PROJECT REPORT

EFFICIENT WATER QUALITY ANALYSIS AND PREDICTION USING MACHINE LEARNING

By team – PNT2022TMID29641

Thanthai Periyar Government Institute of Technology

Abirami A (513119106003)
Abinaya S (513119106002)
Anbarasi C (513119106007)
Hariharan M (513119106501)

1.INTRODUCTION

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.

1.1 PROJECT OVERVIEW

In this project, we are going to implement a water quality prediction using machine learning techniques. In this technique, our model predicts that the water is safe to drink or not using some parameters like pH value, conductivity, hardness, dissolved oxygen etc, . In water environment research, models and conclusions derived from machine learning have been applied to the construction, monitoring, simulation, evaluation, and optimization of various water treatment and management systems. Additionally, machine learning can provide solutions for water pollution control, water quality improvement, and watershed ecosystem security management.

1.2.PURPOSE

The major goal of this project is to use machine learning techniques to predict water quality. Firstly, the water quality index is calculated using the weighted arithmetic index method. To predict the WQI, different regression algorithms are used. 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 project, we propose an alternative approach using machine learning techniques to predict water quality.

2. LITERATURE SURVEY

Paper 1 : Machine Learning Algorithms for Efficient Water Quality Prediction

Publication Year: 26 August 2021

Author: Mourade Azrour, Jamal Mabrouki, Ghizlane Fattah, Azedine Guezzaz & Faissal Aziz

Journal Name: Original Article Summary: Water is an essential resource for human existence. In fact, more than 60% of the human body is made up of water. In recent years, water pollution has become a serious problem affecting water quality. Therefore, to design a model that predicts water quality which is very important to control water pollution, as well as to alert users in case of poor quality detection. This method 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

Paper 2 : Water Quality Factor Prediction Using Supervised Machine Learning

Publication Year: 2018

Author: Kathleen Joslyn

Journal Name: Research Experiences for Undergraduates on Computational Modeling Serving the City

Summary: This method is used 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.

Paper 3 : Predicting and Analyzing Water Quality using Machine Learning

Publication Year: 2016

Author: Yafra Khan, Chai Soo See

Journal Name: IEEE

Summary: Water is the most important thing for human being. The effects of un-clean water are leads to many health issues and impacting every aspect of life. Therefore, management of water resources is very crucial in order to optimize the quality of water. The effects of water contamination can be tackled efficiently if data is analyzed and water quality is predicted beforehand. The goal of this method is to develop a water quality prediction model with the help of water quality factors using Artificial Neural Network (ANN) and time-series analysis.

Paper 4 : Efficient Prediction of Water Quality Index (WQI) Using Machine Learning Algorithms

Publication year: 2021

Author: Md. Mehedi Hassan, Laboni Akter, Mushfiqur Rahman

Journal name: Research Article

Summary: 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 method is 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 obtained 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.

Paper 5 : Water Quality Prediction using Machine Learning Method

Publication Year: 19 January 2018

Author: Ali Heidar Nasrolahi, Amir Hamze Haghiabi

Journal Name: Water Quality Research Journal

Summary: This method investigates the performance of artificial intelligence techniques including artificial neural network (ANN), group method of data handling (GMDH) and support vector machine (SVM) for predicting water quality components. To develop the ANN and SVM, different types of transfer and kernel functions were tested, respectively. The evaluation of the accuracy of the applied models according to the error indexes declared that SVM was the most accurate model.

Paper 6 : Ground Water Quality Prediction using Machine Learning Algorithms

Publication year: 25 January 20

Author: S. Vijay & Dr. K. Kamaraj

Journal Name: International Journal of Research and Analytical Reviews

Summary: 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 physio-chemical

characteristics of ground water quality. The bore well from which the samples were collected are extensively used for drinking purpose. The water quality parameters such as PH, TDC, EC, chloride, sulphate, nitrate, carbonate, bicarbonate, metal ions etc. This method focused on predicting water quality by using Machine learning classifier algorithm C5.0.

Paper 7 : A Real Time Water Quality Monitoring Using Machine Learning Algorithm

Publication Year: 2020

Author: S. Angel Vergina, Dr. S. Kayalvizhi, Dr. R.M. Bhavadharini

Journal Name: European Journal of Molecular & Clinical Medicine

Summary: Water quality parameter is of much importance in our day to day lives. Prediction of water quality will help to reduce water pollution and guard our human health. An intelligent process of monitoring the quality of water automatically detects the condition of water through IoT by processing sensors data and instantly provides notification to water analyst, when the quality of water is abnormal. With the initiation of Machine to Machine Communication analyzing and communicating the data becomes simple and efficient.

Paper 8 : Water Quality Monitoring System Using IoT and Machine Learning

Publication Year: 2018 Author: Nikhil Kumar Koditala, Purnendu Shekar Pandey

Journal name: IEEE

Summary: World Economic Forum ranked drinking water crisis as one of the global risk, due to which around 200 children are dying per day. Drinking unsafe water alone causes around 3.4 million deaths per year. Despite the advancements in technology, sufficient quality measures are not present to measure the quality of drinking water. By focusing on the above issue, this method proposes a low-cost water quality monitoring system using emerging technologies such as IoT, Machine Learning and Cloud Computing which can replace traditional way of quality monitoring. This helps in saving people of rural areas from various dangerous diseases such as fluorosis, bone deformities etc. This method also has a capacity to control temperature of water and adjusts it so as to suit environment temperature.

Paper 9 : Prediction of water quality parameters using machine learning models

Publication Year: 03 June 2021

Author: Atefeh Nouraki , Mohammad Alavi, Mona Golabi

Journal Name: Environmental Science and Pollution Research

Summary: Accurate water quality predicting has an essential role in improving water management and pollution control. The machine learning models have been successfully implemented for modelling total dissolved solids (TDS), sodium absorption ratio (SAR) and total hardness (TH) content in aquatic ecosystems with insufficient data. However, due to multiple pollution sources and complex behaviors of pollutants, this method is used to predict the TDS, SAR, and TH levels in water.

Paper 10 : Smart IoT and Machine Learning-based Framework for Water Quality Assessment and Device Component Monitoring

Publication Year: 14 February 2022

Author: Akashdeep Bhardwaj, Vishal Dagar, Muhammad Owais Khan

Journal Name: Research Gate

Summary: Water is the most important natural element present on earth for humans, yet the availability of pure water is becoming scarce and decreasing. An increase in population and rise in temperatures are two major factors contributing to the water crisis worldwide. Sensors, embedded and smart devices in water plants require proactive monitoring for optimal performance. This methods presents an IoT-based real-time framework to perform water quality management, monitor, and alert for taking actions based on contamination and toxic parameter levels, device and application performance as the first part of the proposed work. Machine learning models analyze water quality trends and device monitoring and management architecture.

