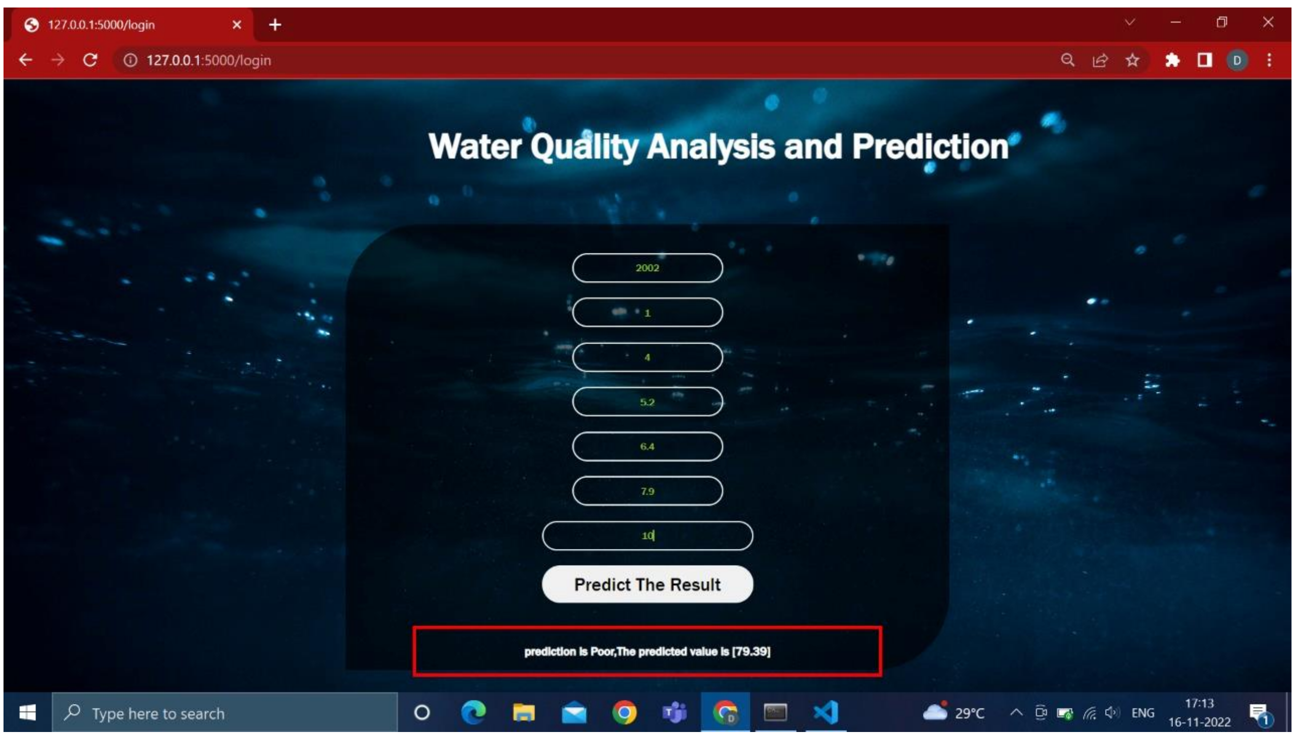


8.TESTING

8.1 TEST CASES 1



TEST CASE 2



8.2 USER ACCEPTANCE TESTING

1. Purpose of Document

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

2. 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	Subtotal
By Design	2	4	2	8
Duplicate	0	0	0	0
Fixed	2	2	0	4
Not Reproduced	0	0	0	0
Skipped	0	0	1	1
Won't Fix	0	0	0	0
Totals	4	6	1	13

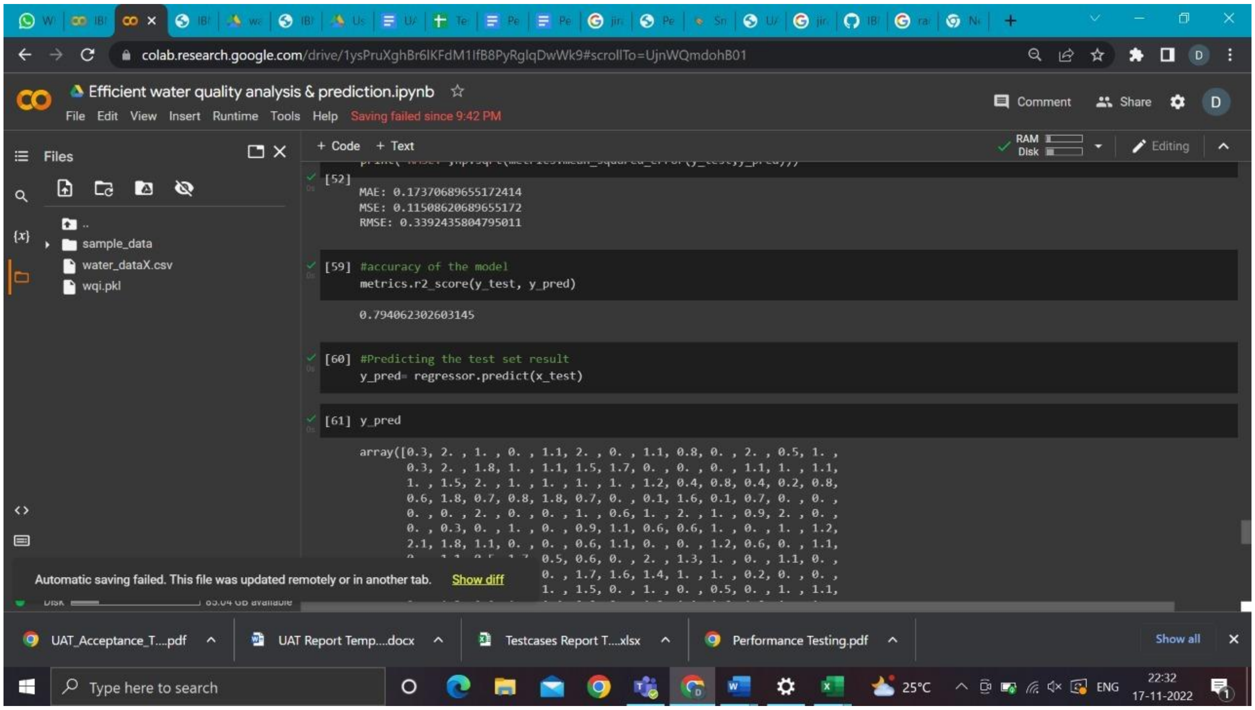
3. Test Case Analysis

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

Verify user is able to input valid inputs	5	0	0	5
Verify user is able to predit output for valid inputs	5	0	0	5
Verify user is able to get ouput with InValid input	5	0	0	5
Verify user is able to get ouput without any input	5	0	0	5

9.RESULT

9.1 PERFORMANCE METRICS



Testing : 25%
Training : 75%

10. 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. If you’re wanting to create a solid foundation on which to build a broader water management plan, then investing in water quality testing should be your first point of action. 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.

