PROJECT REPORT

EFFICIENT WATER QUALITY ANALYSIS AND PREDICTION USING MACHINE LEARNING

By team - PNT2022TMID30589

SINDHUJA.R (613019104074)

VAISHNAVI.M (613019104090)

TAMIL SELVI .K.B(613019104087)

SANTHIYA.K (613019104067)

INDEX

1.INTRODUCTION

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. 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 illnesses. The most common of these diseases in Pakistan are diarrhoea, typhoid, gastroenteritis, cryptosporidium infections, some forms of hepatitis and giardiasis intestinal worms. In Pakistan, water borne diseases, cause a GDP loss of 0.6–1.44% every year. This makes it a pressing problem, particularly in a developing country like Pakistan. 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 calculation, 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) and water quality class (WQC). 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. In this way, expensive and cumbersome lab analysis with specific sensors can be avoided in further similar analyses. A series of representative supervised prediction (classification and regression) algorithms were tested on the dataset worked here. The complete methodology is proposed in the context of water quality numerical analysis.

1.2 PURPOSE

Water makes up about 70% of the earth's surface and is one of the most important sources vital to sustaining life. Rapid urbanization and industrialization have led to a deterioration of water quality at an alarming rate, resulting in harrowing diseases. Water quality has been conventionally estimated through expensive and time-consuming lab and statistical analyses, which render the contemporary notion of real-time monitoring moot. The alarming consequences of poor water quality necessitate an alternative method, which is quicker and inexpensive. With this motivation, this research explores a series of supervised machine learning algorithms to estimate the water quality index (WQI), which is a singular index to describe the general quality of water, and the water quality class (WQC), which is a distinctive class defined on the basis of the WQI. The proposed methodology employs four input parameters, namely, temperature, turbidity, Ph and total dissolved solids. Of all the employed algorithms, gradient boosting, with a learning rate of 0.1 and polynomial regression, with a degree of 2, predict the WQI most efficiently, having a mean absolute error (MAE) of 1.9642 and 2.7273, respectively. Whereas multi-layer perceptron (MLP), with a configuration of (3, 7), classifies the WQC most efficiently, with an accuracy of 0.8507. The proposed methodology achieves reasonable accuracy using a minimal number of parameters to validate the possibility of its use in real time water quality detection systems.

2. LITERATURE REVIEW

2.1 EXISTING PROBLEM

The basic idea of this research is to devise a comprehensive methodology that analyzes and predicts the water quality of particular regions with the help of certain water quality parameters. These parameters include physical, biological, or chemical factors which influence water quality. There are certain quality standards set up by international organizations like the World Health Organization (WHO) and the Environmental Protection Agency (EPA), which serve as a benchmark for determining the quality of water. In its document "Efficient Water Quality Analysis and Prediction using Machine Learning", EPA mentions a total of 101 parameters that affect water quality in one way or another. However, some parameters have a greater and more visible effect on water quality than others.

TITLE: IMPROVING THE ROBUSTNESS OF BEACH WATER QUALITY MODELING USING AN ENSEMBLE MACHINE LEARNING

AUTHOR: Wang et al (2021)

This study demonstrates the utility of using a model stacking approach for predictive modeling of beach water quality. Since model stacking averages out noise from its base models, it is theoretically more promising than individual models in generating predictions with greater accuracy and robustness. The results from this study suggest that the model stacking algorithm has promise for improving the reliability of predictive modeling for beach microbial water quality of other sites with similar hydrogeological and environmental conditions such as other beaches along the Great Lakes. A comprehensive test needs to be done to understand the strength and weaknesses of individual base models and the stacking approach. This study indicated that the model stacking approach mayimprove the robustness of beach water quality modeling.

TITLE: ACCURATE PREDICTION SCHEME OF WATER QUALITY IN SMARTMARICULTURE WITH A DEEP BI-S-SRU LEARNING NETWORK

AUTHOR: J. Liu, C. Yu, Z. Hu et al (2020)

This paper proposed the process and model for the accurate prediction of key water quality parameters (pH, water temperature, and dissolved oxygen). Firstly, the collected water quality data is repaired and corrected by the improved preprocessing method, and then the data is filtered and denoised by the wavelet transform method. After preprocessing, the data received by remote transmission can be recovered well. Next, we construct the Bi-S-SRU (Bi-directional Stacked SRU) deep learning prediction model by importing a pretreated dataset weighted with the discovered correlation coefficients. The experimental results demonstrate that our proposed prediction model canachieve higher prediction accuracy and stability compared with RNN-based and SRU-based prediction

models. The experimental results also show that the Bi-S-SRU-based prediction method is only slightly higher in time complexity than the traditional RNN-based or LSTM-based prediction method.

TITLE: ASSESSMENT OF SURFACE WATER QUALITY BY USING SATELLITE IMAGES FUSION BASED ON PCA METHOD IN THE LAKE GALA, TURKEY

AUTHOR: E. Batur and D. Makita (2019)

In this paper, the PCA model is presented to integrate surface water reflectance values from satellite images to monitor Gala Lake's surface water quality. The values of Chl-a, DO, TSS, SDD, TDS, and pH values calculated by the PCA method were found to be highly correlated with the measured water quality parameters. The results obtained were found to be directly proportional to the number of sensors. L8 OLI and S2A have higher spectral resolution than GK2 images. However, the high temporal resolution of GK2 allows the desired region to be displayed at more frequent intervals, allowing for better monitoring of the instantaneous changes in surface water quality. Therefore, longer measurements should be made and analyzed for a model covering all periods.

TITLE: SURFACE WATER POLLUTION DETECTION USING THE INTERNET OF THINGS

AUTHOR: Shafi et al (2018)

In this paper, the proposed an IoT-based solution to monitor water quality in real-time. The proposed system provides remote monitoring of water quality assessment along with water flow control via a mobile app. Four machine learning algorithms including Support Vector Machine (SVM), k Nearest Neighbor (kNN), single layer neural network, and deep neural network have been applied for the classification of water quality and experimental results revealed that deep neural network outperforms all other algorithms with an accuracy of 93. This system has the potential to effectively utilize to overcome the challenges of water quality in the agriculture sector and various industries.

TITLE: IMPROVING WATER QUALITY INDEX PREDICTION IN PERAK RIVER BASIN MALAYSIA THROUGH A COMBINATION OF MULTIPLE NEURAL NETWORKS

AUTHOR: Ahmad et al (2017)

In this paper, they proposed a reliable real-time prediction model for WQI developed through a selective combination of multiple neural networks by excluding COD and BOD from model inputs asthey cannot be measured in real-time. Single and multiple FANN are used in this paper to model the WQI in the Perak River basin. The selective combination schemes provide models with bettergeneralization capability compared to combining all neural networks. The bootstrap aggregated models with selective

combination provide a real-time WQI prediction tool without delay as only real-time measurements are used as model inputs.

TITLE: ARTIFICIAL INTELLIGENCE FOR THE PREDICTION OF WATER QUALITY INDEXIN GROUNDWATER SYSTEMS

AUTHOR: Mohamad Sakizadeh (2016)

One of the problems of ANN's modeling in environmental studies which suffers from the problem of the small data records is the danger of over-fitting the model to the training data resulting in poor generalization of the model for the data out-of-the training data range. This study's results proved that this problem can be obviated by using some algorithms like Bayesian regularization and Ensemble methods. The prediction of water quality index (WQI) was successfully implemented by Bayesian regularization and Ensemble averaging methods, though the performance of Bayesian regularization was roughly better, with minimum test error indicating the good generalization ability of these methods in this field. The poor generalization ability is a problem that has been overlooked by most of the research all around the world although it is an important issue that should be taken into account.

TITLE: THE USE OF COMBINED NEURAL NETWORKS AND GENETIC ALGORITHMS FORTHE PREDICTION OF RIVER WATER QUALITY

AUTHOR: Ding et al (2014)

In this paper, they propose a water quality prediction model that combines PCA, BPNN, and GA. Using the BPNN model to study water classification and prediction can overcome disadvantages including the large workload of traditional evaluation methods and strong subjectivity. This model possesses objectivity, universality, and practicality. PCA converts the multi-indices into a few aggregative indices with little original data information loss and reduces the input data to speed the training process. Using GA to optimize network parameters can effectively prevent the search process from converging to local optimum solutions, optimize global optimal network parameters, and significantly improve the accuracy of water quality prediction. This model can obtain high training speed and good prediction rate and can be extended to other classification problems.

Our Ideology

The estimated water quality in our work is based on nine parameters: ph, Hardness, Solids, Chloramines, Sulfate, Conductivity, Organic carbon, Trihalomethanes, Turbidity, and pH, which are tested according to World Health Organization (WHO) standards.

The proposed methodology improves on these notions and the methodology being followed isdepicted in Figure 1.

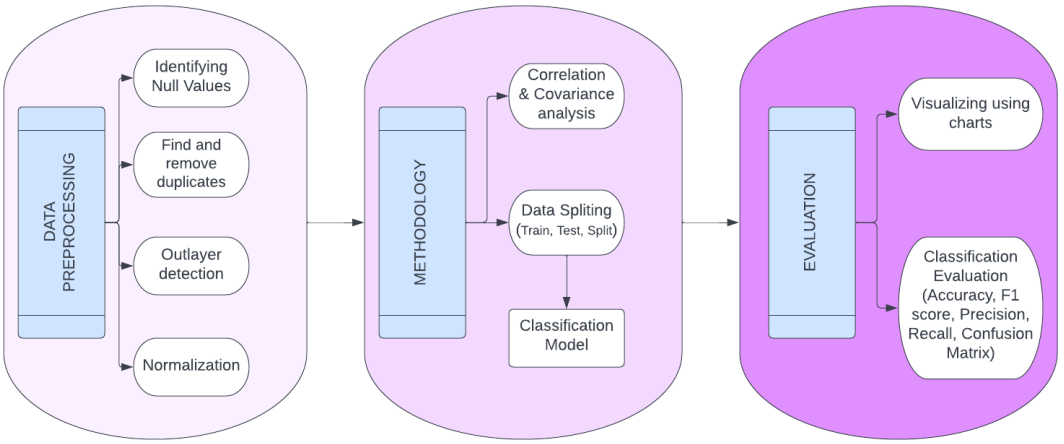


Figure 1

COMPARATIVE ANALYSIS OF LITERATURE SURVEY:

S.No	Year	Researcher	Title	Parameters	Algorithm	Remarks
01	2021	Wang et al	Improving the robustness of beach water quality modeling using an ensemble machine learning	turbidity, temperature, Culturable fecal indicator bacteria such as Escherichia coli (E. coli)	Partial least square,sparse partial least square, random forest, Bayesiannetwork, Akhand linear regression	Highest accuracy of 82.3% with ensemble machine learning algorithm
02	2020	J. Liu, C. Yu, Z. Huet al	Accurate prediction scheme of water quality in smart mariculture with a deep Bi-S-SRU learning network	Salinity, chlorophyll, turbidity, Water Temperature, PH, Dissolved Oxygen(DO)	LSTM, SRU, RNN, LSTM, SRU and Bi-S-SRU	Highest accuracy of 94.42% using a Bi-S-SRU
03	2019	E. Batur and D. Makita	Assessment of surface water quality by using satellite images fusion based on PCA method in the Lake Gala, Turkey	DO, SDD, TDS, and pH Chl-a and TSS	MLR, SVM, ANN,AND PCA	Highest accuracy of 92% using a PCA-based RSR model
04	2018	Shafi et al	Surface Water Pollution Detection using the Internet of Things	turbidity, temperature and pH	Support Vector Machines (SVM), Neural Networks (NN),Deep Neural Networks(Deep NN), and k Nearest Neighbors (kNN)	Highest accuracy of 93% with Deep NN

05	2017	Ahmad etal	Improving water quality index prediction in Perak River basin Malaysia through a combination of multiple neural networks	Nitrate, PH, Electrical conductivity, Dissolved oxygen, total coliform, Biochemical Oxygen Demand	feedforward artificial neural network; forward selection; backward elimination; artificial neural network;multiple neural networks	Highest accuracy of 92.7% using a selective combination methods
06	2016	Sakizadeh	Artificial intelligence for the prediction of water quality index in groundwater systems	EC, TDS, Mn, Cu, Cr(VI), Turbidity, pH, Ca, Mg, Total hardness, Sulfate, Fe, Fluoride Phosphate, Nitrate, Nitrite	ANN with Bayesian regularization	Highest accuracy of 80% using an Artificial Neural Network
07	2014	Ding et al	The Use of Combined Neural Networks and Genetic Algorithms for Prediction of River Water Quality	pH, NH3-N, TN, Cr6+, TP, CODMn, BOD5, TCN, COD, Cd, Cu, Zn, Pb, Hg, As, Se, F-, sulfide, dissolved oxygen, and LAS, etc.	Genetic Algorithm (GA), and Back Propagation Neural Network (BPNN)	The highest accuracy of Non polluted and polluted of 88.9% and 93.1% with PCA technique

2.2 REFERENCES

1. Wang Et Al (2021) - Improving the Robustness of Beach Water Quality Modeling usingan Ensemble Machine Learning
2. J. Liu, C. Yu, Z. Hu Et Al (2020) - Accurate Prediction Scheme of Water Quality in SmartMariculture with A Deep Bi-S-Sru Learning Network
3. E. Batur and D. Makita (2019) - Assessment of Surface Water Quality by using Satellite Images Fusion based on PCA Method in the Lake Gala, Turkey
4. Shafi Et Al (2018) - Surface Water Pollution Detection using The Internet of Things
5. Ahmad Et Al (2017) - Improving Water Quality Index Prediction in Perak River BasinMalaysia Through a Combination of Multiple Neural Networks
6. Mohamad Sakizadeh (2016) - Artificial Intelligence for the Prediction of Water Quality Index in Groundwater Systems
7. Ding Et Al (2014) - The Use of Combined Neural Networks and Genetic Algorithms forthe Prediction of River Water Quality

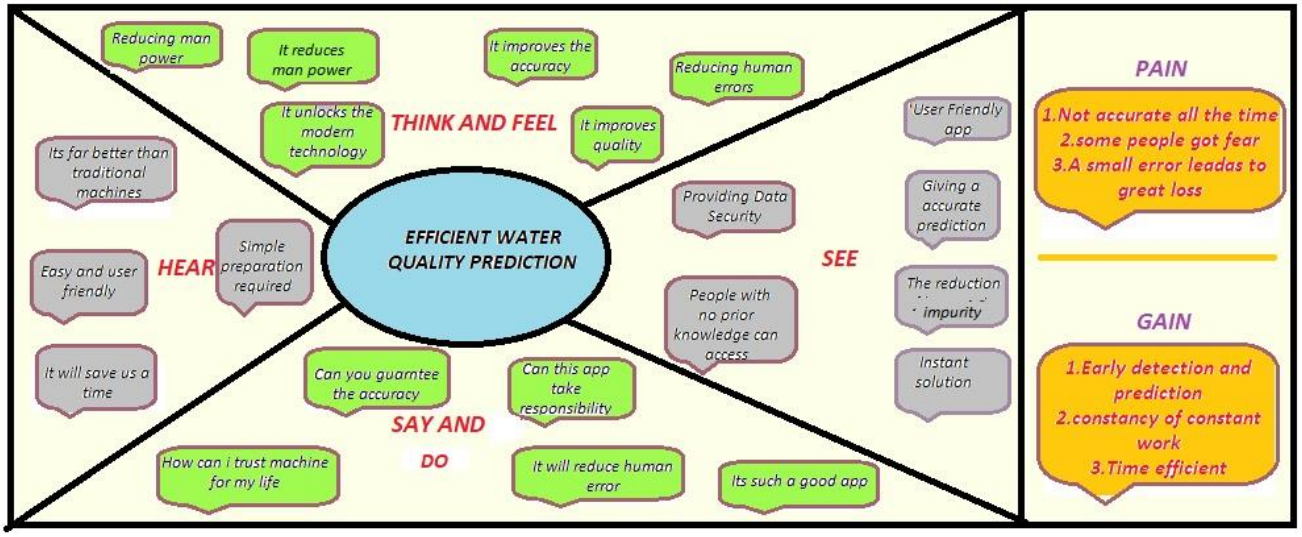
2.3 PROBLEM STATEMENT DEFINITION

What is the aim of the project?	The proposed methodology achieves reasonable accuracy using a minimal number of parameters to validate the possibility of its use in real-time water quality detection systems.
What are the boundaries of the problem?	There is no boundary limit for the issue because if anyone drinks unpurged or contaminated water, they will be affected.
What is the issue?	The most important behavioral risk factors of this disease can only be identified by taking samples of the contaminated water and then researching that water by using datasets and then only we can find the issue.
Where is the issue coming from?	It majorly occurs to the people on the riverside who use the river water. If the water had any harmful chemicals present, it would affect the people with a disease.
Why is it important that we fix the problem?	It is very crucial to develop an application that detects the disease because rapid urbanization and industrialization have led to a deterioration of water quality at an alarming rate, resulting in harrowing diseases. Water quality has been conventionally estimated through expensive, time-consuming lab and statistical analyses. In this, we are simply doing the project to find the chemicals using data science.

Which solution can be used to address this issue?	This study aims to predict water quality components using Bi-S-SRU (Bi-directional Stacked SRU) deep learning prediction model.
What methodology was used to solve the issue?	The estimated water quality in our work is based on nine parameters: pH, Hardness, Solids, Chloramines, Sulfate, Conductivity, Organic carbon, Trihalomethanes, Turbidity, and pH.