Paper 11 : WATER QUALITY PREDICTION USING MACHINE LEARNING

Publication Year: 05 May 2022

Author: Sai Sreeja Kurra, Sambangi Geethika Naidu

Journal Name: International Research Journal of Modernization in Engineering Technology and Science

Summary: The major goal of this method is to use machine learning techniques to measure water quality. The following water quality parameters were utilised to assess the overall water quality in terms of pH, Hardness, Solids, Chloramines, Sulfate, Conductivity, Organic Carbon, Trihalomethanes, Turbidity. To estimate the water quality class, the

method used two types of classification algorithms: Decision Tree (DT)and K- Nearest Neighbor (KNN).

Paper 12 : Machine Learning Methods for Better Water Quality Prediction

Publication Year: November 2019

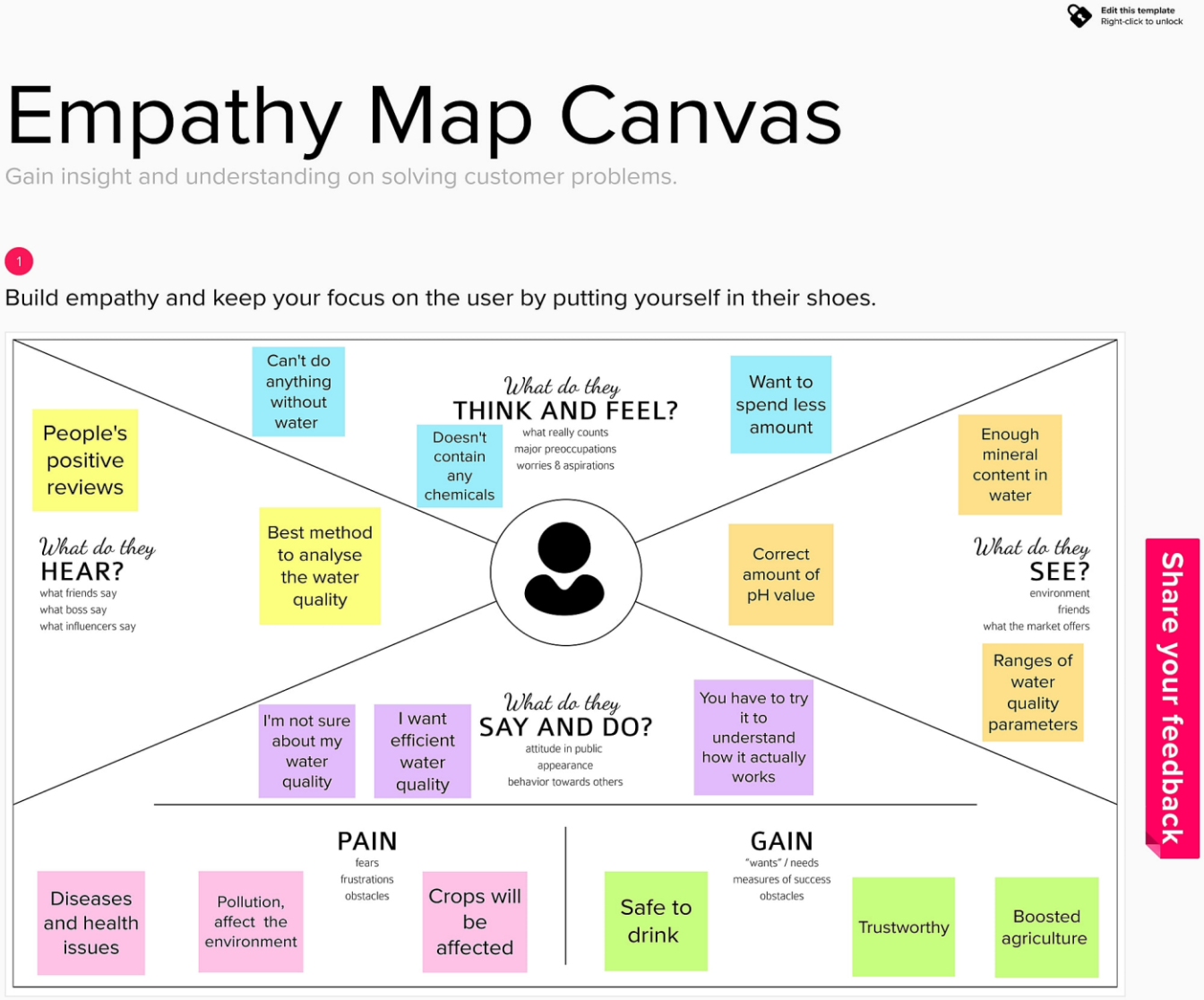
Author: Ali Najah Ahmed, Faridah Binti Othman

