EFFICIENT WATER QUALITY ANALYSIS AND PREDICTION USING MACHINE LEARNING

IBM-PROJECT-PNT2022TMID25561

NALAIYATHIRAN PROJECT BASED LEARNING ON PROFESSIONAL READINESS FOR INNOVATION , EMPLOYING AND ENTERPRENEURSHIP

PORJECT REPORT BY:

N GAYATHRI-210919104012

SNR SHAMYUKTHA-210919104040

M MOHANESWARI-210919104025

A S MELITTA-210919104701

BACHELOR OF ENGINEERING

COMPUTER SCIENCE ENGINEERING

LOYOLA INSTITUTE OF TECHNOLOGY

CHENNAI-600123

ANNA UNIVERSITY::CHENNAI 600 023

BONAFIDE CERTIFICATE

Certified that this project “Efficient water quality analysis and predication using machine learning “ is the bonafide work of “ Gayathri N, Mohaneswari M, Shamyuktha SNR, Melitta A S,who carried out the project work under my supervision.

SIGNATURE	SIGNATURE
Dr.G.Bhuvaneswari,	Mr.Noor Mohammed,
HEAD OF THE DEPARTMENT,	ASSISTANT PROFESSOR,
Computer Science Engineering,	Computer ScienceEngineering,
Loyola institute of technology	Loyola institute of technology,
Palanchur,	Palanchur,
Chennai-600123	Chennai-600 123

TABLE OF CONTENTS

ABSTRACT

This study 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 of Tireh River located in the southwest of Iran. To develop the ANN and SVM, different types of transfer and kernel functions were tested, respectively. Reviewing the results of ANN and SVM indicated that both models have suitable performance for predicting water quality components. During the process of development of ANN and SVM, it was found that tansig and RBF as transfer and kernel functions have the best performance among the tested functions. Comparison of outcomes of GMDH model with other applied models shows that although this model has acceptable performance for predicting the components of water quality, its accuracy is slightly less than ANN and SVM. The evaluation of the accuracy of the applied models according to the error indexes declared that SVM was the most accurate model. Examining the results of the models showed that all of them had some over-estimation properties. By evaluating the results of the models based on the DDR index, it was found that the lowest DDR value was related to the performance of the SVM model.

INTRODUCTION

PROJECT OVERVIEWWater is the most important source for sustaining all kinds of life. Naturalwater resources and aquifers are being polluteddue to indiscriminate urbanization and industrialization; as a result, it may be contaminated with physical, chemical, and biological impurities. As reported, 80% of the diseases are water borne diseases. Several criteria are used to measurethe quality of water, including the quantity of salt (or salinity), bacteria levels, the percentage of dissolved oxygen or the amount of particles suspended in the water (turbidity). Good water quality implies that harmful substances (pollutants) are absent from the water, and needed substances (oxygen, nutrients) are present. The traditional and common estimation of water quality has been Laboratory analysis which is time consuming and not very practical. This method can be processed efficiently by applying machine learning algorithms and big data tools. Machine learning (ML) is a field of inquiry devoted to understanding and building methods that 'learn', that is, methods that leverage data to improve performance on some set of tasks. Machine learning (ML) is a topic of study focused on analyzing and developing "learning" methods, or methods that use data to enhance performance on a certain set of tasks. With the use of machine learning (ML), which is a form of artificial intelligence (AI), software programs can predict outcomes more accurately without having to be explicitly instructed to do so. In order to forecast new output values, machine learning algorithms use historical data as input. A data analysis technique called machine learning automates the creation of analytical models.

PURPOSE

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. The purpose of this project is to Predict Water Quality by considering all water quality standard indicators.

LITERATURE SURVEY

Azamathulla, H. M. 2013 – A Review on Application of Soft Computing Methods in Water Resources Engineering A2 –Yang, Xin-She. In: Metaheuristics in Water, Geotechnical and Transport Engineering (A. H. Gandomi, S. Talatahari & A. H. Alavi, eds). Elsevier, Oxford, pp. 27–41