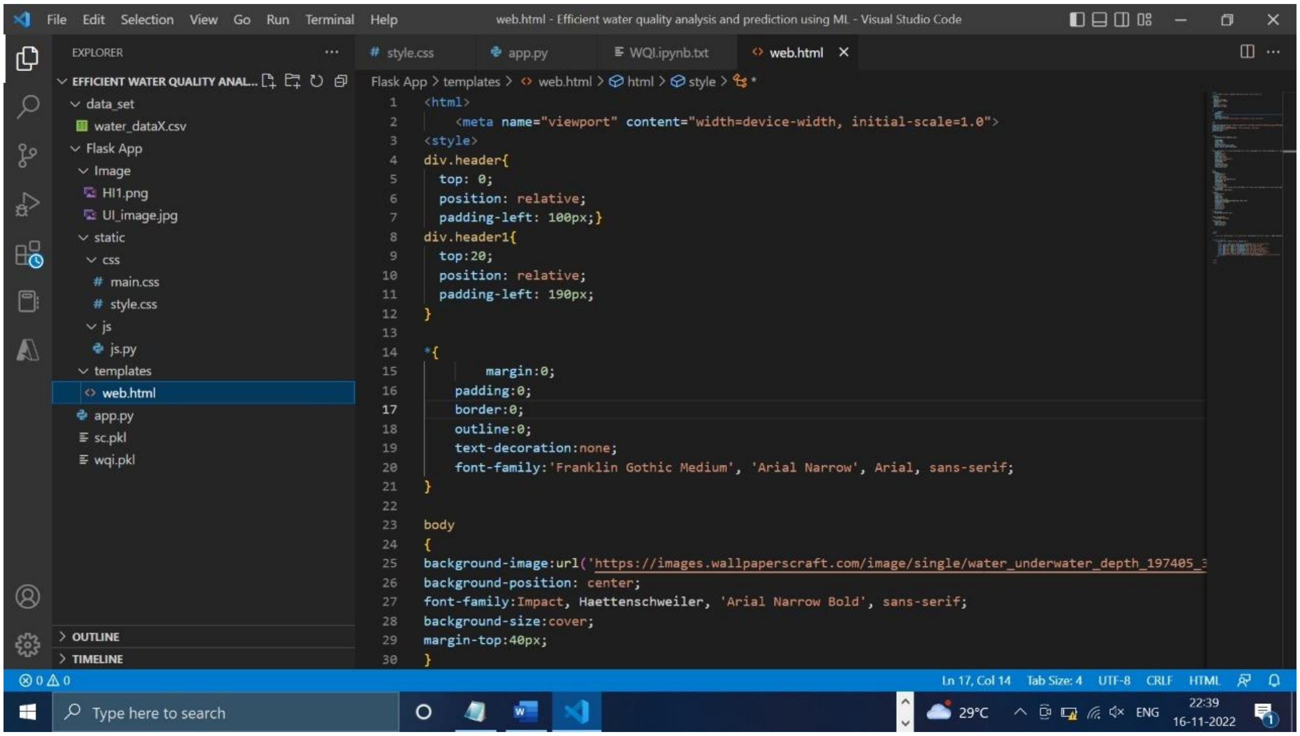
DISADVANTAGES

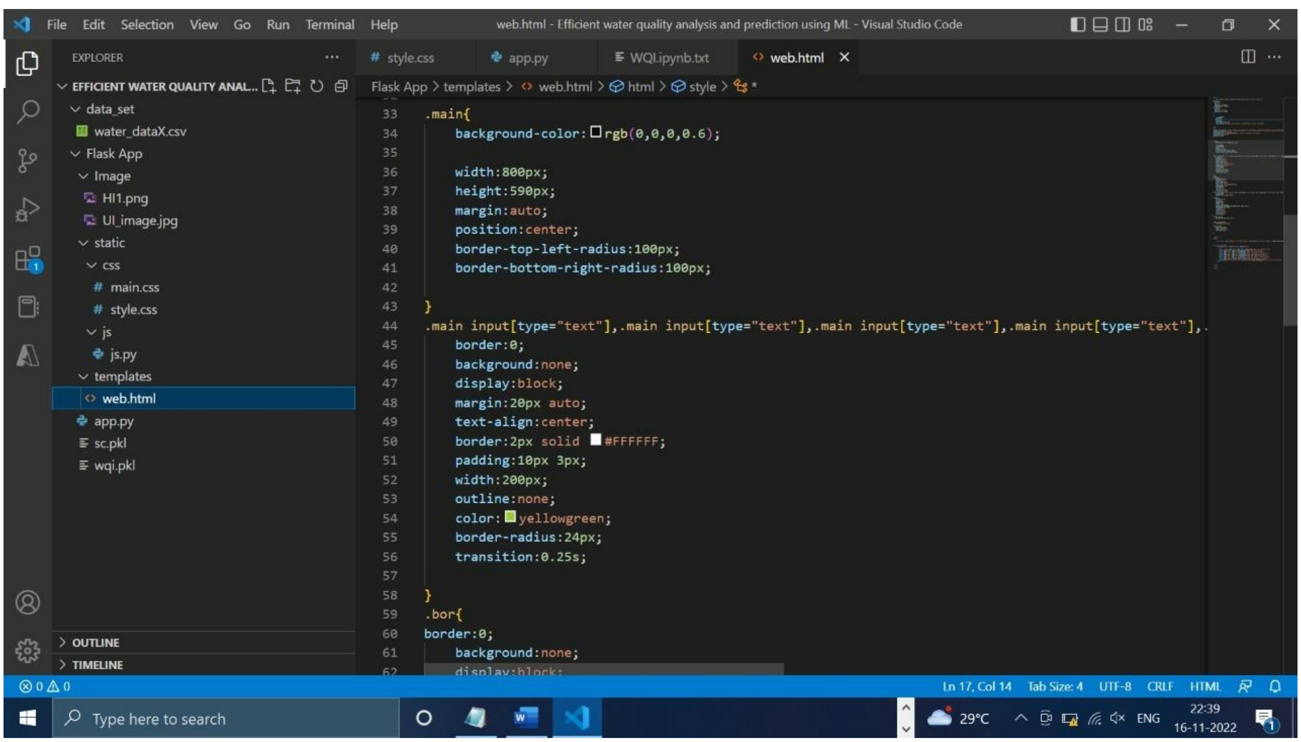
Training necessary Somewhat difficult to manage over time and with large data sets
Requires manual operation to submit data, some configuration required
Costly, usually only feasible under Exchange Network grants Technical expertise and network server required
Requires manual operation to submit data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