Journal Name: Journal of Hydrology

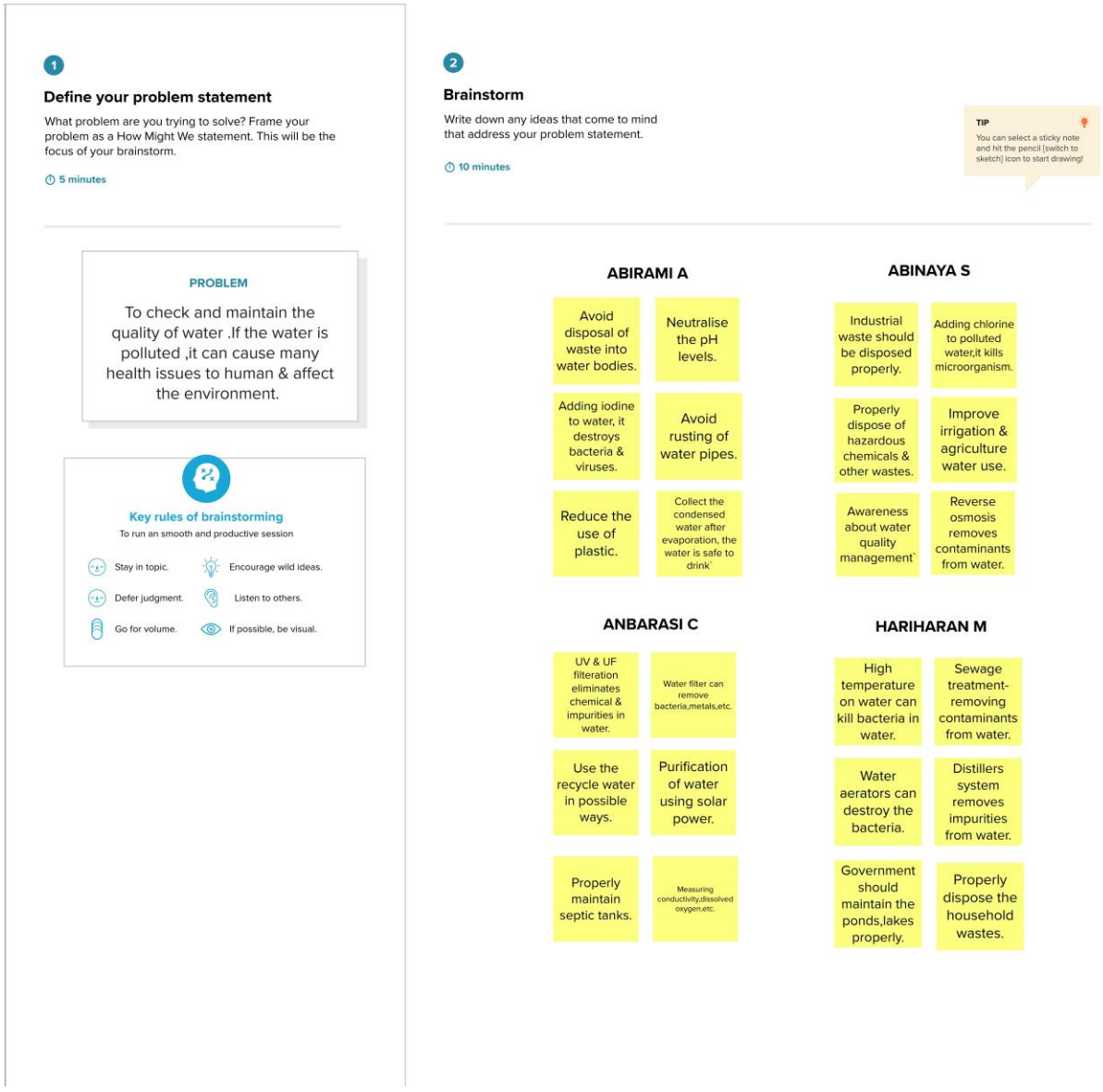
Summary: In any aquatic system analysis, the modelling water quality parameters are of considerable significance. The traditional modelling methodologies are dependent on datasets that involve large amount of unknown or unspecified input data and generally consist of time consuming processes. Therefore, setting up of a water quality prediction model for better water resource management is of critical importance and will serve as a powerful tool. The different modelling approaches that have been implemented include: Adaptive Neuro-Fuzzy Inference System (ANFIS), Radial Basis Function Neural Networks (RBF-ANN), and Multi? Layer Perceptron Neural Networks (MLP-ANN). In the domain of interests, the water quality parameters primarily include ammoniacal nitrogen (AN), suspended solid (SS) and pH.

3. IDEATION & PROPOSED SOLUTION

3.1 EMPATHY MAP CANVAS



3.2 IDEATION AND BRAINSTORMING

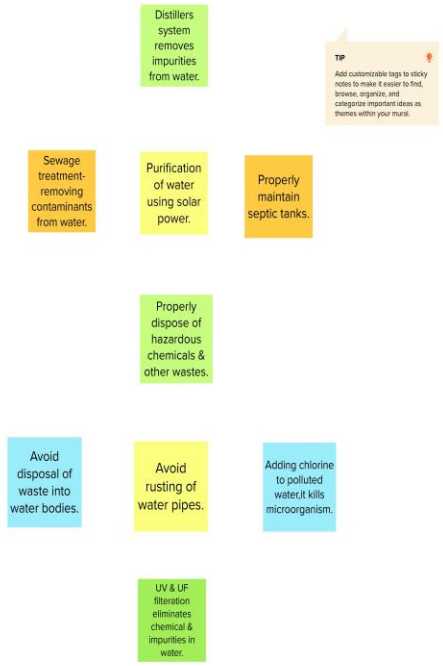


3

Group ideas

Take turns sharing your ideas while clustering similar or related notes as you go. Once all sticky notes have been grouped, give each cluster a sentence-like label. If a cluster is bigger than six sticky notes, try and see if you and break it up into smaller sub-groups.

20 minutes

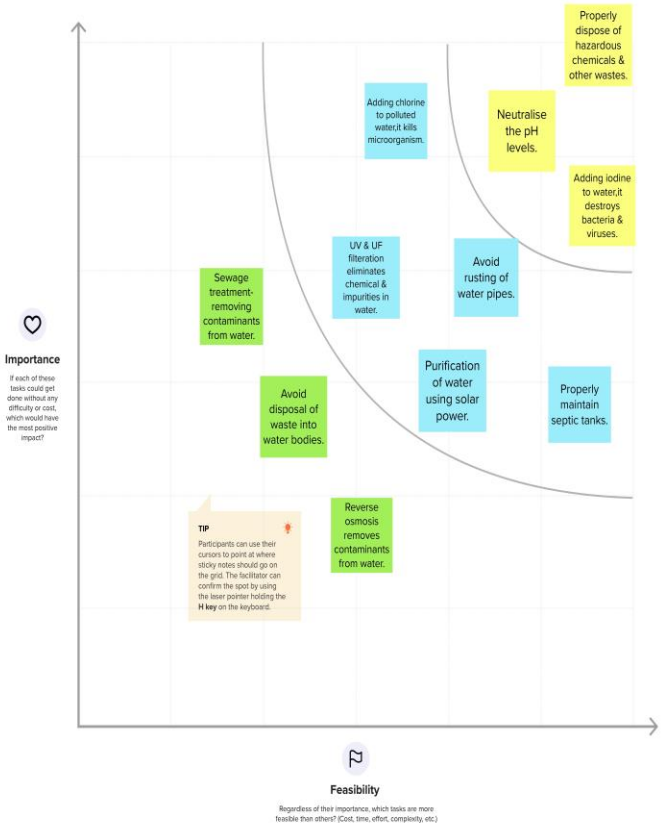


4

Prioritize

Your team should all be on the same page about what's important moving forward. Place your ideas on this grid to determine which ideas are important and which are feasible.