3. IDEATION & PROPOSED SOLUTION

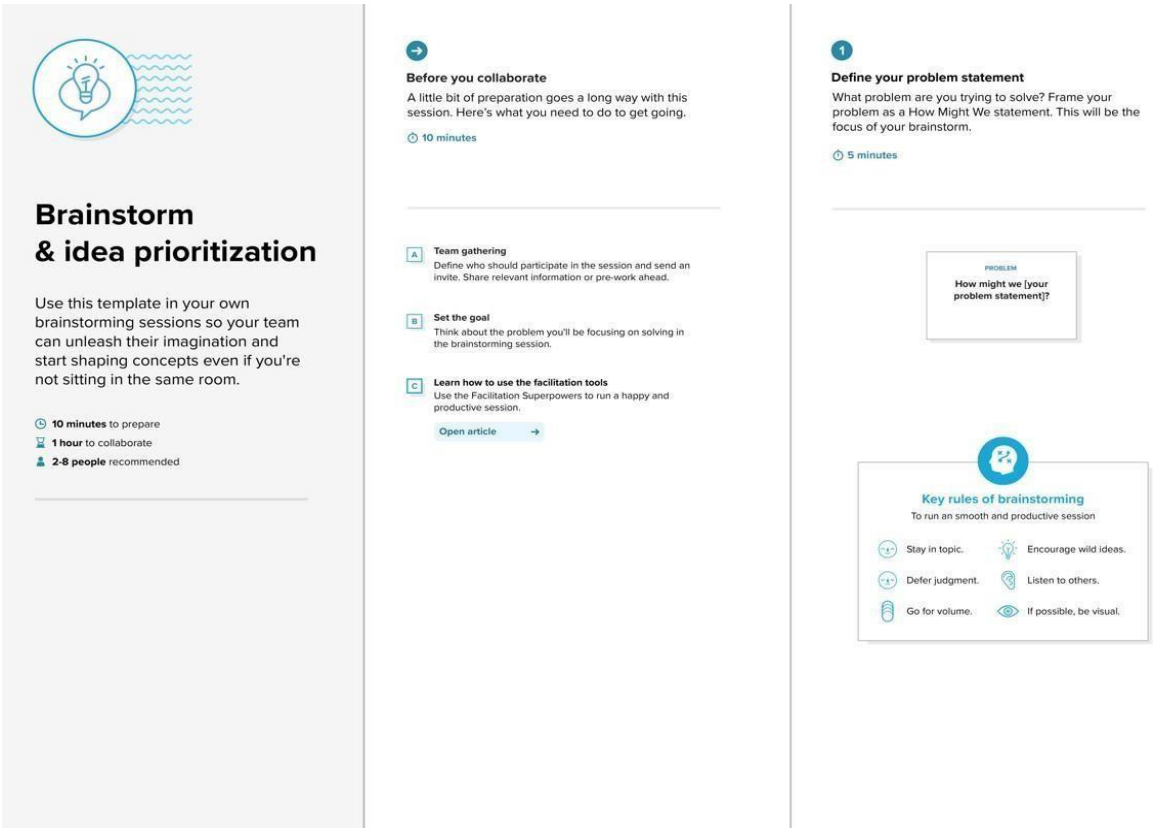
3.1 EMPATHY MAP CANVAS



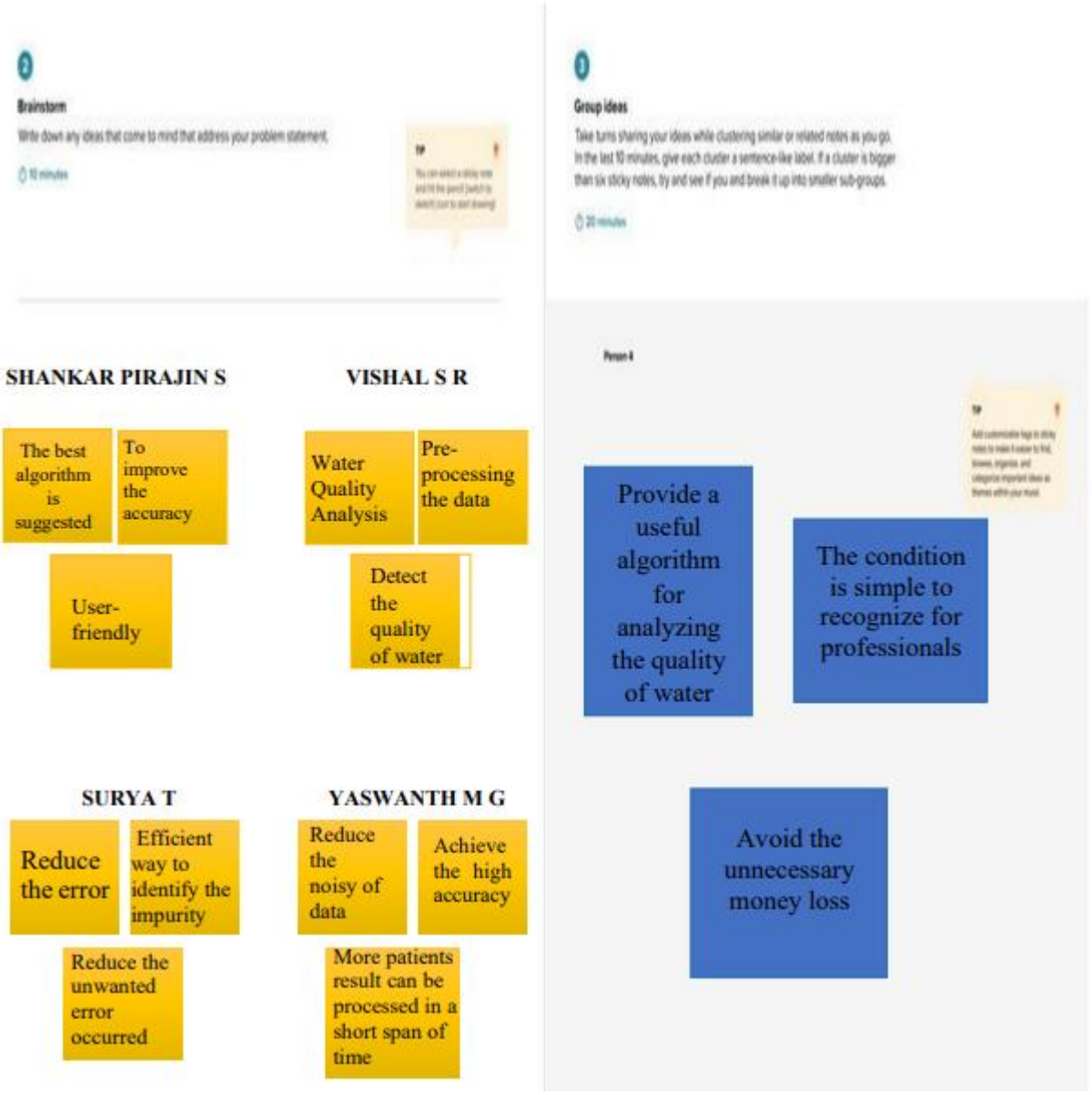
3.2 IDEATION & BRAINSTORMING

Brainstorm & Idea Prioritization Template:

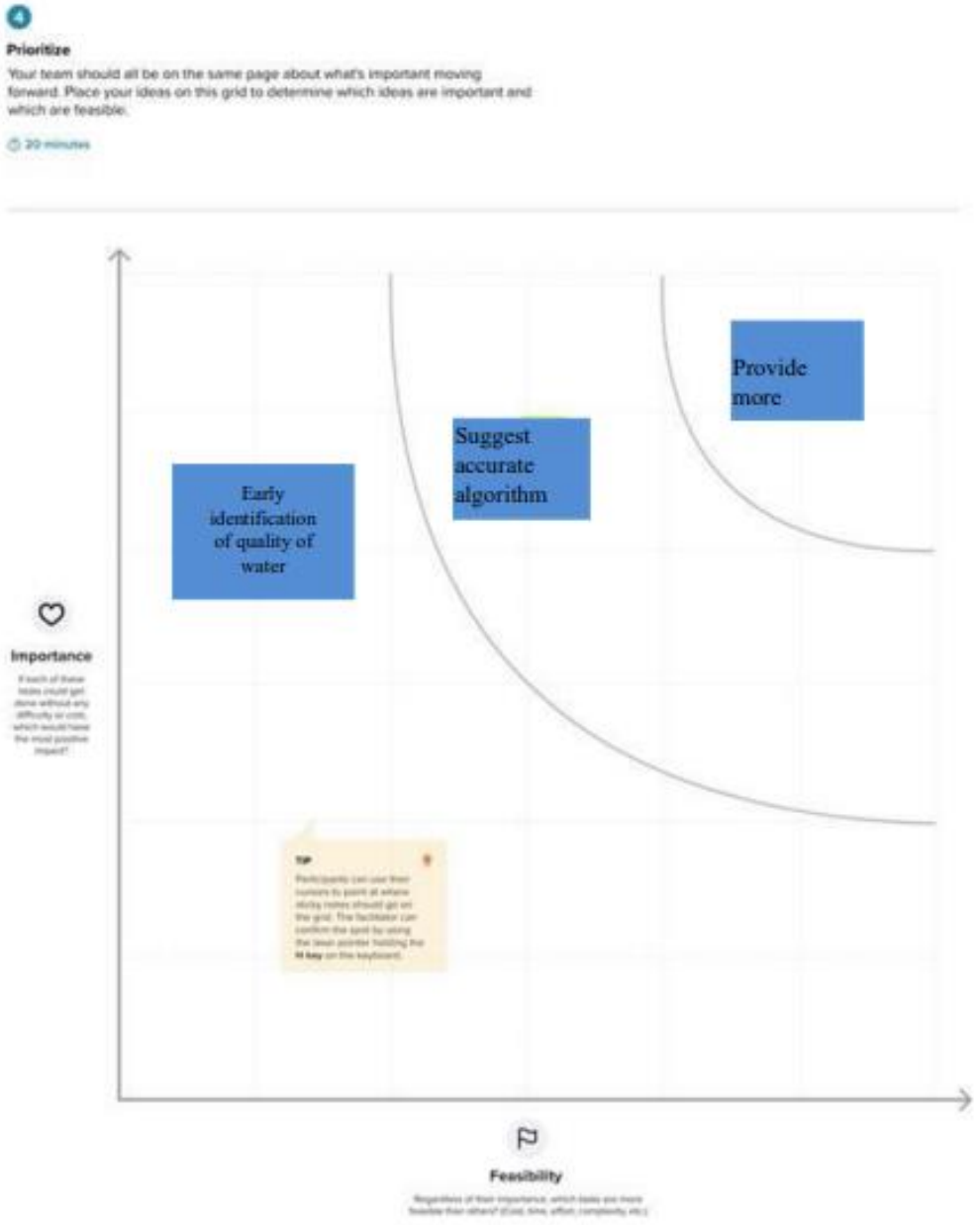
Step-1: Team Gathering, Collaboration and Select the Problem Statement



Step-2: Brainstorm, Idea Listing and Grouping



Step-3: Idea Prioritization



Step

3.3 PROPOSED SOLUTION

S.No.	Parameter	Description
1.	Problem Statement (Problem to be solved)	• People living in urban areas are concerned about the quality of water. Prevention and control of waterborne diseases are dependent on the quality of water as an environmental determinant. The water on the planet makes up about 70% of its surface and is one of the most important sources for life Water quality has deteriorated at an alarming rate because of rapid urbanization and industrialization. Water quality has been conventionally estimated through expensive and time- consuming lab and statistical analyses, which render the contemporary notionof real-time monitoring moot. • It is, however, challenging to predict the quality of urban water since it varies nonlinearly in urban spaces and depends on a range of factors, including meteorology, water use patterns, and land uses. Hence, this project aims at developing a Machine Learning (ML) model that predicts water quality by taking into account all the indicators of water quality.
2.	Idea / Solution description	• A proposed model uses PH, DO, andother water quality standard indicators to predict water quality. • On our dataset, we need to train the datasets to run smoothly and see an incremental improvement in prediction rate using Random Forest Regression • Our plan is to build a web application that is integrated with the model. It provides a user interface where the user can enter predictions. On the UI, predictions are displayed based on the values entered into the saved model.

3.	Novelty / Uniqueness	• Using the model, it is possible to determine whether the water is suitable for drinking. Therefore, it contributes to the maintenance of health.
4.	Social Impact / Customer Satisfaction	• Water makes up about 70% of the earth's surface and is one of the most important sources vital to sustaining life. Rapid urbanization and industrialization have led to a deterioration of water quality at an alarming rate, resulting in harrowing diseases. • Most of the research either employed manual lab analysis, not estimating the water quality index standard, or used too many parameters to be efficient enough. • With machine learning techniques, the implementation was done by the Water Quality Index (WQI). • Web app is developed as UI is provided for the customer/user where he has to enter the values for predictions.
5.	Business Model (Revenue Model)	• A web application that is integrated to the model built. A UI is provided for the uses where he has to enter the values for predictions. The enter values are given to the saved model and prediction is showcased on the UI and deploy it on IBM cloud. • We can sell it for the prediction of water in various environments if the model preforms well ,also can make the app as premium one.
6.	Scalability of the Solution	• The proposed can be implemented in realtime water quality analysis by getting water sample using devices(Internet Of Things). • Real time applications can be used in various places like schools, colleges etc. • Machine learning model integrated with IOT can make users more comfortable and to use in realtime.

3.4 PROBLEM SOLUTION FIT

<table>
<tr>
<td>

CS

1.CUSTOMER SEGMENT(s)

People, Residential, Commercial,Lab Testing

</td>
<td>

CC

6.CUSTOMER CONSTRAINTS

Water is essential for every one to sustain. If the water is impure it may cause diseases with this application it can be avoided.

</td>
<td>

AS

5.AVAILABLE SOLUTION

we need to train the datasets to run smoothly and see an incremental improvement in the prediction rate using Random Forest Regression algorithm on our dataset

</td>
</tr>
<tr>
<td>

J&P

2.JOB-TO-BE DONE/PROBLEMS

Check the quality of water, whetherthe water is drinkable, reason for unusability. Can verify the quality by themselves without expert

</td>
<td>

RC

9.PROBLEM ROOT CAUSE

The major cause of this problemis lack of drinking water and doesn't follow the proper diet and doesn't have proper awareness is also being a root cause.