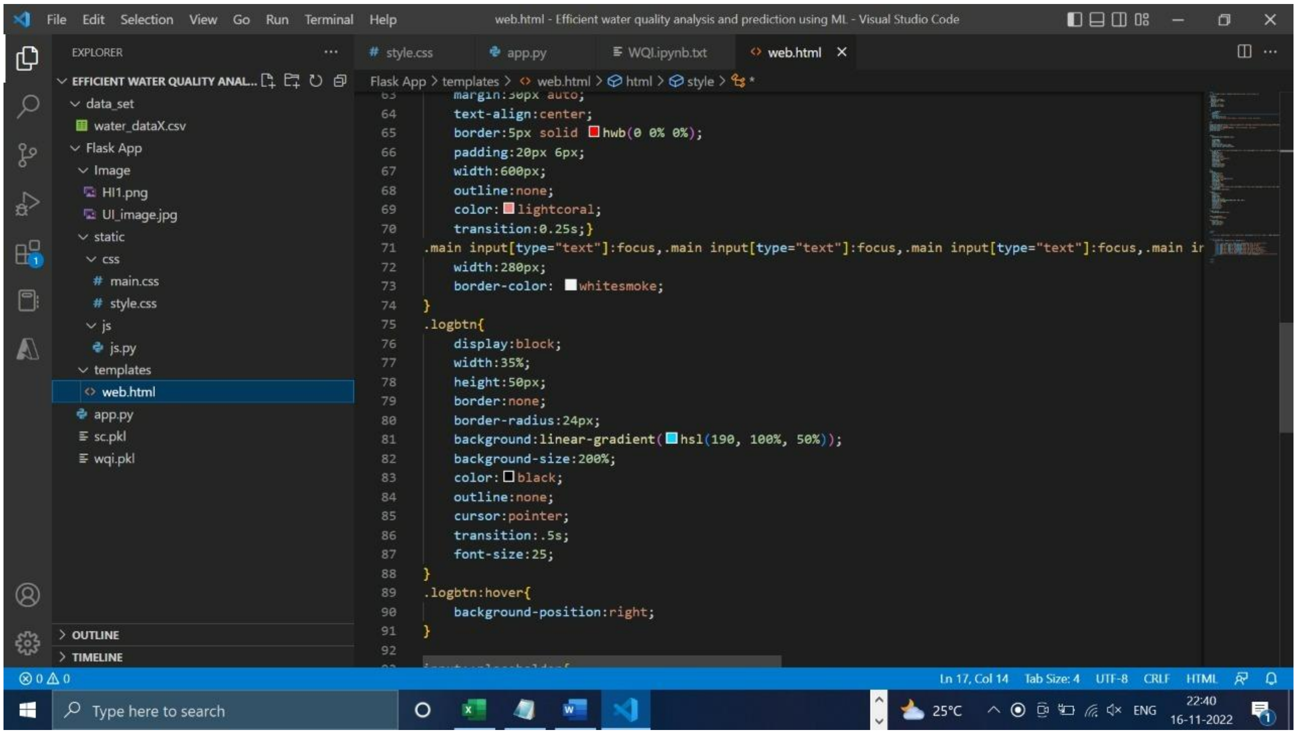
11. CONCLUSION

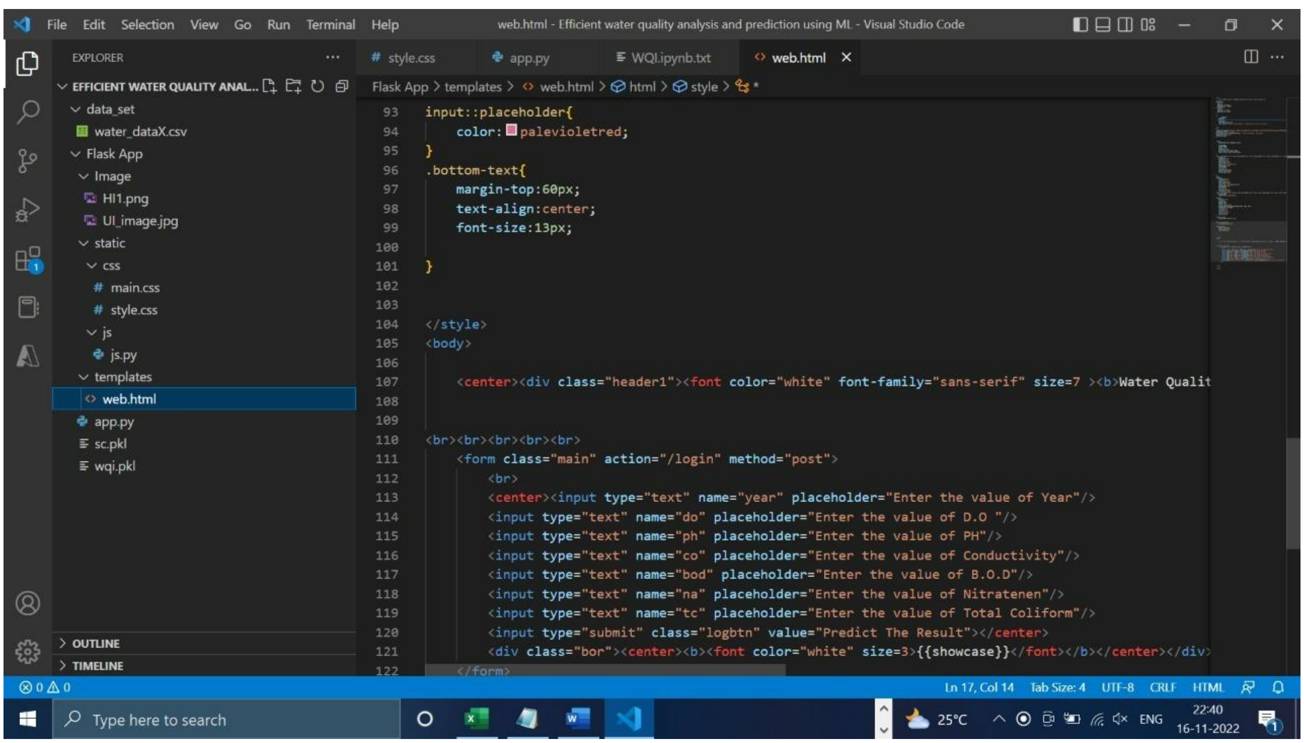
WQI determines the quality of water, which is one of the most important resources for existence. Traditionally, testing water quality required an expensive and time-consuming lab analysis. This study looked into an alternative machine learning method for predicting water quality using only a few simple water quality criteria. To estimate, a set of representative supervised machine learning algorithms was used. It would detect water of bad quality before it was released for consumption and notify the appropriate authorities It will hopefully reduce the number of individuals who drink low-quality water, lowering the risk of diseases like typhoid and diarrhea. In this case, using a prescriptive analysis based on projected values would result in future capabilities to assist decision and policy makers.