20 minutes



3.3 PROPOSED SOLUTION

S. No.	Parameter	Description
1.	Problem Statement (Problem to be solved)	Water is polluted by several ways. Due to rapidly growing urbanization, the pure and safe drinking water supply is a big challenge in urban areas. In India still most of the people use simple water purifier or filter, that is may not be purified. Contaminants in water can cause a health issues for human.
2.	Idea / Solution description	It is difficult to check the quality of water practically at every time. To make it easy, we can use automatic real time monitoring system to analyse and monitor the quality of water. By using Water Quality Index (WQI), we can analyse and predict the water is safe to drink or not. We can measure the water quality parameters such as pH, turbidity, etc.
3.	Novelty / Uniqueness	The uniqueness of this method, when the water is unsafe to drink, it alerts or send messages to the user or customer.
4.	Social Impact / Customer Satisfaction	In this method, Customer can immediately identify abrupt changes in the monitored water quality parameters, and there is a possible to minimize the eradicate risks and dangers.
5.	Business Model (Revenue Model)	This method is a cost efficient approach for fast and accurate measurement of water quality. To analyse the water quality time consuming is less.
6.	Scalability of the Solution	In this method, monitoring process is easy. Customer or user can easily check the quality of water.

3.4 PROBLEM SOLUTION FIT

Define CS, fit into

1. CUSTOMER SEGMENT(S) — CS

Who is your customer?
i.e. working parents of 0-5 y.o. kids.

People of Urban areas who looking for pure and quality water.

6.CUSTOMER CONSTRAINTS — CC

What constraints prevent your customers from taking action or limit their choices of solutions? i.e. spending power, budget, no cash, network connection, available devices.

- Lack of knowledge about water quality standards.
- Checking or testing the water quality in laboratories may be accurate or not.

5. AVAILABLE SOLUTIONS — AS

Which solutions are available to the customers when they face the problem or need to get the job done? What have they tried in the past? What pros & cons do these solutions have? i.e. pen and paper is an alternative to digital notetaking.

Finding the Water Quality Index(WQI).

PROS:

- Measure the water quality parameter such as pH,hardness,etc.
- Simplifies the complex dataset into easily understandable.

CONS:

- Doesn't provide the absolute measure of degree of pollution.

Explore AS, differentiate

Focus on J&P, tap into BE,

2. JOBS-TO-BE-DONE / PROBLEMS — J&P

Which jobs-to-be-done (or problems) do you address for your customers? There could be more than one; explore different sides.

JOBS – TO – BE DONE:

- To analyse and predict the water quality parameters.
- Check whether the water is safe to drink or not.

PROBLEMS:

- Contaminants in water can cause a major health issues.
- Ground water will be affected.

9. PROBLEM ROOT CAUSE — RC

What is the real reason that this problem exists? What is the back story behind the need to do this job?i.e. customers have to do it because of the change in regulations.

- Improper disposal of industrial and household waste.
- Usage of plastic.
- Global warming.
- Highly usage of inorganic fertilizers.

7. BEHAVIOUR — BE

What does your customer do to address the problem and get the job done? i.e. directly related: find the right solar panel installer, calculate usage and benefits; indirectly associated: customers spend free time on volunteering work (i.e. Greenpeace)

- Try to get a pure water in the possible ways.
- Ensure the quality of water.

Focus on J&P, tap into BE,

Identify strong TR & EM

3. TRIGGERS — TR

What triggers customers to act? i.e. seeing their neighbour installingsolar panels, reading about a more efficient solution in the news.

- Insufficient pure drinking water supply.
- Spread of water – borne diseases.

4. EMOTIONS: BEFORE / AFTER — EM

How do customers feel when they face a problem or a job and afterwards? i.e. lost, insecure > confident, in control - use it in your communication strategy & design.

BEFORE:

- Insecure.
- Feeling frustrated.

AFTER:

- Feeling safe.
- Hygienic.

10. YOUR SOLUTION — SL

If you are working on an existing business, write down your current solution first,fill in the canvas, and check how much it fits reality.
If you are working on a new business proposition, then keep it blank until you fill inthe canvas and come up with a solution that fits within customer limitations, solves a problem and matches customer behavior.

By collecting the water samples from a water resources , then the gathered data is used to analyse and predict the quality of water using Machine Learning techniques.

8.CHANNELS of BEHAVIOUR — CH

8.1 ONLINE

What kind of actions do customers take online? Extract online channels from #7

Customer can use the web application to process the data.

8.2 OFFLINE

What kind of actions do customers take offline? Extract offline channels from #7and use them for customer development.

- Using water filter or purifiers.
- Boiling the drinking water.