</td>
<td>

BE

7.BEHAVIOUR

We will be building a web applicationthat is integrated to the model built. The enter values are given to the saved model and predictionis showcased on the UI

</td>
</tr>
<tr>
<td>

TR

3. TRIGGERS

Using this application, user can avoid the fear of water quality. Since the user knows the quality of water they are going to use.

EM

4. EMOTIONS:BEFORE/AFTER

Before: There are no application to predict the water quality.

After: By using this easy to predictthe quality of water using some a parameters.

</td>
<td>

SL

10.YOUR SOLUTION

The heart of the project depends upon the prediction ofthe quality of the water. As abundant as algorithms are present in order to achieve such a goal, it is mandatory to select the best and the most efficient algorithm to finalize the predicted value.

</td>
<td>

CH

8.CHANNEL OF BEHAVIOUR

Online: The application Notify the user with data preprocessing information

Offline: Water quality has been conventionally estimated through expensive and time-consuming lab and statistical analyses, which render the contemporary notion of real-time monitoring moot.

</td>
</tr>
</table>

4. REQUIREMENT ANALYSIS

4.1 Functional Requirements:

Following are the functional requirements of the proposed solution.

FR No.	Functional Requirement (Epic)	Sub Requirement (Story / Sub-Task)
FR-1	User Interface	A detailed description of water quality should be provided.
FR-2	User Form	Values and measures require to predict the Water quality should be given as input in the form.
FR-3	Machine Learning ModelDeployment	Develop the Machine Learning Regression Modelto predict the Water Quality Index (WQI). Develop the Machine Learning Classification Model to predict the Water Quality Classification (WQC).
FR-4	Testing The Water Samples	Provides an option to test any kind of water samples with the required parameters and to calculate the Water Quality Index and impurities present
FR-5	Reporting	If any issues are faced by the customer or user it will be directly notified to the developer

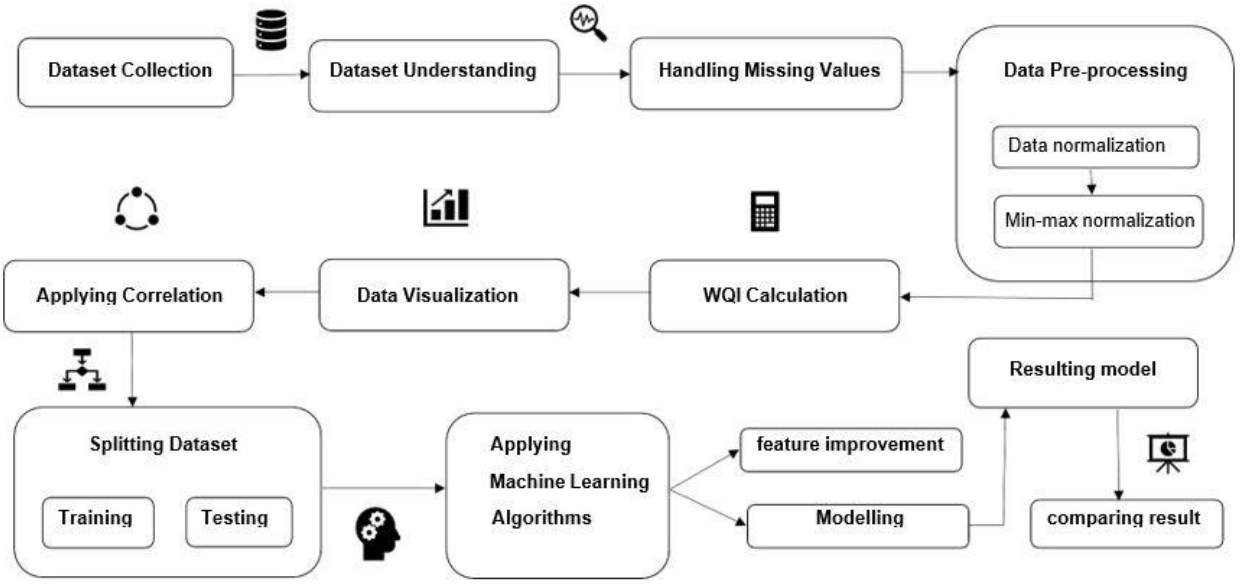
4.2 Non-functional Requirements:

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

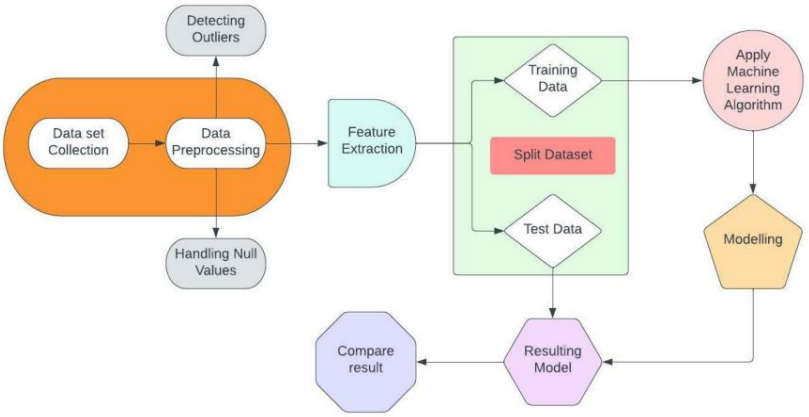
FR No.	Non-Functional Requirement	Description
NFR-1	**Usability**	Customers can access the system more efficiently and in a simpler way. The customers can have the opportunity to view a better interpretation of results. The customers are also recommended the purification techniques based on the impurities.
NFR-2	**Security**	All the predicted information is accessed only by the authenticated users
NFR-3	**Reliability**	It should be reliable in producing effective and efficient water quality prediction results. It should ensure the trust and belief among people that this water quality prediction system produces correct results when used.
NFR-4	**Performance**	The system should be consistent in producing the prediction results of the Water Quality Index (WQI) and also needs to ensure better throughput and response time compared to other systems.
NFR-5	**Availability**	The system can be utilized by the customers 24/7 and it should be availed to test any kind of water samples anywhere
NFR-6	**Scalability**	It can be used by a wide variety of users like testing agencies, private and public laboratories, restaurants and hotels, and peoplewho wish to test the quality of water they consume. The system should also be compatible enough to be integrated with future technologies also.

5.PROJECT DESIGN

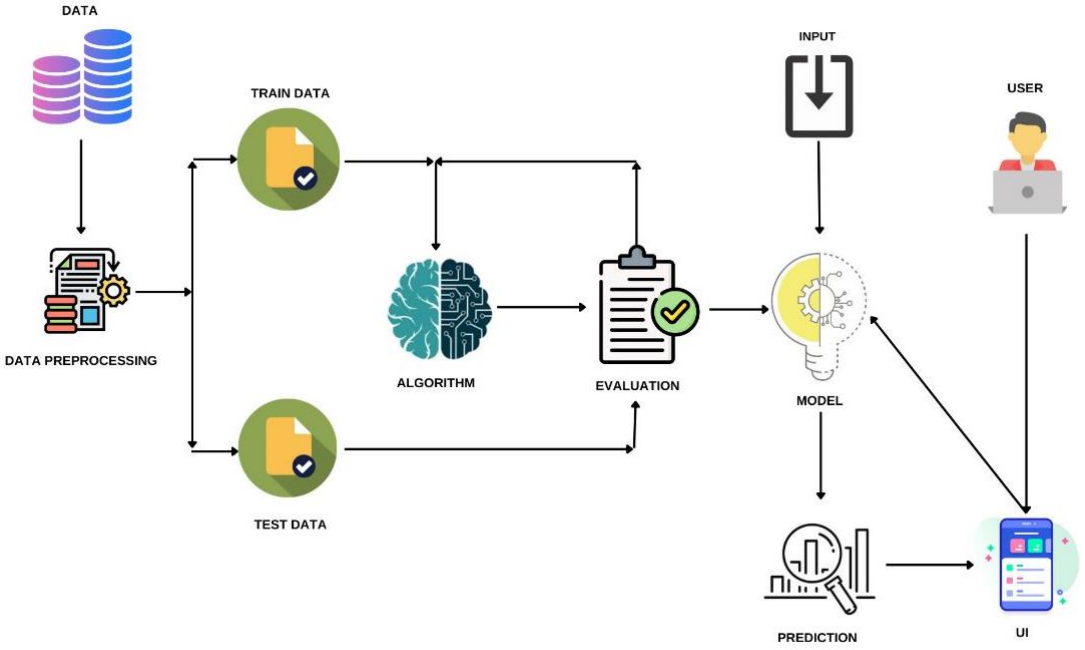
5.1 DATA FLOW DIAGRAMS



5.2 SOLUTION AND TECHNICAL ARCHITECTURE



Solution Architecture



Technical Architecture

Table-1: Components & Technologies:

S. No	Component	Description	Technology
1.	User Interface	User interacts by using web user interface.	HTML, CSS and Python Flask
2.	Application Logic-1 (Login)	User can able to login if that person is already registered tothe site.	HTML, CSS and Python Flask
3.	Application Logic-2 (Register)	User needs to be registered ifthat person is new to the site.	HTML, CSS and Python Flask.
4.	Application Logic-3(Reporting Form)	User needs to click on the reporting form in order to getthe prediction result	Front end- HTML, CSS and Python Flask. Back end – Query Languages, Python.
5.	Database	Data Type-String, Numeralvalues.	Query Languages such as MySQL, NoSQL etc.
6.	Cloud Database	Database Service on Cloud.	IBM DB2, IBM Cloud ant etc.
7.	File Storage	File storage requirements.	Local File-system.
8.	External API-1	Anyone can access the detailswith some restrictions to the personal details of other users.	Web API.
9.	External API-2	Accessibility.	Aadhar API.
10.	Machine Learning Model	Predict the result based on the training and testing dataset.	Data RecognitionModel, etc.
11.	Infrastructure (Server / Cloud)	Application Deployment onLocal System.	Local System.

Table-2: Application Characteristics:

S. No	Characteristics	Description	Technology
1.	Open-Source Frameworks	Frameworks are used for predictive data analysis, providing clear and actionableerror messages.	Tensor flow, Sci-kit learn, Keras.
2.	Security Implementations	OTP will be sent to the registered email id. Unauthorized users could not access the user's details.	Email Verification.
3.	Scalable Architecture	Scalability is improved for implementing the three-tier architecture.	Three tier architecture.
4.	Availability	For enhancing the high availability, load balancer isneeded.	Load Balancer.
5.	Performance	The model could be able to process large number of datasets.	Load Balancer.