12.SOURCE CODE


web.html - Efficient water quality analysis and prediction using ML - Visual Studio Code
<html>
<meta name="viewport" content="width=device-width, initial-scale=1.0">
<style>
div.header{
top: 0;
position: relative;
padding-left: 100px;}
div.header1{
top:20;
position: relative;
padding-left: 190px;
}
*{
margin:0;
padding:0;
border:0;
outline:0;
text-decoration:none;
font-family:'Franklin Gothic Medium', 'Arial Narrow', Arial, sans-serif;
}
body
{
background-image:url('https://images.wallpaperscraft.com/image/single/water_underwater_depth_197405_
background-position: center;
font-family:Impact, Haettenschweiler, 'Arial Narrow Bold', sans-serif;
background-size:cover;
margin-top:40px;
}

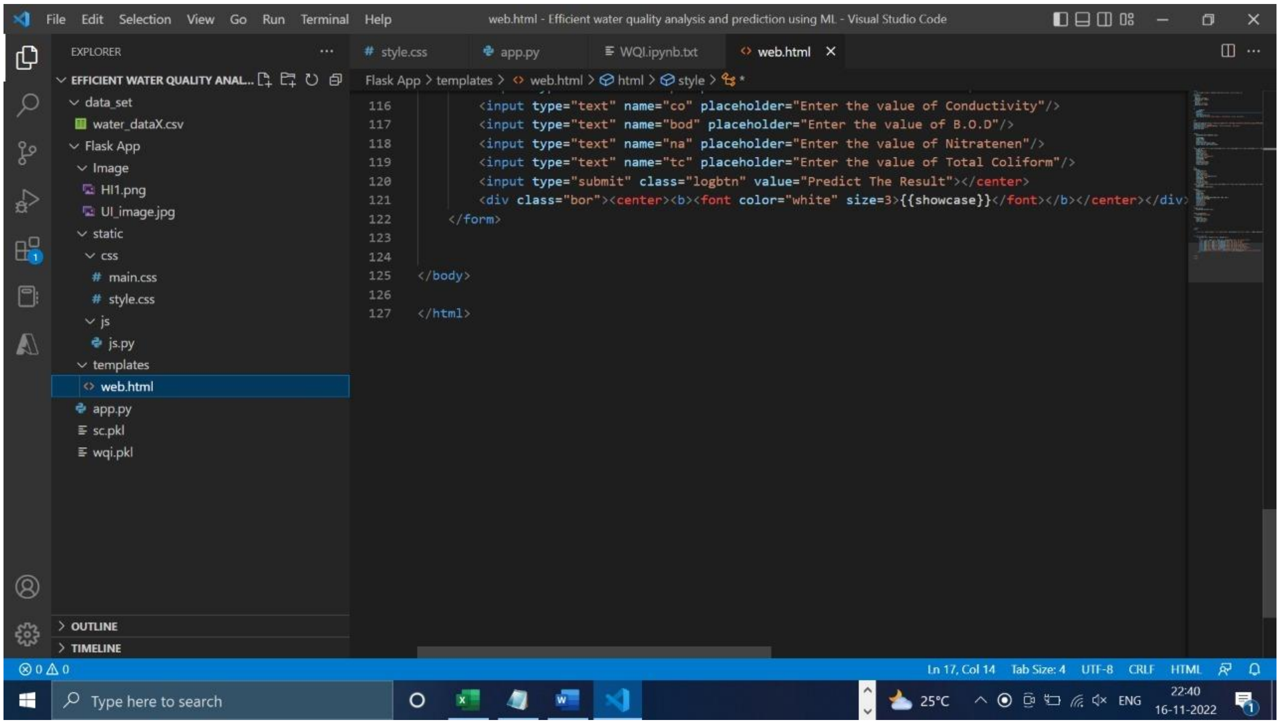

web.html - Efficient water quality analysis and prediction using ML - Visual Studio Code
.main{
background-color:rgb(0,0,0,0.6);
width:800px;
height:590px;
margin:auto;
position:center;
border-top-left-radius:100px;
border-bottom-right-radius:100px;
}
.main input[type="text"],.main input[type="text"],.main input[type="text"],.main input[type="text"],.
border:0;
background:none;
display:block;
margin:20px auto;
text-align:center;
border:2px solid #FFFFFF;
padding:10px 3px;
width:200px;
outline:none;
color:yellowgreen;
border-radius:24px;
transition:0.25s;
}
.bor{
border:0;
background:none;