Extract online & offline CH of BE

4. REQUIREMENT ANALYSIS

4.1 FUNCTIONAL REQUIREMENT

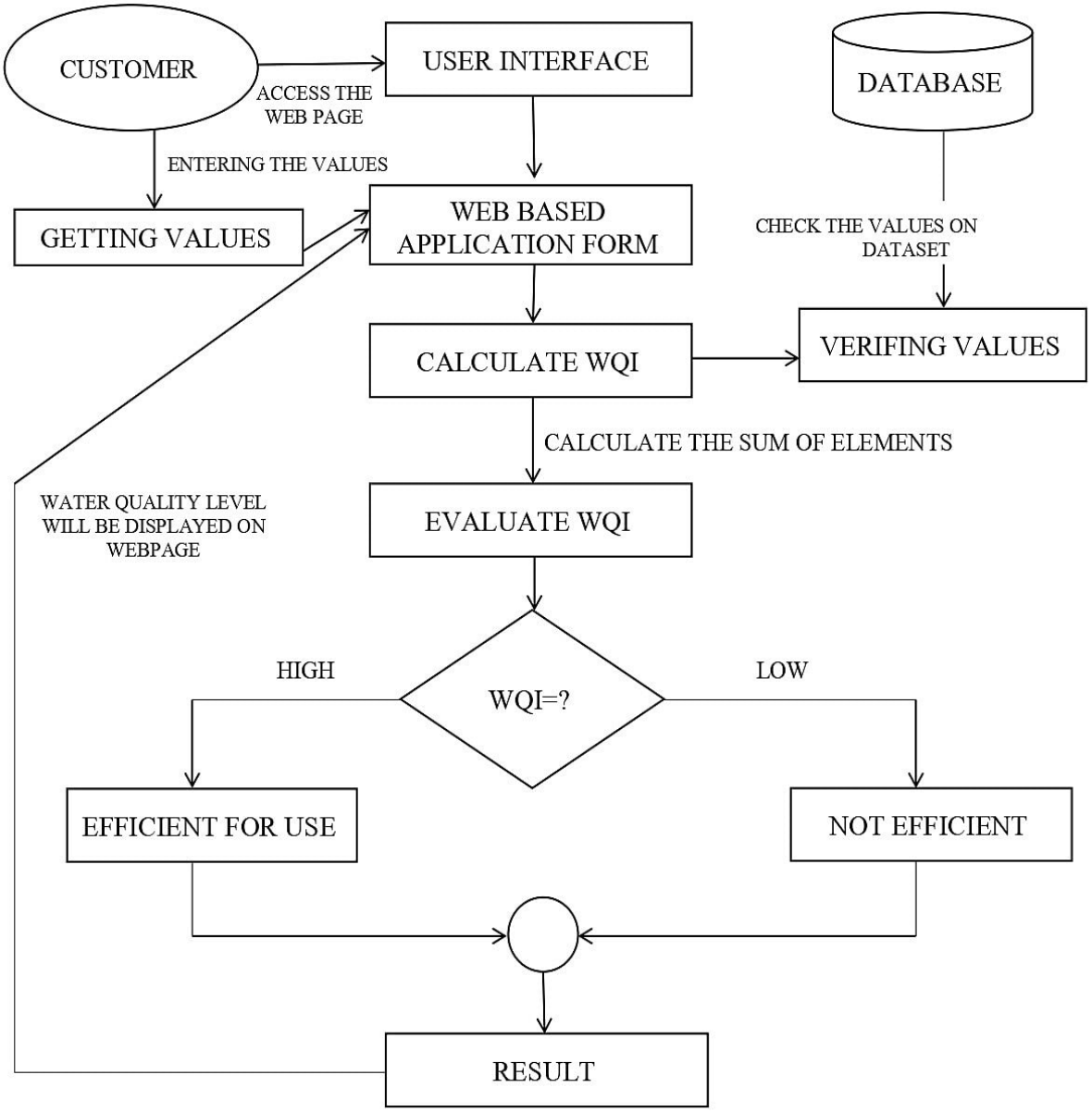
FR No.	Functional Requirement (Epic)	Sub Requirement (Story / Sub-Task)
FR-1	Data collection	Collect the data from various water resources.
FR-2	Data Pre-processing	Collected data can be pre-processed using machine learning algorithm.
FR-3	Model Building	Build and train the model on IBM Cloud.
FR-4	Quality Analysis	User can analyse the water quality using some parameters such as pH, dissolved oxygen, conductivity, etc.
FR-5	Reporting	Result of the water quality analysis will be displayed to the user.

4.2 NON - FUNCTIONAL REQUIREMENT

NFR No.	Non-Functional Requirement	Description
NFR-1	Usability	User can analyse and predict the quality of water.
NFR-2	Security	It is highly secured as each process is verified.
NFR-3	Reliability	It accurately reports the uncertainty in the prediction.
NFR-4	Performance	Analysing or predicting the quality of water will be easy comparing manual method.
NFR-5	Availability	Analysis can be made at any time through online.
NFR-6	Scalability	User can measure and analyse the quality of water. It provides pollution free and pure water.