5.3 USER STORIES

User Type	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptance criteria	Priority	Release
People (web user)		USN-1	As a user, I can understand the detailed description of water quality on the home page	I can access the web page	High	Sprint-1
	Input form	USN-2	As a user, I can enter the details required to analysis the water quality with use of form provided in the web page.	I can give inputs in the form and it is processed and visualize the water quality.	High	Sprint-2
		USN-3	As a user, I can contact the Customer care (people at the water resource organisation) toknow the details of water	I can contact people with Whatsapp, instagram, twitter, mail and also I canmake call	Medium	Sprint-3

6. PROJECT PLANNING AND SCHEDULING

6.1 SPRINT PLANNING AND ESTIMATION

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint-1	Data Preparation	USN-1	Collecting water dataset andpre-processing it	10	High	S Shankar Pirajin T Surya
Sprint-1	Model Building	USN-2	Create an ML model to predictwater quality	5	Medium	S Shankar Pirajin T Surya S R Vishal M G Yaswanth
Sprint-1	Model Evaluation	USN-3	Calculate the performance, error rate, and complexity of the ML model and evaluate the dataset based on the parameter that the dataset consists of.	5	Medium	
Sprint-2	Model Deployment	USN-4	As a user, I need to deploy the model andneed to find the results.	20	Medium	
Sprint-3	Web page (Form)	USN-5	As a user, I can use the application byentering the water dataset to analyze or predict the results.	20	Medium	S Shankar Pirajin T Surya S R Vishal M G Yaswanth
Sprint-4	Dashboard	USN-6	As a user, I can predict the water quality by clicking the submit button and the application will show whether the water isefficient for use or not.	20	High	S R Vishal M G Yaswanth

Project Tracker:

Sprint	Total Story Points	Duration	Sprint Start Date	Sprint End Date	Story Points Completed	Sprint Release Date
Sprint-1	20	6 Days	23 Oct 2022	28 Oct 2022	20	29 Oct 2022
Sprint-2	20	7 Days	29 Oct 2022	04 Nov 2022	20	05 Nov 2022
Sprint-3	20	7 Days	05 Nov 2022	11 Nov 2022	20	12 Nov 2022
Sprint-4	20	8 Days	12 Nov 2022	19 Nov 2022	20	19 Nov 2022

Velocity:

Sprint 1: 1 user stories x 20 story points = 20

Sprint 2: 1 user stories x 20 story points = 20

Sprint 3: 1 user stories x 20 story points = 20

Sprint 4: 1 user stories x 20 story points = 20

Total = 80

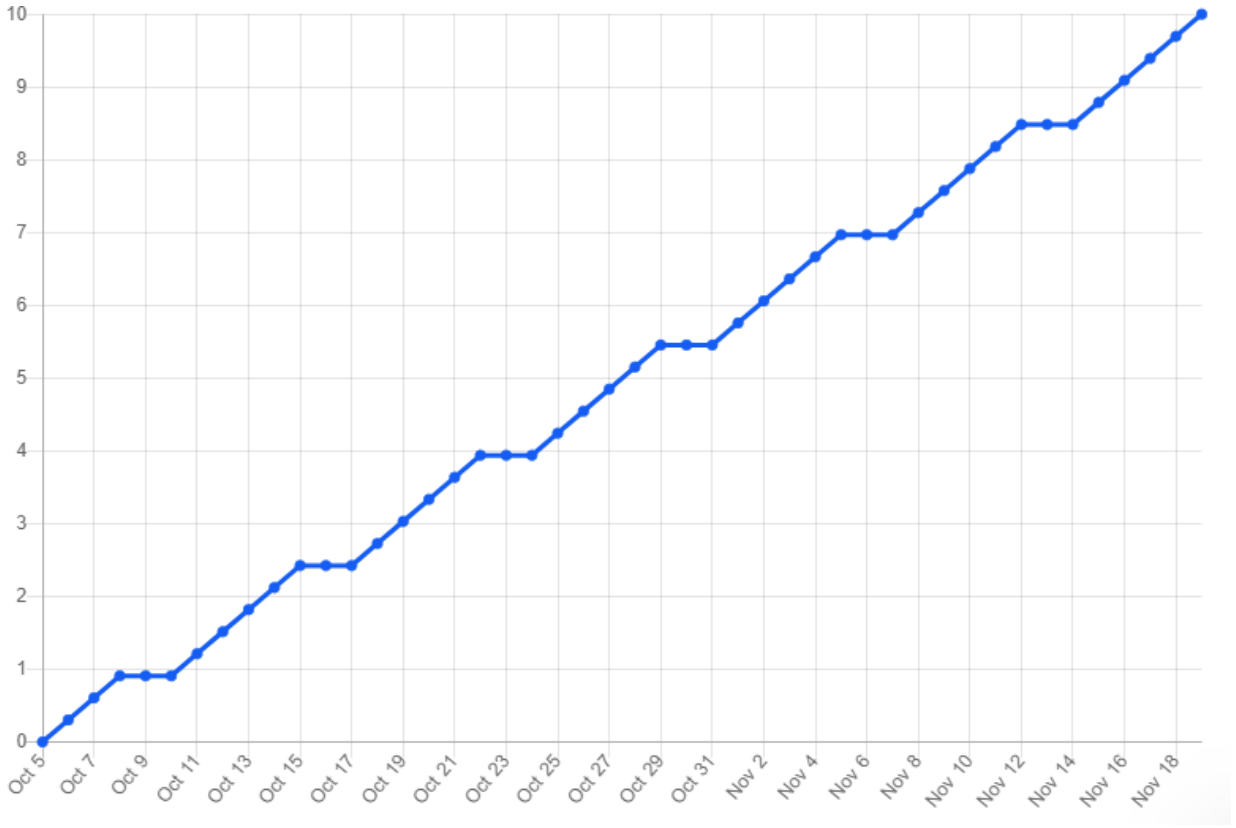
The average sprint velocity is 80 ÷ 4 = **20.**

6.2 PROJECT DELIVERY SCHEDULE

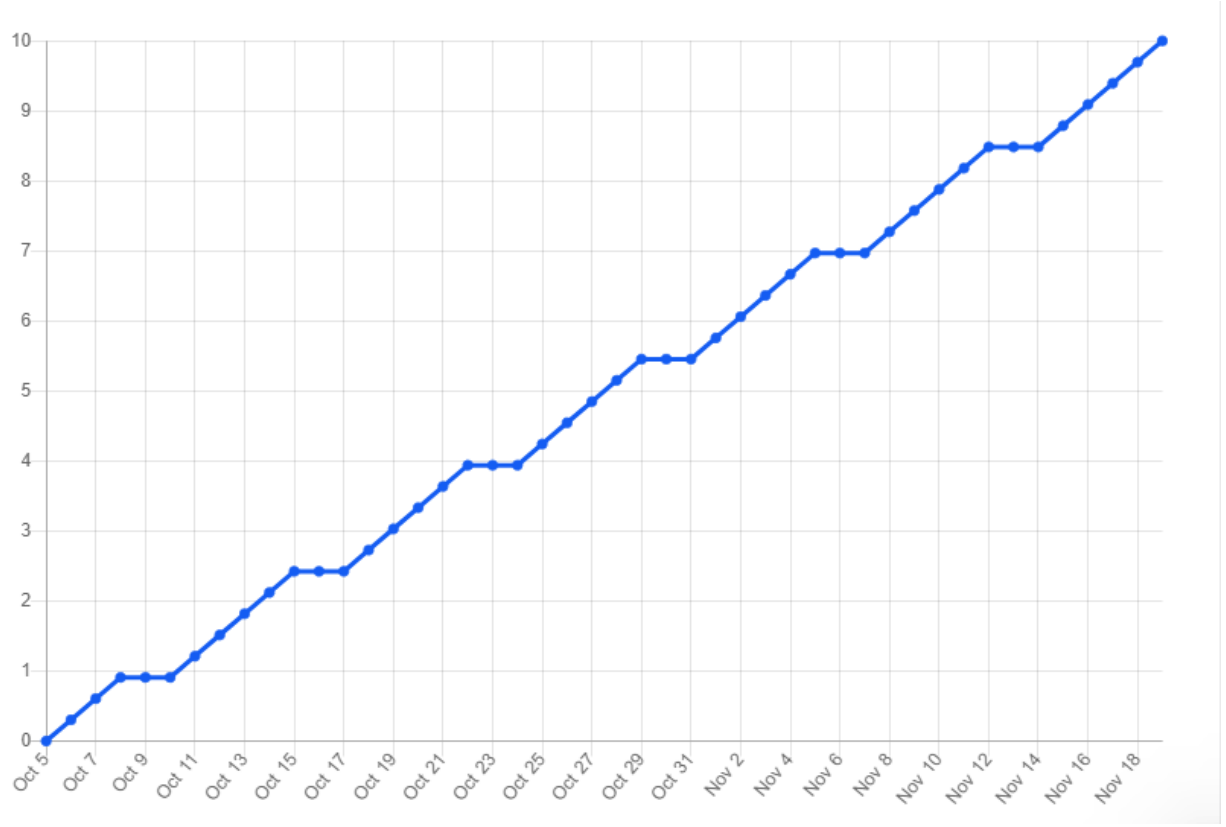
TITLE	DESCRIPTION	DATE
Literature Survey & Information Gathering	Literature survey on the selected project & gathering information by referring the technical papers, research publications, journals etc.	1 SEPTEMBER 2022
Prepare Empathy Map	Prepare Empathy Map Canvasto capture the user Pains and Gains, prepare list of problem Statements that are to be solved by this project.	7 SEPTEMBER 2022 & 9 SEPTEMBER 2022
Ideation	List the ideas by organizing a brainstorming session and prioritize the top three ideas based on the feasibility and importance.	15 SEPTEMBER 2022
Proposed Solution	Prepare the proposed solution document, which includes novelty, feasibility ofidea, revenue model, social impact, scalability of solution,etc.	22 SEPTEMBER 2022
Problem Solution Fit	Prepare problem - solution fit document.	30 SEPTEMBER 2022
Solution Architecture	Prepare solution architecture document.	30 SEPTEMBER 2022
Customer Journey	Prepare the customer journey maps to understandthe user interactions and experiences with the application (entry to exit).	6 OCTOBER 2022
Functional Requirement	Prepare the functional requirement document.	11 OCTOBER 2022
Data Flow Diagrams and User_Stories	Prepare the Data flow diagrams and User Stories for the problem	14 OCTOBER 2022
Technology Stack Architecture	Prepare the Technology Stack Architechture	17 OCTOBER 2022
Prepare Milestone &ActivityList	Prepare the milestones and activity list of theproject.	21 OCTOBER 2022
Project Development Phase	Develop Project Development Phase which include Sprint 1, Sprint 2, Sprint 3, Sprint 4	ON PROGRESS….