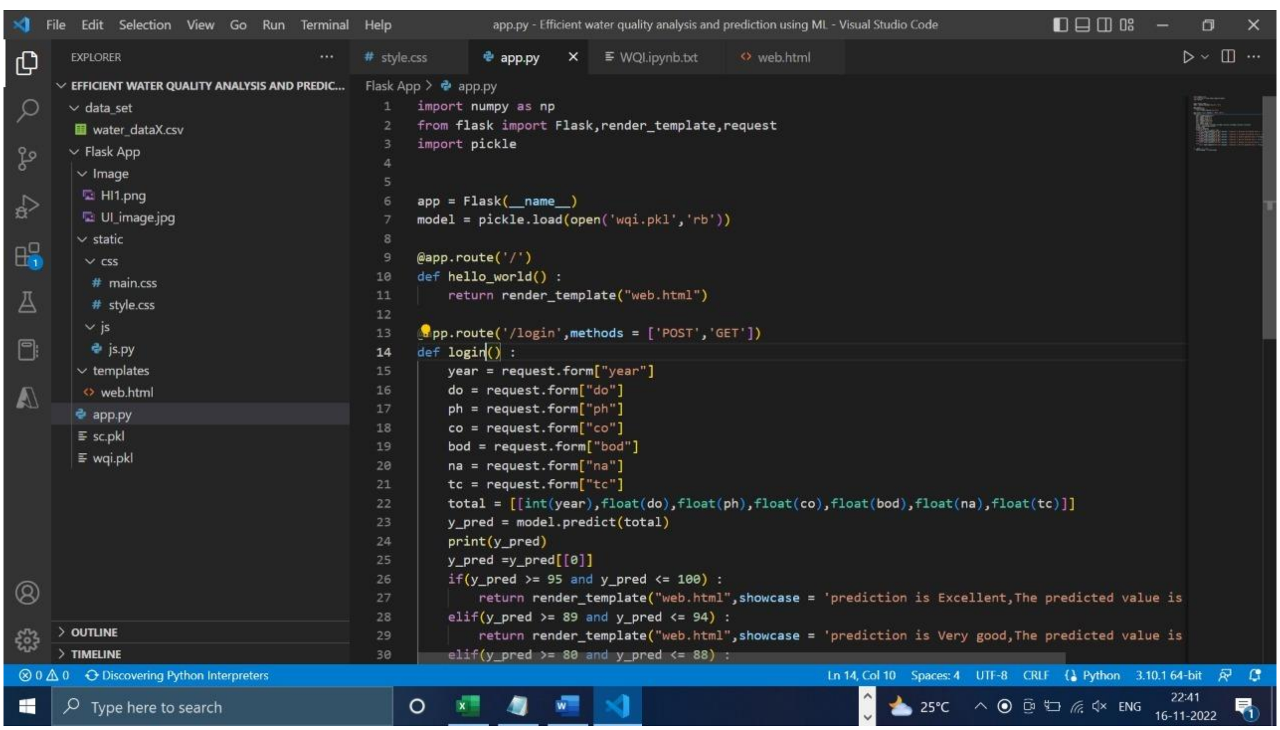

web.html - Efficient water quality analysis and prediction using ML - Visual Studio Code
EXPLORER
EFFICIENT WATER QUALITY ANAL...
data_set
water_dataX.csv
Flask App
Image
HI1.png
UI_image.jpg
static
css
main.css
style.css
js
js.py
templates
web.html
app.py
sc.pkl
wqi.pkl
OUTLINE
TIMELINE
style.css
app.py
WQI.ipynb.txt
web.html
Flask App > templates > web.html > html > style
text-align:center;
border:5px solid hwb(0 0% 0%);
padding:20px 6px;
width:600px;
outline:none;
color:lightcoral;
transition:0.25s;}
.main input[type="text"]:focus,.main input[type="text"]:focus,.main input[type="text"]:focus,.main in
width:280px;
border-color: whitesmoke;
}
.logbtn{
display:block;
width:35%;
height:50px;
border:none;
border-radius:24px;
background:linear-gradient(hsl(190, 100%, 50%));
background-size:200%;
color:black;
outline:none;
cursor:pointer;
transition:.5s;
font-size:25;
}
.logbtn:hover{
background-position:right;
}
Ln 17, Col 14 Tab Size: 4 UTF-8 CRLF HTML
Type here to search
25°C
ENG
22:40
16-11-2022

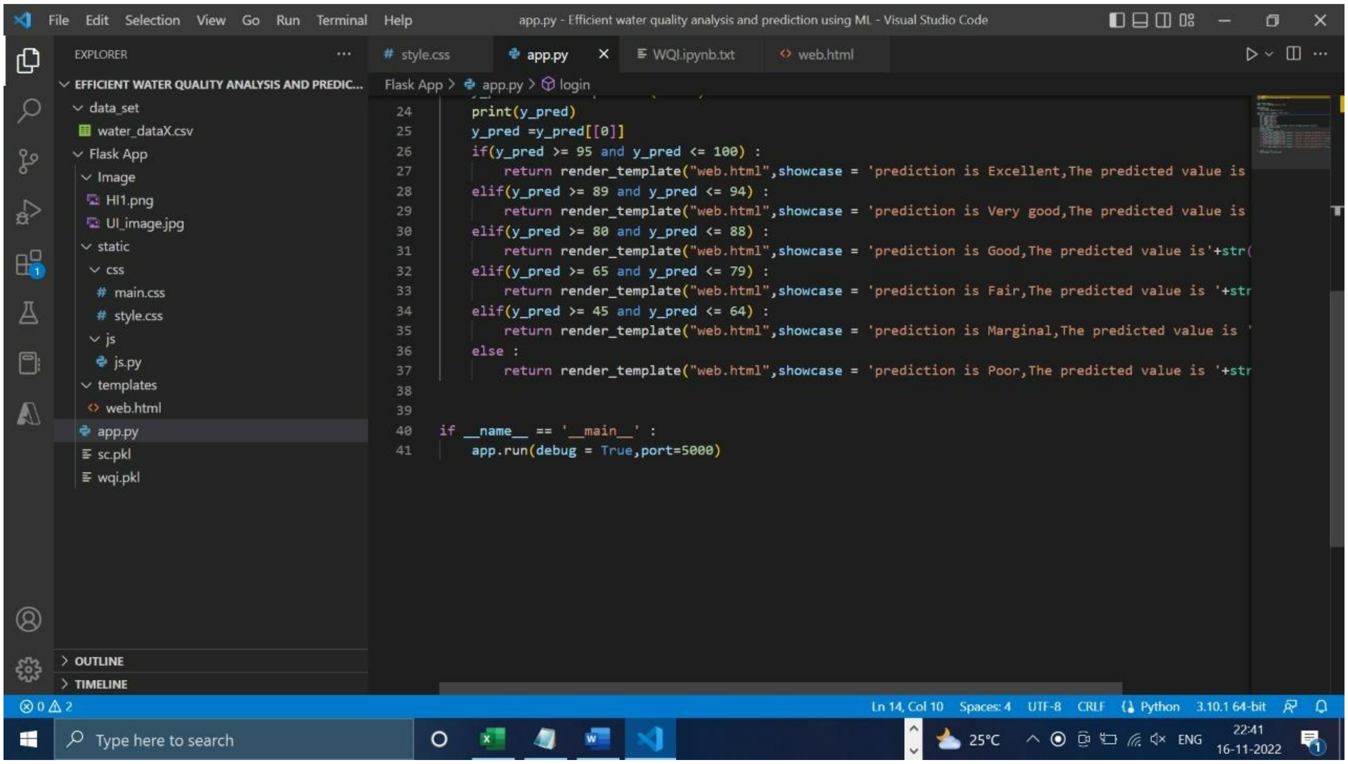

web.html - Efficient water quality analysis and prediction using ML - Visual Studio Code
EXPLORER
EFFICIENT WATER QUALITY ANAL...
data_set
water_dataX.csv
Flask App
Image
HI1.png
UI_image.jpg
static
css
main.css
style.css
js
js.py
templates
web.html
app.py
sc.pkl
wqi.pkl
OUTLINE
TIMELINE
style.css
app.py
WQI.ipynb.txt
web.html
Flask App > templates > web.html > html > style
input::placeholder{
color:palevioletred;
}
.bottom-text{
margin-top:60px;
text-align:center;
font-size:13px;
}
</style>
<body>
<center><div class="header1">Water Qualit