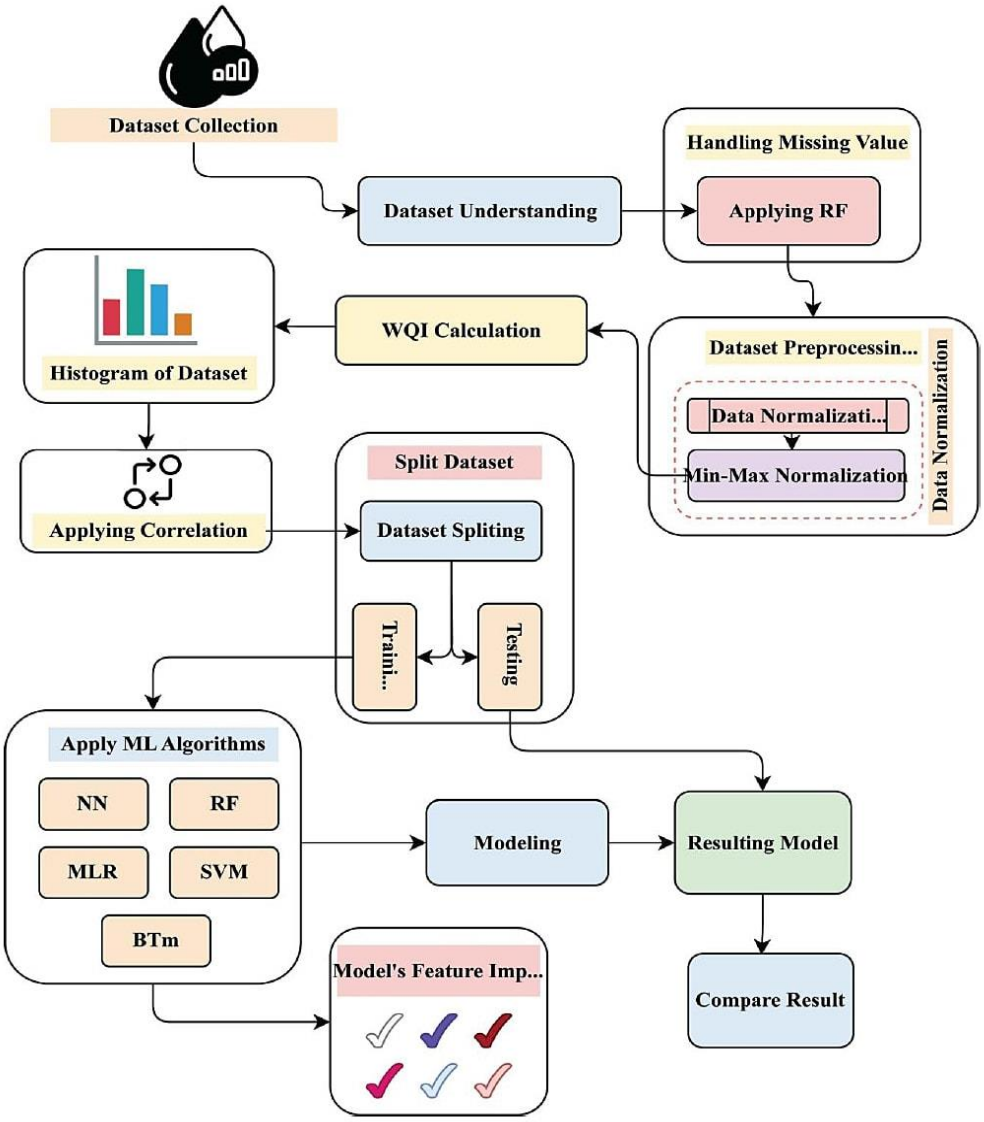
5.PROJECT DESIGN

5.1 DATA FLOW DIAGRAM

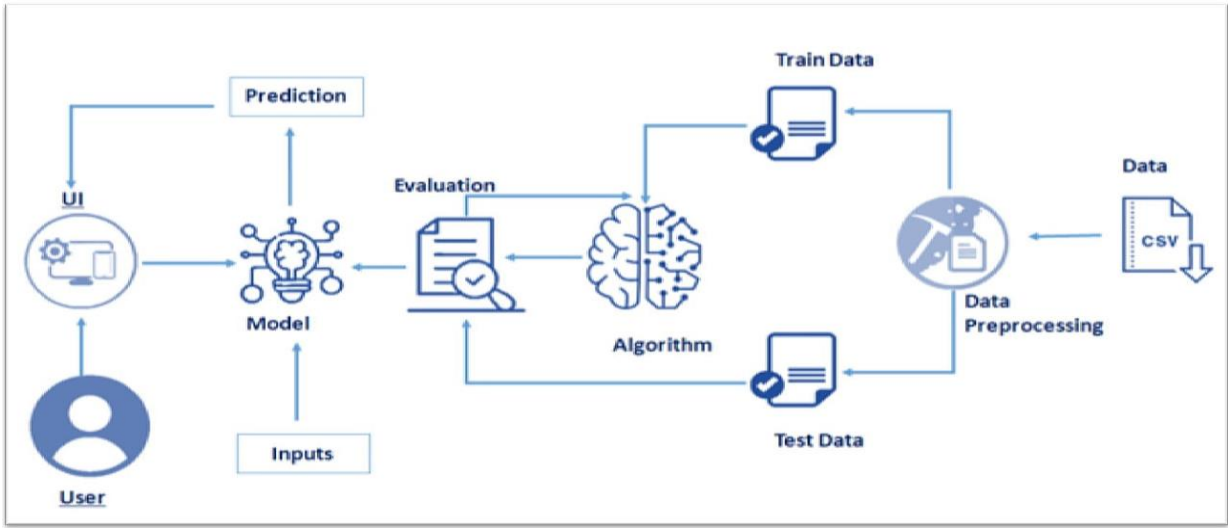


5.2 SOLUTION AND TECHNINCAL ARCHITECTURE

SOLUTION ARCHITECTURE



TECHNINCAL ARCHITECTURE



5.3 USER STORIES

User type	Functional requirement	User story no	User story/Task	Acceptance criteria	Priority	Release
Customer	Access web page	USN - 1	As a user, I can directly analyse the quality of water without login.	I can access the web page through at any time.	High	Sprint 1
Customer		USN - 2	As a user, I can access the application to analyse the water quality.	I can access the web page through at any time.	High	Sprint 1
Customer	Input Evaluation	USN - 3	As a user, I can enter values to predict the water quality.	I can see the quick evaluation.	High	Sprint 2
Customer	Prediction	USN – 4	As a user, I will know the quality of water by using prediction.	I can check the water quality is good or not.	High	Sprint 2
Customer	Result	USN - 5	As a user, I can see the result about water quality.	The result will be displayed on the web page.	High	Sprint 3

6. PROJECT PLANNING AND SCHEDULING

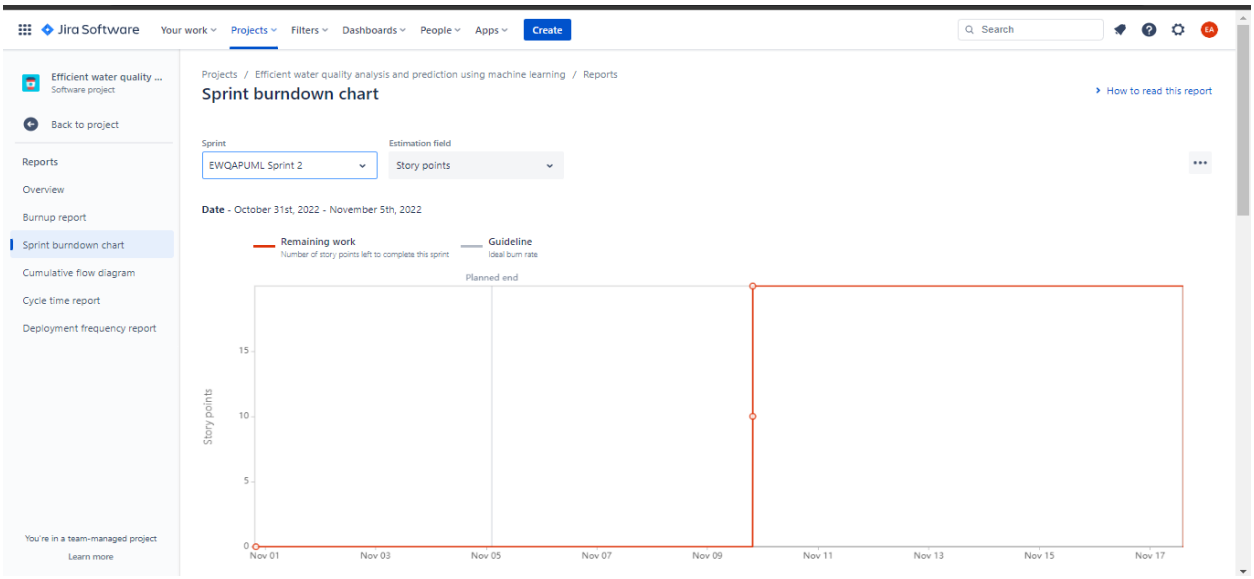
6.1 SPRINT PLANNING AND ESTIMATION

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint-1	Data collection	USN-1	Download or collect the dataset to perform pre-processing.	2	High	Abirami A
Sprint-1	Data-preprocessing	USN-2	Data pre-processing formats the data and handle the missing values.	1	High	Anbarasi C
Sprint-2	Model Building	USN-3	Calculate the Water Quality Index (WQI) using given formula for every parameter.	2	High	Hariharan M
Sprint-2	Training and Testing	USN-4	Training the model using ML algorithms and testing the performance of the model.	2	High	Abirami A
Sprint-3	IBM Deployment	USN-5	Deploy the model on IBM cloud.	2	High	Abinaya S
Sprint-4	Implementing the web page	USN-6	Implementing the web page for collecting the data from user. And it provides the details about quality of water.	2	High	Anbarasi C