6.3 REPORTS FROM JIRA

BURNDOWN CHART



BURNUP CHART



7.CODING AND SOLUTIONING

7.1 FEATURE 1 (RANDOM FOREST ALGORITHM MODEL)

Random Forest Classifier is used to train and test the model for analyzing the quality of water with the help of collected and pre- processed dataset collections. NumPy is a library for the Python programming language, adding support for large, multi-dimensional arraysand matrices, along with a large collection of high-level mathematical functions to operate on these arrays. Moreover, NumPy forms the foundation of the Machine Learning stack. Pandas is an open-source Python package that is most widely used for data science/data analysis and machine learning tasks. Sea born is a Python data visualization librarybased on matplotlib. It provides a high-level interface for drawing attractiveand informative statistical graphics. For a brief introduction to the ideas behind the library, you can read the introductory notes or the paper. Matplotlib is a comprehensive library for creating static, animated, and interactive visualizations in Python. Matplotlib makes easy things easy andhard things possible. Create publication quality plots. Make interactive figures that can zoom, pan, update. EDA is applied to investigate the data and summarize the key insights. It will give you the basic understanding ofyour data, it is distribution, null values and much more. You can either explore data using graphs or through some python functions. There will betwo types of analysis. Descriptive statistics are brief informational coefficients that summarize a given data set, which can be either a representation of the entire population or a sample of a population. Descriptive statistics are broken down into measures of central tendency and measures of variability. Measures of central tendency include themean, median, and mode, while measures of variability include standard deviation, variance, minimum and maximum variables, kurtosis, and Skewness. Label Encoding refers to converting the labels into a numeric form to convert them into the machine-readable form. Machine learning algorithms can then decide in a better way how those labels must be operated. It is an important pre-processing step for the structured dataset in supervised learning. “Pickling” is the process whereby a Python object hierarchy is converted into a byte stream, and “unpickling” is the inverse operation, whereby a byte stream is converted back into an object hierarchy. XG Boost is an optimized distributed gradient boosting library designed to be highly efficient, flexible, and portable. It implements machine

learning algorithms under the Gradient Boosting framework.

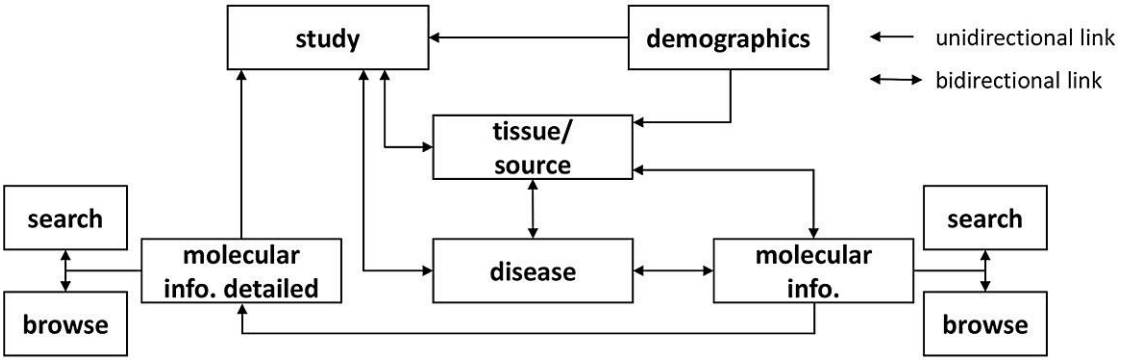
7.2 FEATURE 2(FLASK CONNECTIVITY)

The framework is the basis upon which software programs are built. It serves as a foundation for software developers, allowing them to create a variety of applications for certain platforms. It is a set of functions and predefined classes used to connect with the system software and handle inputs and outputs. It simplifies the life of a developer while giving them theability to use certain extensions and makes the online applications scalable and maintainable. Flask is a web application framework written in Python. A Web Application Framework or a simply a Web Framework represents a collection of libraries and modules that enable web application developers to write applications without worrying about low-level details such as protocol, thread management, among other examples. Flask is a web application framework written in Python. It was developed by Armin Ronacher, who led a team of international Python enthusiasts called Poocco. Flask is based on the Werkzeg WSGI toolkit and the Jinja2 template engine. Both are Pocco projects. The Web Server Gateway Interface (Web Server Gateway Interface, WSGI) has been used as a standard for Python web application development. WSGI is the specification of a common interface between web servers and web applications. Flask is often referred to as a micro-framework. It is designed to keep the core of the application simple and scalable. Instead of an abstraction layer for database support, Flask supports extensions to add such capabilities to the application. Unlike the Django framework, Flask is very Pythonic. It's easy to get started with Flask, because it doesn't have ahuge learning curve. HTML stands for Hyper Text Markup Language. HTMLis the standard markup language for creating Web pages. HTML describes the structure of a Web page. HTML consists of a series of elements. HTMLelements tell the browser how to display the content. Flask is used for developing web applications using python, implemented on Werkzeug and Jinja2. Advantages of using Flask framework are: There is a built-in development server and a fast debugger provided. The model deployed using Flask is used to predict the quality of water. Hypertext markup language (HTML) is the basic

language used to create documents for the Web and, along with HTTP (hypertext transfer protocol) and URLs (universal resource locators), is one of the three main protocols of the Web.Hypertext is text that contains hyperlinks. A hyperlink is an automated cross-reference to another location on the same document or to another document which, when selected by a user, causes the computer to display the linked location or document within a concise period. A markup language is a set of tags that can be embedded in digital text to provide additional information about it, including its content, structure and appearance. This information facilitates automated operations on the text,including formatting it for display, searching it and even modifying it. Some type of markup language is employed by every word processing program and by nearly every other program that displays text, although such languages and their tags are typically hidden from the user.HTML consistsof a set of predefined tags that can be embedded in text by web site designers in order to indicate the details of how web pages are rendered (i.e., converted into a final, easily usable, form) by web browsers. These details include paragraphing, margins, fonts (including style and size), columns, colors (background and text), links, the location of images, text flow around images, tables, and user input form elements (such as spacesfor adding text and submit buttons).

7.3 DATABASE SCHEMA

In the recent decades, the evolution of omics technologies has led to advances in all biological fields, creating a demand for effective storage, management and exchange of rapidly generated data and research discoveries. To address this need, the development of databases of experimental outputs has become a common part of scientific practice in order to serve as knowledge sources and data-sharing platforms, providing information about genes, transcripts, proteins or metabolites. In this review,we present omics databases available currently, with a special focus on their application in kidney research and possibly in clinical practice. Databases are divided into two categories: general databases with a broad information scope and kidney-specific databases distinctively concentrated on kidney pathologies. In research, databases can be used as a rich sourceof information about pathophysiological mechanisms and molecular targets. In the future, databases will support clinicians with their decisions,providing better and faster diagnoses and setting the direction towards more preventive, personalized medicine. We also provide a test case demonstrating the potential of biological databases in comparing multi- omics datasets and generating new hypotheses to answer a critical and common diagnostic problem in nephrology practice. In the future,employment of databases combined with data integration and data miningshould provide powerful insights into unlocking the mysteries of kidney disease, leading to a potential impact on pharmacological intervention andtherapeutic disease management.



8.TESTING

8.1 TEST CASES

Test case ID	Feature Type	Component	Test Scenario	Steps To Execute	Test Data	Expected Result	Actual Result	Status
Home Page_tc _1	Functional	Home Page	Verify useris able to see the dashboard of the webpage	1. to ensure that user can able to see information about water quality by clicking info 2. Verify the prediction button to analyse the quality a.Info button b.Predict button	_	Home and the buttons should be displayed	Working as expected	Pass
info page_tc _2	Functional	Info Page	Verify user is able to see the information of the webpage	1. to ensure that user can able to see information about water quality by clicking info 2. Verify Register/Signup page accepts only unique email a.Info button b.Predict button	_	Information should be displayed tothe user	Working as expected	Pass
Prediction Page_tc _5	UI	prediction page	Verify useris able to see the description and predict button	1. Enter URL and clickgo 2. Click on predicbutton 3. Enter D.O,Ph,conductivity in the text box 4.Enter nitratenen and give the total coliform to predict the water level 5.Click on predict button	_	Application should show correct prediction level of the water and user should ensure the quality by referring thegiven parameter table.	Working as expected	Pass
Prediction Page_tc _5	Functional	prediction page	Verify useris able to predict by giving letters	1. Enter URL and clickgo 2. Click on predicbutton 3. Enter D.O,Ph,conductivity in the text box	_	Application should not show any value because user	Working as expected	Pass

				4. Enter nitratenen asa letter 5. Click on predict button		entered letter		
Prediction Page_tc_5	Functional	prediction page	Verify the parameter table is listed below	1. parameters should be displayed by the result of water quality 2. displayed range should be shown in table	_	Application should show range of the result predicted by the webpage	Working as expected	Pass
Prediction Page_tc_5	Functional	predictio npage	Verify user can leave any field unfilled	1. Enter URL and click go 2. Click on predic button 3. Enter D.O,Ph,conductivity in the text box 4.Leave nitratenen field as empty 5.Click on predict button	_	Application should not show any value because user left a field as empty	Working as expected	Pass

Test Scenarios:

1 Verify user is able to see home page

2 Verify user is able to predict the WQI or not?

3 Verify user is able to navigate to information page?

4 Verify user is able to enter values to input field?

5 Verify Prediction page elements?

6 Verify user is able to enter any text in the input field?

7 Verify user is able to see the prediction value and the result?

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 wereresolved

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	8	2	4	10	37
Not Reproduced	0	0	1	0	1
Skipped	0	0	0	0	0
Won't Fix	0	5	2	1	8
Totals	19	24	17	16	58

3. Test Case Analysis

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

Section	Total Cases	Not Tested	Fail	Pass
Home Page	7	0	0	7
Client Application	51	0	0	51
Prediction	2	0	0	2
Pop ups	3	0	0	3
URL port	9	0	0	9
Final Report Output	4	0	0	4
Redirection	2	0	0	2

9.RESULTS

9.1 PERFORMANCE METRICES

TITLE	DESCRIPTION	DATE
Literature Survey & Information Gathering	Literature survey on the selected project & gathering information by referring the technical papers, research publications, journals etc.	1 SEPTEMBER 2022
Prepare Empathy Map	Prepare Empathy Map Canvasto capture the user Pains and Gains, prepare list of problem Statements that are to be solved by this project.	7 SEPTEMBER 2022 & 9 SEPTEMBER 2022
Ideation	List the ideas by organizinga brainstorming session and prioritize the top three ideas based on the feasibility and importance.	15 SEPTEMBER 2022
Proposed Solution	Prepare the proposed solution document, which includes novelty, feasibility ofidea, revenue model, social impact, scalability of solution,etc.	22 SEPTEMBER 2022
Problem Solution Fit	Prepare problem - solution fit document.	30 SEPTEMBER 2022
Solution Architecture	Prepare solution architecture document.	30 SEPTEMBER 2022
Customer Journey	Prepare the customer journey maps to understandthe user interactions and experiences with the application (entry to exit).	6 OCTOBER 2022
Functional Requirement	Prepare the functional requirement document.	11 OCTOBER 2022
Data Flow Diagrams and User_Stories	Prepare the Data flow diagrams and User Stories for the problem	14 OCTOBER 2022
Technology Stack Architecture	Prepare the Technology Stack Architechture	17 OCTOBER 2022
Prepare Milestone &ActivityList	Prepare the milestones and activity list of the project.	21 OCTOBER 2022
Project Development Phase	Develop Project Development Phase which include Sprint 1, Sprint 2, Sprint 3, Sprint 4	ON PROGRESS….