Chen, X., Chen, Y., Shimizu, T., Niu, J., Nakagami, K. i., Qian, X., Jia, B., Nakajima, J., Han, J. & Li, J. 2017 Water resources management in the urban agglomeration of the Lake Biwa region, Japan: an ecosystem services-based sustainability assessment. Sci. Total Environ. 586 (Suppl. C), 174–187..
Dehghani, M., Saghafifian, B., Nasiri Saleh, F., Farokhnia, A. & Noori, R. 2014 Uncertainty analysis of stream flow drought forecast using artificial neural networks and Monte-Carlo simulation. Int. J. Climatol. 34 (4), 1169–1180.
Heddam, S. 2016 c New modelling strategy based on radial basis function neural network
(RBFNN) for predicting dissolved oxygen concentration using the components of the Gregorian calendar as inputs: case study of Clackamas River, Oregon, USA. Model. Earth Syst. Environ. 2 (4), 162–167.
Gocic, M., Shamshirband, S., Razak, Z., Petkovic´, D., Ch, S. & Trajkovic, S. 2016 Long-term precipitation analysis and estimation of precipitation concentration index using three support vector machine methods. Adv. Meteorol. doi:10. 1155 /2016/ 7912357.
Jaddi, N. S. & Abdullah, S. 2017 A cooperative-competitive masterslave global-best harmony search for ANN optimization and water-quality prediction. Appl. Soft Computer. 51, 209–224.
E. Tipton, L. Hedges, M. Vaden-Kiernan, G. Borman, K. Sullivan, and S. Caverly, "Sample selection in randomized experiments: a new method using propensity score stratified sampling," Journal of Research on Educational Effectiveness, vol. 7, no. 1, pp. 114-135, Jan. 2014.
Parsaie, A. & Haghiabi, A. H. 2017c Improving modelling of discharge coefficient of triangular labyrinth lateral weirs Using SVM, GMDH and MARS techniques. Irrigation and Drainage 66, 636–654.Zahiri, A., Azamathulla, H. M. & Ghorbani, K. 2014 Prediction of local scour depth
downstream of bed sills using soft computing models. In: Computational Intelligence Techniques in Earth and Environmental Sciences (T. Islam, P. K. Srivastava, M. Gupta, X. Zhu & S. Mukherjee, eds). Springer Netherlands, Dordrecht, Switzerland, pp. 197–208
Motagh, M., Shamshiri, R., Haghshenas Haghighi, M., Wetzel, H.-U., Akbari, B., Nahavandchi, H., Roessner, S. & Arabi, S. 2017 Quantifying groundwater exploitation induced subsidence in the Rafsanjan plain, southeastern Iran, using InSAR time-series and in situ measurements. Eng. Geol. 218, 134–151.
L. Li, P. Jiang, H. Guang, L. Dong, G. Wu, and H. Wu " Water quality prediction based

on recurrent neural network and improved evidence theory: a case study of Qiantang River, China," Environmental Science and Pollution Research, vol. 26, no. 19, pp. 19879-19896, Mar. 2019.

K. Kadam, V. M. Wagh, A. A. Muley, B. N. Umrikar, and R. N. Sankhua, "Prediction of Water quality index using artificial neural network and multiple linear regression modeling approach in Shivganga River basin, India," Modeling Earth Systems and Environment, vol. 5, no.3, pp. 951-96, Mar. 2019.

Y. Khan and S. Chai, "Ensemble of ANN and ANFIS for water quality prediction and analysis - a data driven approach," Journal of Telecommunication, Electronic and Computer Engineering , vol. 9, no. 2, pp. 117-122, 2017.

R. Barzegar, J. Adamowski, and A. A. Moghaddam, "Application of wavelet-artificial intelligence hybrid models for water quality prediction: a case study in Aji-Chay River, Iran," Stochastic Environmental Research and Risk Assessment, vol. 30, no. 7, pp. 1797 1819, Jan. 2016.

Azad, H. Karami, S. Farzin, A. Saeedian, H. Kashi, and F. Sayyahi, "Prediction of Water Quality Parameters Using ANFIS Optimized by Intelligence Algorithms (Case Study: Gorganrood River)," KSCE Journal of Civil Engineering, vol. 22, no. 7, pp. 2206-2213, Sep. 2017.

K.Venkatesan, S. Ahmad, W. Johnson, J. R. Batista, "Systems dynamic model to forecast salinity load to the Colorado River due to urbanization within the Las Vegas Valley," Science of the Total Environment of Research on Educational Effectiveness, vol. 469, no. 13, pp. 2616 2625, Apr. 2011.