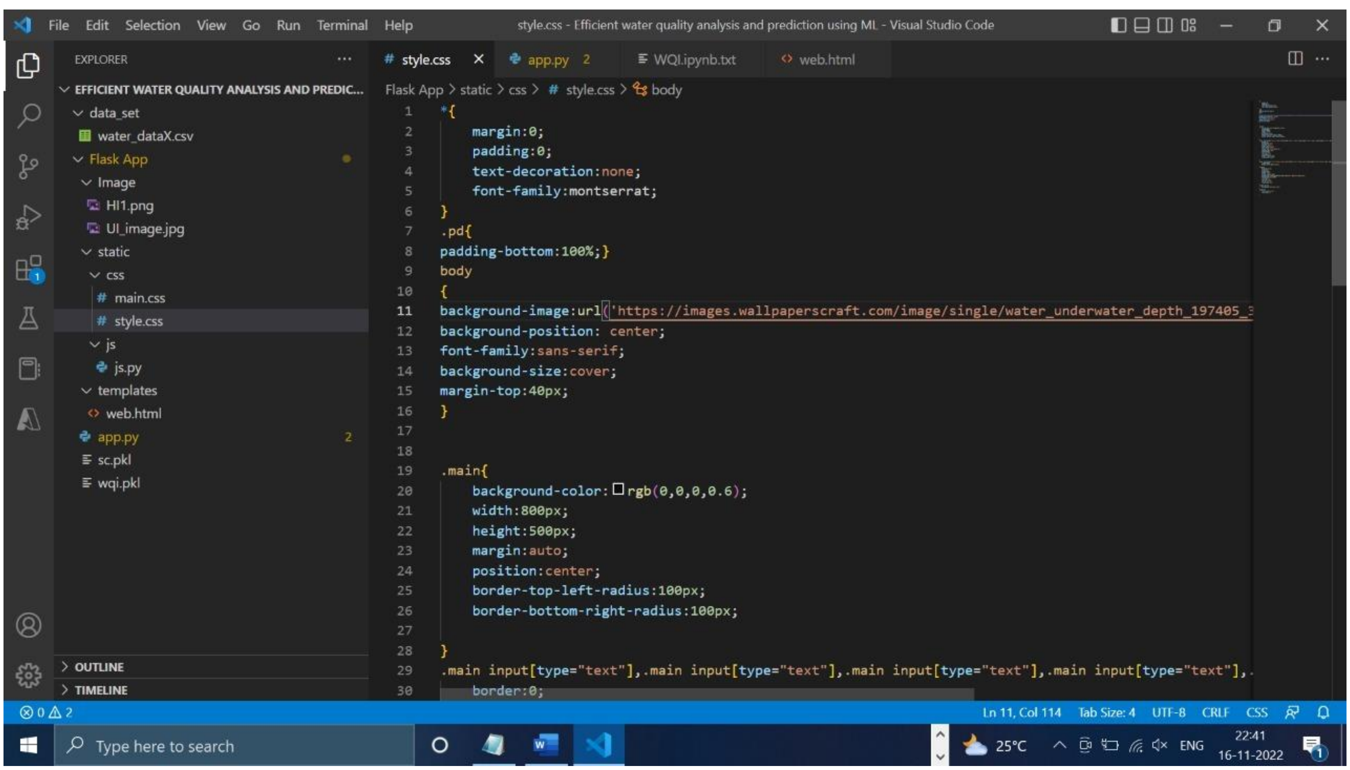
<form class="main" action="/login" method="post">

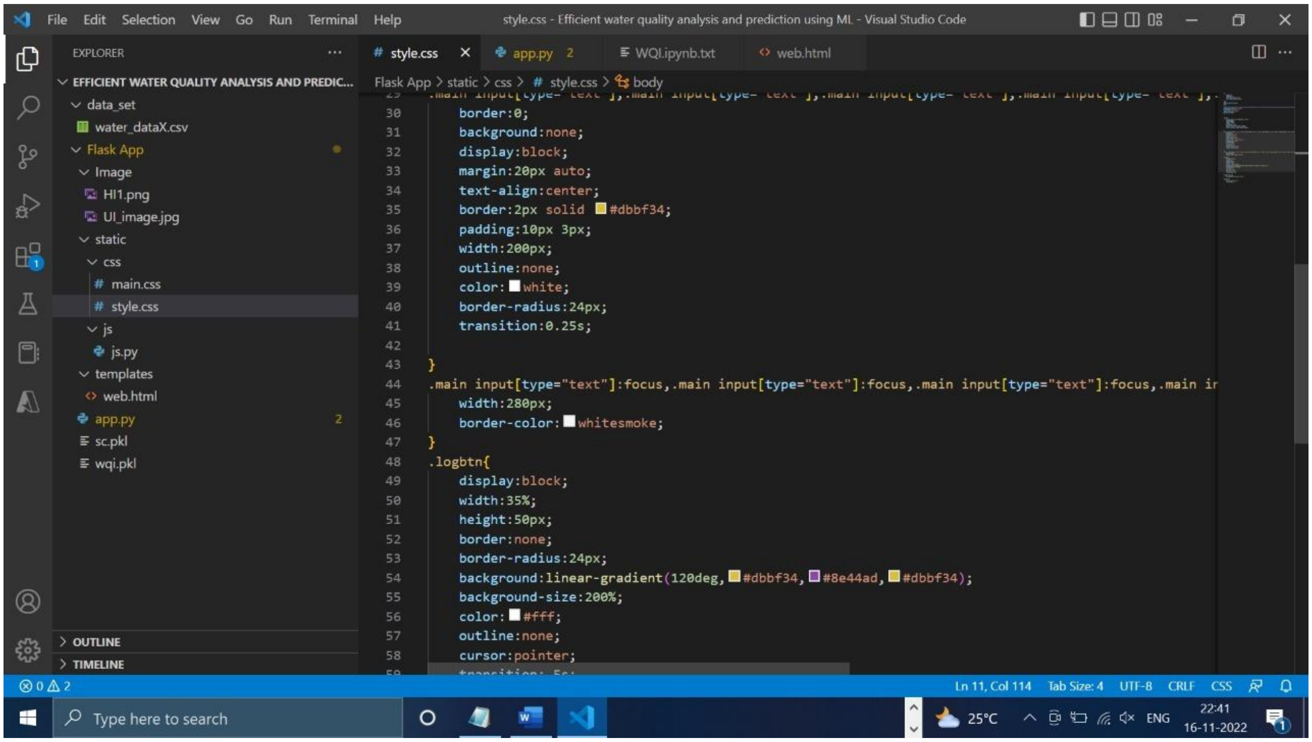
<center><input type="text" name="year" placeholder="Enter the value of Year"/>
<input type="text" name="do" placeholder="Enter the value of D.O "/>
<input type="text" name="ph" placeholder="Enter the value of PH"/>
<input type="text" name="co" placeholder="Enter the value of Conductivity"/>
<input type="text" name="bod" placeholder="Enter the value of B.O.D"/>
<input type="text" name="na" placeholder="Enter the value of Nitratenen"/>
<input type="text" name="tc" placeholder="Enter the value of Total Coliform"/>
<input type="submit" class="logbtn" value="Predict The Result"></center>
<div class="bor"><center>{{showcase}}</center></div>
Ln 17, Col 14 Tab Size: 4 UTF-8 CRLF HTML
Type here to search
25°C
ENG
22:40
16-11-2022