10. ADVANTAGES AND DISADVANTAGES

ADVANTAGES

1.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.

2.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.

3.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

1.Training necessary Somewhat difficult to manage over time and with large data sets

2.Requires manual operation to submit data, some configuration required

3.Costly, usually only feasible under Exchange Network grants Technical expertise and network server required

4.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

Water is one of the most essential resources for survival and its quality is determined through WQI. Conventionally, to test water quality, one has to go through expensive and cumbersome lab analysis. This research explored an alternative method of machine learning to predict water quality using minimal and easily available water quality parameters. The data used to conduct the study were acquired from PCRWR and contained 663 samples from 12 different sources of Rawal Lake, Pakistan. A set of representative supervised machine learning algorithms were employed to estimate WQI. This showed that polynomial regression with a degree of 2, and gradient boosting, with a learning rate of 0.1, outperformed other regression algorithms by predicting WQI most efficiently, while MLP with a configuration of (3, 7) outperformed other classification algorithms by classifying WQC most efficiently. In this paper, the performance of artificial intelligence techniques were evaluated to predict the water quality components of Tireh River (Iran). To this end most dataset related well-known components, such as pH, SO_4, Na, Ca, Cl, Mg, HCO_3 etc., were collected. Results indicated that the applied models have suitable performance for predicting water quality.

12. FUTURE SCOPE

In future works, we propose integrating the findings of this research in a large-scale IoT-based online monitoring system using only the sensors of the required parameters. The tested algorithms would predict the water quality immediately based on the real-time data fed from the IoT system. The proposed IoT system would employ the parameter sensors of pH, turbidity, temperature and TDS for parameter readings and communicate those readings using an Arduino microcontroller and ZigBee transceiver. It would identify poor quality water before it is released for consumption and alert concerned authorities. It will hopefully result in curtailment of people consuming poor quality water and consequently de-escalate harrowing diseases like typhoid and diarrhea. In this regard, the application of a prescriptive analysis from the expected values would lead to future facilities to support decision and policy makers. More data sources are required to verify the reliability and robustness of the proposed models. So far, the water quality dataset from the LVW collected by Southern Nevada Water Authority and Las Vegas Wash Coordination Committee, and dataset collected from Boulder Basin have been used as the experimental dataset. In the future, more efforts will be made to find more datasets to build a more reliable water quality prediction model.

13.APPENDIX

SOURCE CODE:

Python Code

```
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('/',methods=['GET'])
def home():
    return render_template("index.html")
@app.route('/login',methods = ['POST'])
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)
    y_pred = y_pred[0]
    if(y_pred >= 95 and y_pred <=100):
        return render_template("index.html",showcase = "Excellent, The Predicted
Value is "+str(y_pred))
    elif(y_pred >=89 and y_pred <=94):
        return render_template("index.html",showcase = "Very Good, The Predicted
Value is "+str(y_pred))
```

```
    elif(y_pred >=80 and y_pred <=88):
        return render_template("index.html",showcase = "Good, The Predicted Value is
"+str(y_pred))
    elif(y_pred>=65 and y_pred<=79):
        return render_template("index.html",showcase = "Fair, The Predicted Value is
"+str(y_pred))
    elif(y_pred>=45 and y_pred<=64):
        return render_template("index.html",showcase = "Marginal, The Predicted
Value is "+str(y_pred))
    else:
        return render_template("index.html",showcase = "Poor, The Predicted Value is
"+str(y_pred))
if __name__ == '__main__':
    app.run(debug = True,port = 5000)
```

Flask Connectivity

```
import numpy as np
from flask import Flask,render_template,request
import pickle
import requests
API_KEY = "THZC3nURvpmMSRhpBOMUNeqJsJ6p40DZ4pp2FSnNkWfY"
token_response = requests.post('https://iam.cloud.ibm.com/identity/token',
data={"apikey":
 API_KEY, "grant_type": 'urn:ibm:params:oauth:grant-type:apikey'})
mltoken = token_response.json()["access_token"]
header = {'Content-Type': 'application/json', 'Authorization': 'Bearer ' + mltoken}
app = Flask(__name__)
model = pickle.load(open('wqi.pkl','rb'))
@app.route('/',methods=['GET'])
def home():
```

```
    return render_template("index.html")
@app.route('/login',methods = ['POST'])
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)]]
    payload_scoring = {"input_data": [{"fields": [['year','do','ph','co','bod','na','tc']],
"values": total}]}
    response_scoring = requests.post('https://us-
south.ml.cloud.ibm.com/ml/v4/deployments/735973ab-d35c-4182-90f9-
ca418497ced0/predictions?version=2022-11-18', json=payload_scoring,
    headers={'Authorization': 'Bearer ' + mltoken})
    print("Scoring response")
    print(response_scoring.json())
    predictions=response_scoring.json()
    y_pred=predictions['predictions'][0]['values'][0][0]
    if(y_pred >= 95 and y_pred <=100):
        return render_template("index.html",showcase = "Excellent, The Predicted
Value is "+str(y_pred))
    elif(y_pred >=89 and y_pred <=94):
        return render_template("index.html",showcase = "Very Good, The Predicted
Value is "+str(y_pred))
    elif(y_pred >=80 and y_pred <=88):
        return render_template("index.html",showcase = "Good, The Predicted Value is
"+str(y_pred))
    elif(y_pred>=65 and y_pred<=79):
```

```
        return render_template("index.html",showcase = "Fair, The Predicted Value is
"+str(y_pred))
    elif(y_pred>=45 and y_pred<=64):
        return render_template("index.html",showcase = "Marginal, The Predicted
Value is "+str(y_pred))
    else:
        return render_template("index.html",showcase = "Poor, The Predicted Value is
"+str(y_pred))
if __name__ == '__main__':
    app.run(debug = True,port = 5000)
```

Login.php

```
<?php
include 'config.php';
session_start();
error_reporting(0);
if (isset($_SESSION['username'])) {
   header("Location: predict.php");
}
if (isset($_POST['submit'])) {
   $email = $_POST['email'];
   $password = md5($_POST['password']);
   $sql = "SELECT * FROM members WHERE email='$email' AND
password='$password'";
   $result = mysqli_query($conn, $sql);
   if ($result->num_rows > 0) {
      $row = mysqli_fetch_assoc($result);
      $_SESSION['username'] = $row['username'];
      header("Location:predict.php");
   } else {
```

```
        echo "<script>alert('Woops! Email or Password is Wrong.')</script>";
    }
}
?>
<!DOCTYPE html>
<html>
<head>
    <meta charset="utf-8">
    <meta name="viewport" content="width=device-width, initial-scale=1.0">
    <link rel="stylesheet" href="https://stackpath.bootstrapcdn.com/font-awesome/4.7.0/css/font-awesome.min.css">
    <link rel="stylesheet" type="text/css" href="style.css">
    <title>LogIn</title>
</head>
<body>
    <div class="container">
        <form action="" method="POST" class="login-email">
            <p class="login-text" style="font-size: 2rem; font-weight: 800;">Login</p>
            <div class="input-group">
                <input type="email" placeholder="Email" name="email" value="<?php echo $email; ?>" required>
            </div>
            <div class="input-group">
                <input type="password" placeholder="Password" name="password" value="<?php echo $_POST['password']; ?>" required>
            </div>
            <div class="input-group">
                <button name="submit" class="btn">Login</button>
            </div>
            <p class="login-register-text">Don't have an account? <a href="register.php">Register Here</a>.</p>
```

```
        </form>
    </div>
</body>
</html>
```

Config.php

```
<?php
include 'config.php';
session_start();
error_reporting(0);
if (isset($_SESSION['username'])) {
    header("Location: predict.php");
}
if (isset($_POST['submit'])) {
    $email = $_POST['email'];
    $password = md5($_POST['password']);
    $sql = "SELECT * FROM members WHERE email='$email' AND
password='$password'";
    $result = mysqli_query($conn, $sql);
    if ($result->num_rows > 0) {
        $row = mysqli_fetch_assoc($result);
        $_SESSION['username'] = $row['username'];
        header("Location:predict.php");
    } else {
        echo "<script>alert('Woops! Email or Password is Wrong.')</script>";
    }
}
?>
<!DOCTYPE html>
<html>
```

```
<head>
    <meta charset="utf-8">
    <meta name="viewport" content="width=device-width, initial-scale=1.0">
    <link rel="stylesheet" href="https://stackpath.bootstrapcdn.com/font-awesome/4.7.0/css/font-awesome.min.css">
    <link rel="stylesheet" type="text/css" href="style.css">
    <title>LogIn</title>
</head>
<body>
    <div class="container">
        <form action="" method="POST" class="login-email">
            <p class="login-text" style="font-size: 2rem; font-weight: 800;">Login</p>
            <div class="input-group">
                <input type="email" placeholder="Email" name="email" value="<?php echo $email; ?>" required>
            </div>
            <div class="input-group">
                <input type="password" placeholder="Password" name="password" value="<?php echo $_POST['password']; ?>" required>
            </div>
            <div class="input-group">
                <button name="submit" class="btn">Login</button>
            </div>
            <p class="login-register-text">Don't have an account? <a href="register.php">Register Here</a>.</p>
        </form>
    </div>
</body>
</html>
```