J. Wan, M. Huang, Y. Ma, W. Guo, Y. Wang, H. Zhang, W. Li, and X. Sun, "Prediction of effluent quality of a paper mill wastewater treatment using an adaptive network-based fuzzy inference system," Applied Soft Computing, vol. 11, no. 3, pp. 3238-3246, Apr. 2011.

Existing problem

The main problem lies here. For testing the water quality we have to conduct lab tests on the water which is costly and time-consuming as well. So, in this paper, we propose an alternative approach using artificial intelligence to predict water quality. This method uses a significant and easily available water quality index which is set by the WHO(World Health Organisation). The data taken in this paper is taken from the PCPB India which includes 3277 examples of the distinct wellspring. In this paper,WQI(WaterQualityIndex)iscalculatedusingAItechniques.Soinfuturework,wecaninte gratethiswith IoT based framework to study large datasets and to expand our study to a larger scale. By usingthat it can predict the water quality fast and more accurately than any other IoT framework. That IoTframework system uses some limits for the sensor to check the parameters like ph, Temperature, Turbidity, and so on. And further after reading this parameter pass these readings to the Arduino microcontroller and ZigBee handset for further prediction

2.2 References

Srivastava,G.;Kumar,P.Waterqualityindexwithmissingparameters.Int.J.Res.Eng.Technol.2 013,2,609–614.

PCRWR.WaterQualityofFiltrationPlants,MonitoringReport;PCRWR:Islamabad,Pakistan,20 10.Availableonline:htt p://www.pcrwr.gov.pk/Publications/Water%20Quality%20Reports/FILTRTAION%20PLAN TS%20REPOTCDA.pdf (accessedon23August2019).

Sakizadeh,M.Artificialintelligenceforthepredictionofwaterqualityindexingroundwatersyst ems.Model.EarthSyst. Environ.2016,2,8. [CrossRef]

2.3Problem Statement

Definition Access to safe drinking-water is essential to health, a basic human right and a component of effective policy for health protection. This is important as a health and development issue at a national, regional and local level. In some regions, it has been shown that investments in water supply and sanitation can yield a net economic benefit, since the reductions in adverse health effects and health care costs out weigh the costs of under taking the interventions.

EXISITING PROBLEM

The main problem lies here. For testing the waterquality we have to conduct lab testson thewater which is costly and time-consuming as well. So, in this paper, we propose an alternative approachusing artificialintelligence to predict water quality.This method uses a significant and easily availablewater quality index which is set by the WHO(World Health Organisation). The data takenin this paper is taken from the PCPB Indiawhich includes 3277 examples of the distinctwellspring. In this paper,WQI(WaterQualityIndex)iscalculatedusingAItechniques.Soinfuturework,wecaninte gratethiswith IoT based frameworkto study large datasets and to expand our study to a larger scale. By usingthatitcan predictthe water qualityfast and more accurately than any otherIoT framework. ThatIoTframework system uses some limitsfor the sensor to checkthe parameters like ph, Temperature, Turbidity, and so on. And further afterreading this parameter pass these readingsto the Arduinomicrocontroller and ZigBee handsetfor further prediction

REFERANCE

Privastava,G.;Kumar,P.Waterqualityindexwithmissingparameters.Int.J.Res.Eng.Technol.2013,2,609–614.

PCRWR.WaterQualityofFiltrationPlants,MonitoringReport;PCRWR:Islamabad,Pakistan,2010.Availableonline:http://www.pcrwr.gov.pk/Publications/Water%20Quality%20Reports/FILTRTAION%20PLANTS%20REPOT-CDA.pdf(accessedon23August2019).

Sakizadeh,M.Artificialintelligenceforthepredictionofwaterqualityindexingroundwatersystems.Model.EarthSyst. Environ.2016,2,8. [CrossRef]