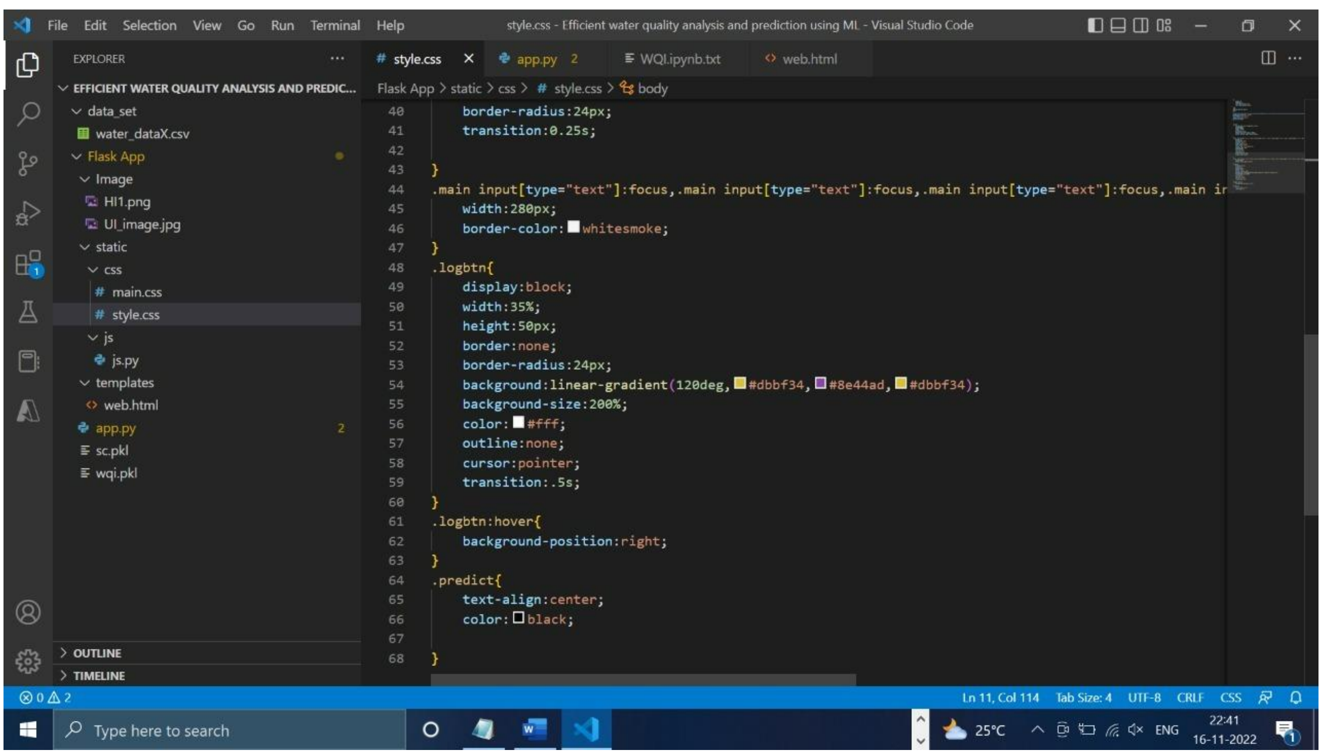

web.html - Efficient water quality analysis and prediction using ML - Visual Studio Code
EXPLORER
EFFICIENT WATER QUALITY ANAL...
data_set
water_dataX.csv
Flask App
Image
HI1.png
UI_image.jpg
static
css
main.css
style.css
js
js.py
templates
web.html
app.py
sc.pkl
wqi.pkl
OUTLINE
TIMELINE
style.css
app.py
WQI.ipynb.txt
web.html
Flask App > templates > web.html > html > style
<input type="text" name="co" placeholder="Enter the value of Conductivity"/>
<input type="text" name="bod" placeholder="Enter the value of B.O.D"/>
<input type="text" name="na" placeholder="Enter the value of Nitratenen"/>
<input type="text" name="tc" placeholder="Enter the value of Total Coliform"/>
<input type="submit" class="logbtn" value="Predict The Result"></center>
<div class="bor"><center>{{showcase}}</center></div>
</form>
</body>
</html>
Ln 17, Col 14 Tab Size: 4 UTF-8 CRLF HTML
Type here to search
25°C ENG 22:40 16-11-2022


app.py - Efficient water quality analysis and prediction using ML - Visual Studio Code
EXPLORER
EFFICIENT WATER QUALITY ANALYSIS AND PREDIC...
data_set
water_dataX.csv
Flask App
Image
HI1.png
UI_image.jpg
static
css
main.css
style.css
js
js.py
templates
web.html
app.py
sc.pkl
wqi.pkl
OUTLINE
TIMELINE
style.css
app.py
WQI.ipynb.txt
web.html
Flask App > app.py
import numpy as np
from flask import Flask,render_template,request
import pickle
app = Flask(__name__)
model = pickle.load(open('wqi.pkl','rb'))
@app.route('/')
def hello_world() :
return render_template("web.html")
@app.route('/login',methods = ['POST','GET'])
def login() :
year = request.form["year"]
do = request.form["do"]
ph = request.form["ph"]
co = request.form["co"]
bod = request.form["bod"]
na = request.form["na"]
tc = request.form["tc"]
total = [[int(year),float(do),float(ph),float(co),float(bod),float(na),float(tc)]]
y_pred = model.predict(total)
print(y_pred)
y_pred =y_pred[[0]]
if(y_pred >= 95 and y_pred <= 100) :
return render_template("web.html",showcase = 'prediction is Excellent,The predicted value is
elif(y_pred >= 89 and y_pred <= 94) :
return render_template("web.html",showcase = 'prediction is Very good,The predicted value is
elif(y_pred >= 80 and y_pred <= 88) :
Discovering Python Interpreters
Ln 14, Col 10 Spaces: 4 UTF-8 CRLF Python 3.10.1 64-bit
Type here to search
25°C ENG 22:41 16-11-2022