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	20	6 Days	24 Oct 2022	29 Oct 2022	20	29 Oct 2022
Sprint-2	20	6 Days	31 Oct 2022	05 Nov 2022	20	05 Nov 2022
Sprint-3	20	6 Days	07 Nov 2022	12 Nov 2022	20	11 Nov 2022
Sprint-4	20	6 Days	14 Nov 2022	19 Nov 2022	20	19 Nov 2022

6.3 REPORTS FROM JIRA



7. CODING AND SOLUTIONING
7.1 FEATURE 1

The proposed system is the machine learning model where we could able to predict the quality of the water from giving the necessary details regarding the water body.In this proposed model, Random forest algorithm is used. With the dataset we will be finding out the water quality index and using that we split the data into the training and testing set. Then the model will be created using the splitted data. After the model is created the accuracy of the model will be determined and model is deployed in the pickle. There is also another method to deploy a model using the IBM cloud.

Code for Random Forest Algorithm

```
from sklearn.ensemble import RandomForestRegressor
regressor = RandomForestRegressor(n_estimators = 10, random_state = 0)
regressor.fit(X_train, Y_train)
y_pred = regressor.predict(X_test)
```

Code for deploy the model in pickle using flask

```
import numpy as np
from flask import Flask,render_template,request
import pickle
app= Flask(__name__)
model=pickle.load(open(r'new1.pkl','rb'))
@app.route('/')
def home() :
 return render_template("webpage.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"]

  tc = request.form["tc"]

  na = request.form["na"]

  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("webpage.html",showcase = 'Excellent, The Predicted Value is'
+ str(y_pred))

  elif(y_pred >= 89 and y_pred<=94):

  return render_template("webpage.html",showcase = 'Very Good, The Predicted Value
is' + str(y_pred))

  elif(y_pred >= 80 and y_pred<=88):

    return render_template("webpage.html",showcase = 'Good, The Predicted Value is' +
str(y_pred))

 elif(y_pred >= 65 and y_pred<=79):

   return render_template("webpage.html",showcase = 'Fair, The Predicted Value is'  +
str(y_pred))

 elif(y_pred >= 45 and y_pred<=64):

 return render_template("webpage.html",showcase = 'Marginal, The Predicted Value is' +
str(y_pred))

 else:

return render_template("webpage.html",showcase = 'Poor, The Predicted Value is' +
str(y_pred))

if __name__ == '__main__':

     app.run(debug = False)
```

7.2 FEATURE 2

Code for model deployed in IBM Cloud

```
!pip install ibm_watson_machine_learning

from ibm_watson_machine_learning import APIClient
wml_credentials={
   "url": "https://us-south.ml.cloud.ibm.com",
   "apikey": "hG98mgRIr5gqmayEXjjBeFTSE0WVRIX1y-GSVPognOpe"
}
wml_client=APIClient(wml_credentials)
wml_client.spaces.list()

SPACE_ID="285f980d-ce64-44b0-8e04-71605abb7bbe"

wml_client.set.default_space(SPACE_ID)

wml_client.software_specifications.list()

import sklearn
sklearn.__version__

MODEL_NAME= 'Deployment'
DEPLOYMENT_NAME ='demo'
```

```
DEMO_MODEL = regressor

software_spec_uid = wml_client.software_specifications.get_id_by_name('runtime-22.1-py3.9')

model_props={
wml_client.repository.ModelMetaNames.NAME: MODEL_NAME,
wml_client.repository.ModelMetaNames.TYPE:"scikit-learn_1.0",
wml_client.repository.ModelMetaNames.SOFTWARE_SPEC_UID:software_spec_uid
}

model_details = wml_client.repository.store_model(
model=DEMO_MODEL,
meta_props=model_props,
training_data=X_train,
training_target=Y_train
)

model_details

model_id=wml_client.repository.get_model_id(model_details)

model_id

deployment_props={
   wml_client.deployments.ConfigurationMetaNames.NAME:DEPLOYMENT_NAME,
   wml_client.deployments.ConfigurationMetaNames.ONLINE:{}
}

deployment = wml_client.deployments.create(
   artifact_uid=model_id,
   meta_props=deployment_props
)
```

8.TESTING
8.1 TEST CASES

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Test case ID	Feature Type	Component	Test Scenario	Pre-Requisite	Steps To Execute	Test Data	Expected Result	Actual Result	Status	Commnets	TC for Automatio	BUG ID	Executed By
2	PredictionID_01	Webpage	Prediction Page	Verify user is able to see the prediction page	Active server with frontend to predict the value	1.Enter URL in the browser 2.Search the URl 3.Verify that the prediction page is displayed or not	http://127.0.0.1:5000/	Prediction page should display	Working as expected	Pass	User can see the prediction page and values bar	NO	----	PNT2022TMID29641
3	PredictionID_02	Webpage	Prediction Page	Verify user can able to enter their values in the correponding value bar	HTML value tag with the valid data	1.Enter the values 2.In the webpage the entering values will be displayed	http://127.0.0.1:5000/	User should able to enter the parameter values	Working as expected	Pass	User can enter the values in the respected value tags in the webpage	NO	---	PNT2022TMID29641
	PredictionID_03	Prediction	Prediction Page	check the entered values are predicted or not and also whether the wqi value displayed	Algorithm with the html and javascript to process the prediction	1.Enter URL(http://127.0.0.1:5000/) and click go 2.Enter the respected values 3.Click on predict button	pH : 6.58 conductivity: 125.7	The perfect wqi value for the corresponding parameter values	Working as expected	Pass	User will know the water quality parameter values to enter in the webpage	Yes	---	PNT2022TMID29641

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 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	Severity 4	Subtotal
By Design	13	3	2	3	21
Duplicate	3	1	3	0	7
External	4	3	0	1	8
Fixed	12	2	4	20	38
Not Reproduced	3	0	1	0	4

Skipped	0	0	1	1	2
Won't Fix	0	5	2	1	8
Totals	35	14	13	26	88

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
Print Engine	7	0	0	7
Client Application	51	0	0	51
Security	2	0	0	2
Outsource Shipping	3	0	0	3
Exception Reporting	9	0	0	9
Final Report Output	4	0	0	4
Version Control	2	0	0	2

9. RESULTS

9.1. PERFORMANCE METRICS

Project team shall fill the following information in model performance testing template.