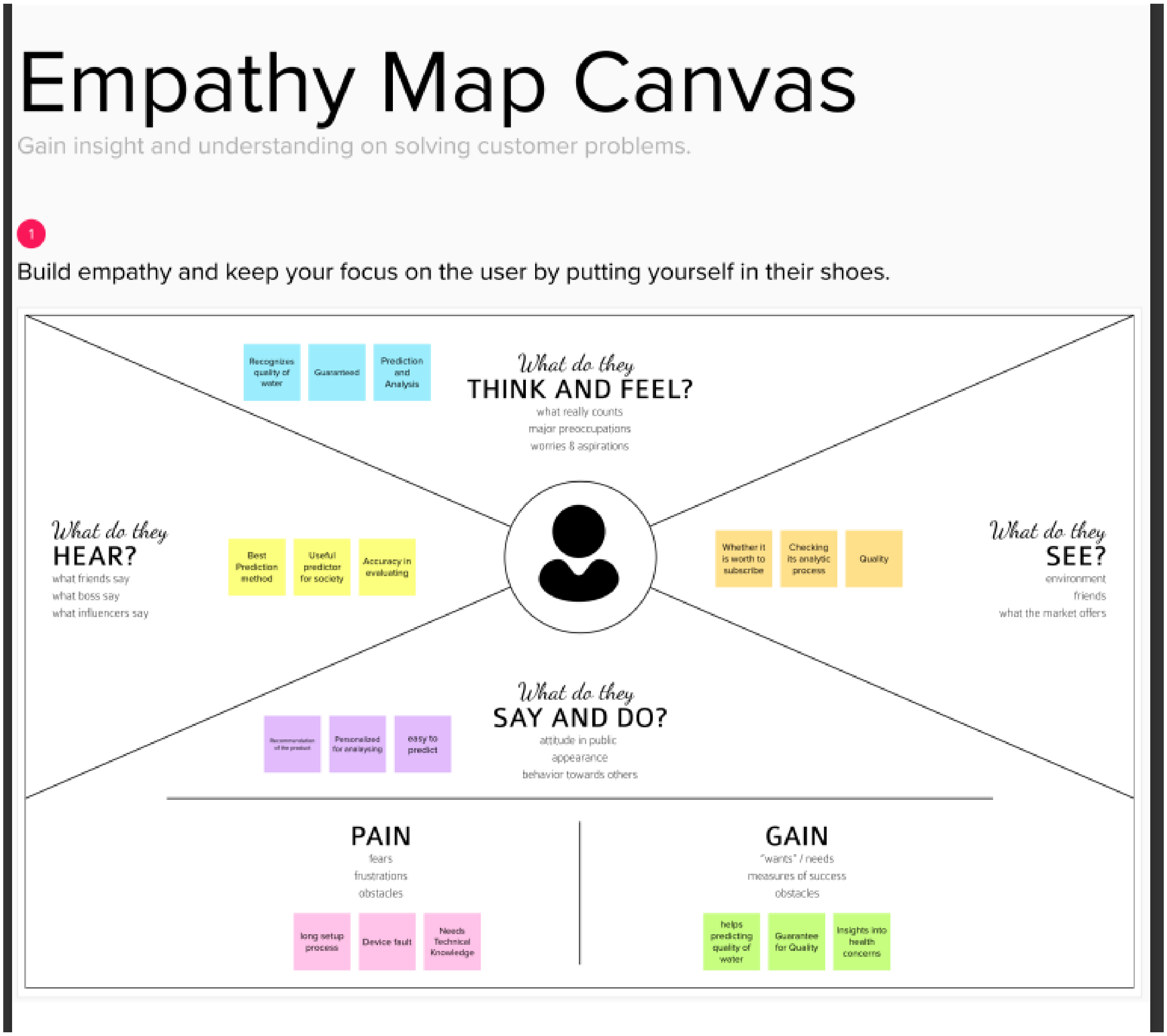
Problem Statement Definition

Access to safe drinking-water is essential to health, a basic human right and a component of effective policy for health protection. This is important as a health and development issue at a national, regional and local level. In some regions, it has been shown that investments in water supply and sanitation can yield a net economic benefit, since the reductions in adverse health effects and health care costs out weighthe costs of under taking theinterventions