File Edit Selection View Go Run Terminal Help
app.py - Efficient water quality analysis and prediction using ML - Visual Studio Code
EXPLORER
EFFICIENT WATER QUALITY ANALYSIS AND PREDIC...
data_set
water_dataX.csv
Flask App
Image
HI1.png
UI_image.jpg
static
css
main.css
style.css
js
js.py
templates
web.html
app.py
sc.pkl
wqi.pkl
OUTLINE
TIMELINE
style.css
app.py
WQI.ipynb.txt
web.html
Flask App > app.py > login
print(y_pred)
y_pred =y_pred[[0]]
if(y_pred >= 95 and y_pred <= 100) :
return render_template("web.html",showcase = 'prediction is Excellent,The predicted value is
elif(y_pred >= 89 and y_pred <= 94) :
return render_template("web.html",showcase = 'prediction is Very good,The predicted value is
elif(y_pred >= 80 and y_pred <= 88) :
return render_template("web.html",showcase = 'prediction is Good,The predicted value is'+str(
elif(y_pred >= 65 and y_pred <= 79) :
return render_template("web.html",showcase = 'prediction is Fair,The predicted value is '+str
elif(y_pred >= 45 and y_pred <= 64) :
return render_template("web.html",showcase = 'prediction is Marginal,The predicted value is '
else :
return render_template("web.html",showcase = 'prediction is Poor,The predicted value is '+str
if __name__ == '__main__' :
app.run(debug = True,port=5000)
Ln 14, Col 10 Spaces: 4 UTF-8 CRLF Python 3.10.1 64-bit
Type here to search
25°C ENG 22:41 16-11-2022


File Edit Selection View Go Run Terminal Help
style.css - Efficient water quality analysis and prediction using ML - Visual Studio Code
EXPLORER
EFFICIENT WATER QUALITY ANALYSIS AND PREDIC...
data_set
water_dataX.csv
Flask App
Image
HI1.png
UI_image.jpg
static
css
main.css
style.css
js
js.py
templates
web.html
app.py
sc.pkl
wqi.pkl
OUTLINE
TIMELINE
style.css
app.py 2
WQI.ipynb.txt
web.html
Flask App > static > css > style.css > body
*{
margin:0;
padding:0;
text-decoration:none;
font-family:montserrat;
}
.pd{
padding-bottom:100%;}
body
{
background-image:url('https://images.wallpaperscraft.com/image/single/water_underwater_depth_197405_
background-position: center;
font-family:sans-serif;
background-size:cover;
margin-top:40px;
}
.main{
background-color:rgb(0,0,0,0.6);
width:800px;
height:500px;
margin:auto;
position:center;
border-top-left-radius:100px;
border-bottom-right-radius:100px;
}
.main input[type="text"],.main input[type="text"],.main input[type="text"],.main input[type="text"],.
border:0;
Ln 11, Col 114 Tab Size: 4 UTF-8 CRLF CSS
Type here to search
25°C ENG 22:41 16-11-2022


File Edit Selection View Go Run Terminal Help
style.css - Efficient water quality analysis and prediction using ML - Visual Studio Code
EXPLORER
EFFICIENT WATER QUALITY ANALYSIS AND PREDIC...
data_set
water_dataX.csv
Flask App
Image
HI1.png
UI_image.jpg
static
css
main.css
style.css
js
js.py
templates
web.html
app.py
sc.pkl
wqi.pkl
OUTLINE
TIMELINE
style.css
app.py
WQI.ipynb.txt
web.html
Flask App > static > css > # style.css > body
border:0;
background:none;
display:block;
margin:20px auto;
text-align:center;
border:2px solid #dbbf34;
padding:10px 3px;
width:200px;
outline:none;
color:white;
border-radius:24px;
transition:0.25s;
}
.main input[type="text"]:focus,.main input[type="text"]:focus,.main input[type="text"]:focus,.main ir
width:280px;
border-color:whitesmoke;
}
.logbtn{
display:block;
width:35%;
height:50px;
border:none;
border-radius:24px;
background:linear-gradient(120deg,#dbbf34,#8e44ad,#dbbf34);
background-size:200%;
color:#fff;
outline:none;
cursor:pointer;
Ln 11, Col 114 Tab Size: 4 UTF-8 CRLF CSS
Type here to search
25°C
ENG
22:41
16-11-2022


File Edit Selection View Go Run Terminal Help
style.css - Efficient water quality analysis and prediction using ML - Visual Studio Code
EXPLORER
EFFICIENT WATER QUALITY ANALYSIS AND PREDIC...
data_set
water_dataX.csv
Flask App
Image
HI1.png
UI_image.jpg
static
css
main.css
style.css
js
js.py
templates
web.html
app.py
sc.pkl
wqi.pkl
OUTLINE
TIMELINE
style.css
app.py
WQI.ipynb.txt
web.html
Flask App > static > css > # style.css > body
border-radius:24px;
transition:0.25s;
}
.main input[type="text"]:focus,.main input[type="text"]:focus,.main input[type="text"]:focus,.main ir
width:280px;
border-color:whitesmoke;
}
.logbtn{
display:block;
width:35%;
height:50px;
border:none;
border-radius:24px;
background:linear-gradient(120deg,#dbbf34,#8e44ad,#dbbf34);
background-size:200%;
color:#fff;
outline:none;
cursor:pointer;
transition:.5s;
}
.logbtn:hover{
background-position:right;
}
.predict{
text-align:center;
color:black;
}
Ln 11, Col 114 Tab Size: 4 UTF-8 CRLF CSS
Type here to search
25°C
ENG
22:41
16-11-2022

13. APPENDIX

REQUIREMENT.TXT

Flask == 2.2.2

numpy == 1.23.4

pandas == 1.5.1

scikit-learn == 1.1.3

matplotlib == 3.6.2

seaborn == 0.12.1

flask-cors==3.0.10

Github Link

https://github.com/IBM-EPBL/IBM-Project-10441-1659180083

Demo Link:

https://youtu.be/PM4XONnHVPU