S.No.	Parameter	Values	Screenshot
1.	**Metrics**	**Regression Model:** MAE – 0.9872005 MSE – 5.55509587 RMSE – 2.35692089 R2 score - 0.96971918	
2.	**Tune the Model**	Hyperparameter Tuning – 0.781970282	

		Validation method	Validation Method
			#Logistic Regression
			pred3= log_reg.predict(x_test)
			s3=accuracy_score(y_test,pred3)*100
			results('LR',s3)
			print("Accuracy score of LR :",accuracy_score(y_test,pred3)*100)
			#Random Forest
			pred2=R_model.predict(x_test)
			s2=accuracy_score(y_test,pred2)*100
			print("Accuracy score of RF :",s2)
			results('RF',s2)
			#KNN
			pred5=modellin.predict(x_test)
			s5=accuracy_score(y_test,pred5)*100
			print("Accuracy score of KNN :",s5)
			results('SVM-LIN',s5)
			#SVM Linear
			pred5=modellin.predict(x_test)
			s5=accuracy_score(y_test,pred5)*100
			print("Accuracy score SVM :",s5)
			results('SVM-LIN',s5)
			#SVM Poly
			pred6 = modelpoly.predict(x_test)
			s6=accuracy_score(y_test,pred6)*100
			print("Accuracy score of SVM Poly :",s6)
			results('SVM poly', s6)
			Accuracy score of LR : 92.67299864314789
			Accuracy score of RF : 97.33152419719585
			Accuracy score of KNN : 92.67299864314789
			Accuracy score SVM : 92.67299864314789
			Accuracy score of SVM Poly : 95.74853007688829

10. ADVANTAGES &DISADVANTAGES

ADVANTAGES:

- ➤ User can analyse the quality of water without any login and registration.
- ➤ Time consuming is less than manual method.
- ➤ No cost is need to predict the water quality only internet is required.
- ➤ User can predict the quality of water at any time and anywhere.
- ➤ This model in software testing evaluates test cases and various error incidents in a short span of time.

DISADVANTAGES:

- ➤ Collecting the dataset is somewhat difficult.
- ➤ When it comes to processing data, the scale of data generated far exceeds the human capacity to understand and analyze it.

11. CONCLUSION

Water is the important thing to all the living beings. We must know the quality of water whether it is safe to drink or not. Using machine learning algorithm we can able to predict the quality of water.Traditionally, one must undergo an expensive and time?consuming lab analysis to test the purity of the water. This project investigated a machine learning approach to predict and analyse the water quality in urban areas.

12. FUTURE SCOPE

Machine learning has been widely used as a powerful tool to solve problems in the water environment because it can be applied to predict the water quality.The fundamental issue that machine learning models with unbalanced datasets encounter is overfitting. In the future, generalised results are expected to be obtained utilising mixed features and a balanced dataset. In addition, we want to forecast water quality and automatically extract features using deep learning and a sizable dataset.

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(r'new1.pkl','rb'))
@app.route('/')
def home() :
  return render_template("webpage.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"]
  tc = request.form["tc"]
  na = request.form["na"]
  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("webpage.html",showcase = 'Excellent, The Predicted Value is' + str(y_pred))
  elif(y_pred >= 89 and y_pred<=94):
    return render_template("webpage.html",showcase = 'Very Good, The Predicted Value is' + str(y_pred))
  elif(y_pred >= 80 and y_pred<=88):
    return render_template("webpage.html",showcase = 'Good, The Predicted Value is' + str(y_pred))
  elif(y_pred >= 65 and y_pred<=79):
    return render_template("webpage.html",showcase = 'Fair, The Predicted Value is' + str(y_pred))
  elif(y_pred >= 45 and y_pred<=64):
    return render_template("webpage.html",showcase = 'Marginal, The Predicted Value is' + str(y_pred))
  else:
    return render_template("webpage.html",showcase = 'Poor, The Predicted Value is' + str(y_pred))


if __name__ == '__main__':
    app.run(debug = False)
```

HTML CODE

```
<!DOCTYPE html>
<html lang="en">
```

```
<head>
   <meta charset="UTF-8">
   <meta http-equiv="X-UA-Compatible" content="IE=edge">
   <meta name="viewport" content="width=device-width, initial-scale=1.0">
   <title>Document</title>
   <link rel="stylesheet" href="../static/css/webstyle.css">

</head>

<body>
   <header>
      <nav>
         <div class="row">
         <div class="row1">
               <img src="../static/css/drop.png" alt="logo">
            </div>
            <div class="row2">
               <h1>Urban Water Quality Prediction</h1>
               <br>
               <br>

            </div>
         </div>
      </nav>
   </header>
   <main>
      <div class="column">
         <form action="/login" method="post">
            <label for=""></label>
            <input type="text" name="year" id="" placeholder="Enter Year">
            <label for=""></label>
            <input type="text" name="do" id="" placeholder="Enter D.O">
            <label for=""></label>
            <input type="text" name="ph" id="" placeholder="Enter PH">
            <label for=""></label>
```

```
            <input type="text" name="co" id="" placeholder="Enter Conductivity">
            <label for=""></label>
            <input type="text" name="bod" id="" placeholder="Enter B.O.D">
            <label for=""></label>
            <input type="text" name="na" id="" placeholder="Enter Nitratenen">
            <label for=""></label>
            <input type="text" name="tc" id="" placeholder="Enter Total Coliform">
            <label for=""></label>
            <div class="last">
               <input type="submit" value="Predict">
            </div>
            <div class="bor">
              <p style="color:black">{{showcase}} </P>
            </div>

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

GITHUB LINK :

https://github.com/IBM-EPBL/IBM-Project-13780-1659529933

PROJECT DEMO VIDEO LINK:

DRIVE LINK:
https://drive.google.com/drive/folders/1x21cVmaEbenudptqrKxY_9TgElNiY6-F

YOUTUBE LINK:
https://youtu.be/eIsy__7gwCY