Register.php

```php
<?php
include 'config.php';
error_reporting(0);
session_start();
if (isset($_SESSION['username'])) {
    header("Location: predict.html");
}
if (isset($_POST['submit'])) {
    $username = $_POST['username'];
    $email = $_POST['email'];
    $password = md5($_POST['password']);
    $cpassword = md5($_POST['cpassword']);
    if ($password == $cpassword) {
        $sql = "SELECT * FROM login WHERE email='$email'";
        $result = mysqli_query($conn, $sql);
        if (!$result->num_rows > 0) {
            $sql = "INSERT INTO login (username, email, password)
                    VALUES ('$username', '$email', '$password')";
            $result = mysqli_query($conn, $sql);
            if ($result) {
                echo "<script>alert('Wow! User Registration Completed.')</script>";
                header("Location: predict.html");
                $username = "";
                $email = "";
                $_POST['password'] = "";
                $_POST['cpassword'] = "";
            }
            else {
                echo "<script>alert('Woops! Something Wrong Went.')</script>";
```

```
            }
        } else {
            echo "<script>alert('Woops! Email Already Exists.')</script>";
        }
        } if ($password != $cpassword) {
        echo "<script>alert('Password Not Matched.')</script>";
    }
}
?>
<!DOCTYPE html>
<html>
<head>
    <meta charset="utf-8">
    <meta name="viewport" content="width=device-width, initial-scale=1.0">
    <link rel="stylesheet" href="https://stackpath.bootstrapcdn.com/font-
awesome/4.7.0/css/font-awesome.min.css">
    <link rel="stylesheet" type="text/css" href="style.css">
    <title>Register</title>
</head>
<body>
    <div class="container">
        <form action="" method="POST" class="login-email">
            <p class="login-text" style="font-size: 2rem; font-weight:
800;">Register</p>
            <div class="input-group">
                <input type="text" placeholder="Username" name="username"
value="<?php echo $username; ?>" required>
            </div>
            <div class="input-group">
                <input type="email" placeholder="Email" name="email" value="<?php
echo $email; ?>" required>
```

```
            </div>
            <div class="input-group">
                <input type="password" placeholder="Password" name="password"
value="<?php echo $_POST['password']; ?>" required>
            </div>
            <div class="input-group">
                <input type="password" placeholder="Confirm Password"
name="cpassword" value="<?php echo $_POST['cpassword']; ?>" required>
            </div>
            <div class="input-group">
                <button name="submit" class="btn">Register</button>
            </div>
            <p class="login-register-text">Have an account? <a href="index.php">Login
Here</a>.</p>
        </form>
    </div>
</body>
</html>
```

Parameter.html

```
<!DOCTYPE html>

<html>

  <head>

    <title>Efficient Water Quality Anaysis</title>

    <link href="https://fonts.googleapis.com/css?family=Roboto:300,400,500,700"
    rel="stylesheet">
```

```
<link rel="stylesheet" href="https://use.fontawesome.com/releases/v5.5.0/css/all.css" integrity="sha384-B4dIYHKNBt8Bc12p+WXckhzcICo0wtJAoU8YZTY5qE0Id1GSseTk6S+L3BlXeVIU" crossorigin="anonymous">

<style>

 html, body {

 min-height: 100%;

 }

 body, div, form, input, select,textarea, p {

 padding: 0;

 margin: 0;

 outline: none;

 font-family: Roboto, Arial, sans-serif;

 font-size: 14px;

 color: #666;

 line-height: 22px;

 }

 h1 {

 position: absolute;

 margin: 0;

 font-size: 32px;

 color: #fff;
```

```
z-index: 2;
}
h5 {
margin: 10px 0;
}
.testbox {
display: flex;
justify-content: center;
align-items: center;
height: inherit;
padding: 400px;
}
form {
width: 100%;
padding: 20px;
border-radius: 6px;
background: #fff;
box-shadow: 0 0 20px 0 #095484;
}
.banner {
position: relative;
```

```
height: 210px;
background-image: url("bg.png");
background-size: cover;
display: flex;
justify-content: center;
align-items: center;
text-align: center;
}
.banner::after {
content: "";
background-color: rgba(0, 0, 0, 0.5);
position: absolute;
width: 100%;
height: 100%;
}
input, select, textarea {
margin-bottom: 10px;
border: 1px solid #ccc;
border-radius: 3px;
}
input {
```

```
  width: calc(100% - 10px);
  padding: 5px;
  }
  select {
  width: 100%;
  padding: 7px 0;
  background: transparent;
  }
  textarea {
  width: calc(100% - 12px);
  padding: 5px;
  }
  .item:hover p, .item:hover i, .question:hover p, .question label:hover,
input:hover::placeholder, a {
  color: #095484;
  }
  .item input:hover, .item select:hover, .item textarea:hover {
  border: 1px solid transparent;
  box-shadow: 0 0 6px 0 #095484;
  color: #095484;
  }
```

```
.item {
position: relative;
margin: 10px 0;
}
input[type="date"]::-webkit-inner-spin-button {
display: none;
}
.item i, input[type="date"]::-webkit-calendar-picker-indicator {
position: absolute;
font-size: 20px;
color: #a9a9a9;
}
.item i {
right: 2%;
top: 30px;
z-index: 1;
}
.question span {
margin-left: 30px;
}
span.required {
```

```
margin-left: 0;
color: red;
}
}
.btn-block {
margin-top: 10px;
text-align: center;
}
button {
width: 150px;
padding: 10px;
border: none;
border-radius: 5px;
background: #095484;
font-size: 16px;
color: #fff;
cursor: pointer;
}
button:hover {
background: #0666a3;
}
```

```
    @media (min-width: 568px) {
    .city-item {
    display: flex;
    flex-wrap: wrap;
    justify-content: space-between;
    }
    .city-item input {
    width: calc(50% - 20px);
    }
    .city-item select {
    width: calc(50% - 8px);
    }
    }
  </style>
</head>
<body>
  <div class="testbox">
  <form action="/">
    <div class="banner">
      <h1>Water Quality Analysis</h1>
    </div>
```

```
    <div class="item">
      <p>Temperature</p>
      <input type="text" name="temp" required/>
    </div>
    <div class="item">
      <p>DO</p>
      <input type="text" name="DO" required/>
    </div>
  <div class="item">
   <p>pH</p>
      <input type="text" name="PH" required/>
    </div>
    <div class="item">
      <p>Conductivity</p>
      <input type="text" name="conductivity" required/>
    </div>
  <div class="item">
   <p>BOD</p>
      <input type="text" name="BOD" required/>
    </div>
    <div class="item">
```

```
    <p>NI</p>
    <input type="text" name="NI" required/>
  </div>

<div class="item">
    <p>Fec_col</p>
    <input type="text" name="Fec_col" required/>
  </div>
<div class="item">
 <p>Tot_col</p>
    <input type="text" name="Tot_col" required/>
  </div>
  <div class="item">
    <p>year</p>
    <input type="text" name="year" required/>
  </div>
    <div class="btn-block">
      <button type="submit" href="/" >SUBMIT</button>
    </div>
 </form>
 </div>
```

```
  </body>
 </html>
```

ContactPage.html

```
<!DOCTYPE html>
<html>
 <head>
  <title>Customer Support</title>
  <link rel="stylesheet" href="https://use.fontawesome.com/releases/v5.4.1/css/all.css" integrity="sha384-5sAR7xN1Nv6T6+dT2mhtzEpVJvfS3NScPQTrOxhwjIuvcA67KV2R5Jz6kr4abQsz" crossorigin="anonymous">
  <link href="https://fonts.googleapis.com/css?family=Roboto:300,400,500,700" rel="stylesheet">
  <style>
   html, body {
   min-height: 100%;
   padding: 0;
   margin: 0;
   font-family: Roboto, Arial, sans-serif;
   font-size: 14px;
   color: #666;
   }
```

```
h1 {
margin: 0 0 20px;
font-weight: 400;
color: #1c87c9;
}
p {
margin: 0 0 5px;
}
.main-block {
display: flex;
flex-direction: column;
justify-content: center;
align-items: center;
min-height: 100vh;
background: CadetBlue;
}
form {
padding: 25px;
margin: 25px;
box-shadow: 0 2px 5px AntiqueWhite;
background: Beige;
```

```
}
.fas {
margin: 25px 10px 0;
font-size: 72px;
color: #fff;
}
.fa-envelope {
transform: rotate(-20deg);
}
.fa-at , .fa-mail-bulk{
transform: rotate(10deg);
}
input, textarea {
width: calc(100% - 18px);
padding: 8px;
margin-bottom: 20px;
border: 1px solid #1c87c9;
outline: none;
}
input::placeholder {
color: #666;
```

```
}
button {
width: 100%;
padding: 10px;
border: none;
background: #1c87c9;
font-size: 16px;
font-weight: 400;
color: #fff;
}
button:hover {
background: #2371a0;
}
@media (min-width: 568px) {
.main-block {
flex-direction: row;
}
.left-part, form {
width: 50%;
}
.fa-envelope {
```

```
    margin-top: 0;
    margin-left: 20%;
    }
    .fa-at {
    margin-top: -10%;
    margin-left: 65%;
    }
    .fa-mail-bulk {
    margin-top: 2%;
    margin-left: 28%;
    }
    }
  </style>
</head>
<body>
  <div class="main-block">
    <div class="left-part">
      <i class="fas fa-envelope"></i>
      <i class="fas fa-at"></i>
      <i class="fas fa-mail-bulk"></i>
    </div>
```

```
      <form action="/">
        <h1>Contact Us</h1>
        <div class="info">
          <input class="fname" type="text" name="name" placeholder="Full name">
          <input type="text" name="name" placeholder="Email">
          <input type="text" name="name" placeholder="Phone number">
        </div>
        <p>Message</p>
        <div>
          <textarea rows="4"></textarea>
        </div>
        <button type="submit" href="/">Submit</button>
      </form>
    </div>
  </body>
</html>
```

Logout.html

```
<!DOCTYPE html>
<head>
<meta charset="utf-8">
<link href="logstyle.css" rel="stylesheet">
```

```
 <link href="log.js" rel="javascript">

<title>Logout</title>

<meta name="description" content="">

<meta name="viewport" content="width=device-width">

</head>

<body ng-app="logoutApp" ng-controller="logoutController">

<div class="background-photo">

<div class="jumbotron">

<div class="container">

<h1>See you soon!</h1>

</div>

</div>

<div class="middle-block">

<div ng-if="!!loadingShowed">

You will be redirect in {{seconds}} seconds.

</div >

<div ng-if="!loadingShowed">

Please, click below.

</div>

<div class="round-class">
```

```
<i ng-class="{'fa fa-3x':true, 'fa-spinner fa-pulse fa-fw':!!loadingShowed, 'fa-sign-in':!loadingShowed}" ></i>

<span class="sr-only">Loading...</span>

</div>

</div>

<div class="second">

<div class="container">

<div class="col-xs-12 col-sm-6">

<div class="row">

<div class="round-class ball">

<i class="fa fa-thumbs-o-up fa-lg" aria-hidden="true"></i></div>

<div class="right-text">

Thanks to using our service!

</div>

</div>

<div class="row">

<div class="round-class ball">

<i class="fa fa-mobile fa-2x" aria-hidden="true"></i></div>

 <div class="right-text">

connect with us through customer support for any queries

  </div></div></div></div></div></div>
```

</body>

</html>

LINKS:

GITHUB - https://github.com/IBM-EPBL/IBM-Project-15632-1659602130

VIDEO LINK - https://youtu.be/zpINgI3LiPo