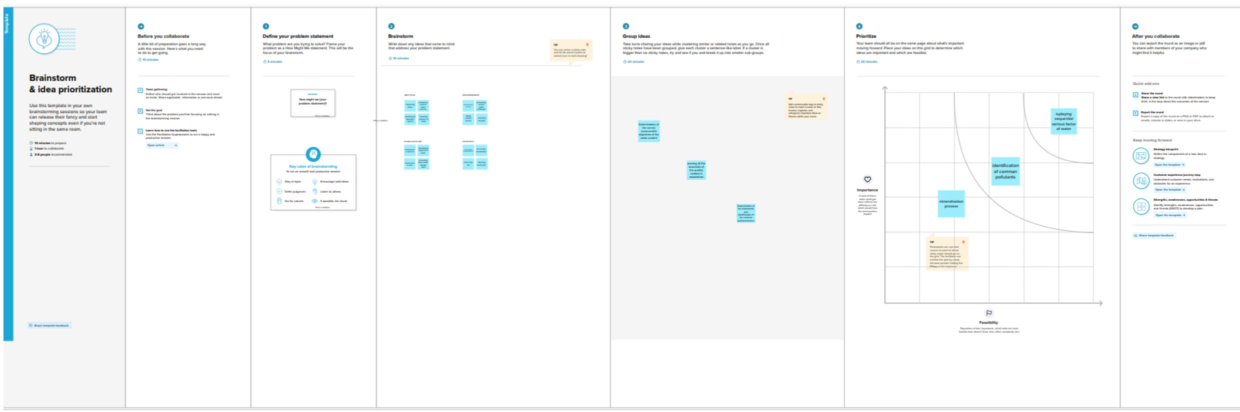
IDEATION & PROPOSEDSOLUTION

Empathy Map Canvas

An empathy map canvas serves as a foundation for outstanding user experiences, which focus on providing the experience customers want rather than forcing design teams to rely on guesswork.Empathy map canvases help identify exactly what it is that users are looking for brandscan deliver.



IDEATION AND BRAIN STORMING



Proposed Solution

Water quality has been conventionally estimated through expensive and time-consuming lab andstatistical analyses, which render the contemporary notion of real-time monitoring moot. The alarming consequences of poor water quality necessitate an alternative method,which is quickerand inexpensive. With this motivation, this research exploresaseries of supervised machinelearning 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 proposedmethodology employs four input parameters, namely, temperature, turbiditypHand total dissolved solids.

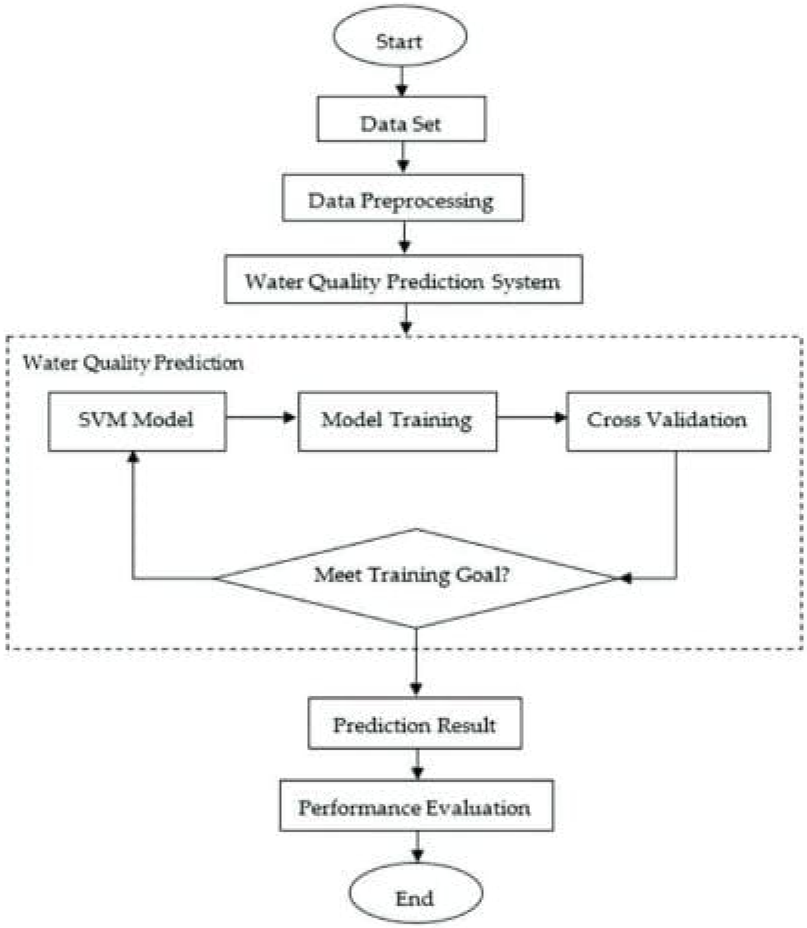
REQUIREMENT ANALYSIS

FR NO.	FUNCTIONAL REQUIREMENT(EPIC)	SUB REQUIREMENTS (STORY\SUB-TASK)
FR-1	User registration	Registration through form Registration through Gmail Registration through Linkedin
FR_2	User Confirmation	Confirmation via Email Confirmation via OTP
FR_3	Executive administraction	Registration of monitoring the environment status and

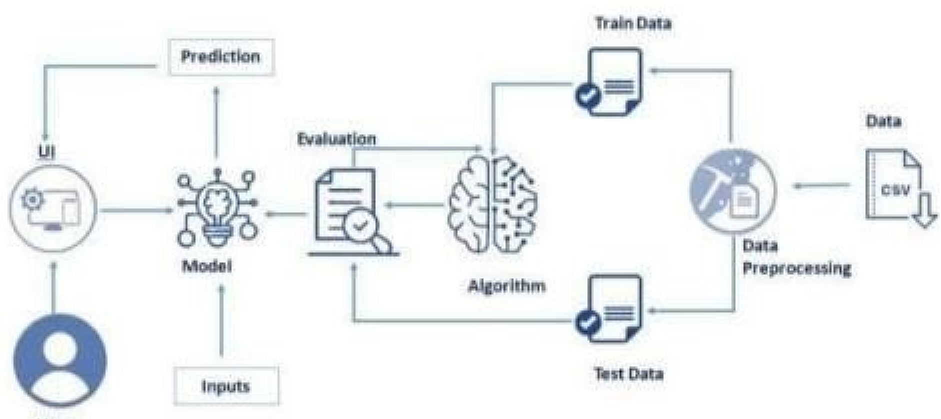
		regulatry complience like pollution event enmergency managent ,and it includes two different functions :early warning /forcast monitoring.
FR-4	Data handling	file contains water quality metrics for different water bodies.
FR_5	Quality analysis	analysis with the acquired information of the water accross various water quality indicator like (PH,turbidoty,TDS,temperatu re ,)using different modes.

5. PROJECT DESIGN

5.1 Data Flow Diagram



Solution &Technical Architecture



PROJECT PLANNING & SCHEDULING

Sprint Planning& Estimation

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint-1	Data Collection	USN-1	Collecting dataset for pre-processing	10	High	Billu Dilip Budharaju Sumanth Aduri Charansai Bandla Venkata Akash Bhuvanendra Chowdary V
Sprint-1		USN-2	Data pre-processing- Used to transform the data into useful format.	10	Medium	Billu Dilip Budharaju Sumanth Aduri Charansai Bandla Venkata Akash Bhuvanendra Chowdary V
Sprint-2	Model Building	USN-3	Calculate the Water Quality Index (WQI) using Regression algorithm of machine learning.	10	High	Billu Dilip Budharaju Sumanth Aduri Charansai Bandla Venkata Akash Bhuvanendra Chowdary V
Sprint-2		USN-4	Splitting the data into training and testing from the entire dataset.	10	Medium	Billu Dilip Budharaju Sumanth Aduri Charansai Bandla Venkata Akash Bhuvanendra Chowdary V
Sprint-3	Training and Testing	USN-5	Training the model using regression algorithm and testing the performance of the model	20	Medium	Billu Dilip Budharaju Sumanth Aduri Charansai Bandla Venkata Akash Bhuvanendra Chowdary V
Sprint-4	Implementation of Web page	USN-6	Implementing the web page for collecting the data from user	10	High	Billu Dilip Budharaju Sumanth Aduri Charansai Bandla Venkata Akash Bhuvanendra Chowdary V
Sprint-4		USN-6	Deploying the model using IBM Cloud and IBM Watson Studio	10	Medium	Billu Dilip Budharaju Sumanth Aduri Charansai Bandla Venkata Akash Bhuvanendra Chowdary V

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		
Sprint-3	20	6 Days	07 Nov 2022	12 Nov 2022		
Sprint-4	20	6 Days	14 Nov 2022	19 Nov 2022		

velocity:

sprint 1average velocity:

average velocity=20/2=10

sprint 2 average velocity :

average velocity=20/2=10

sprint 3 average velocity:

average velocity=20/1=20

sprint 4 average velocity:

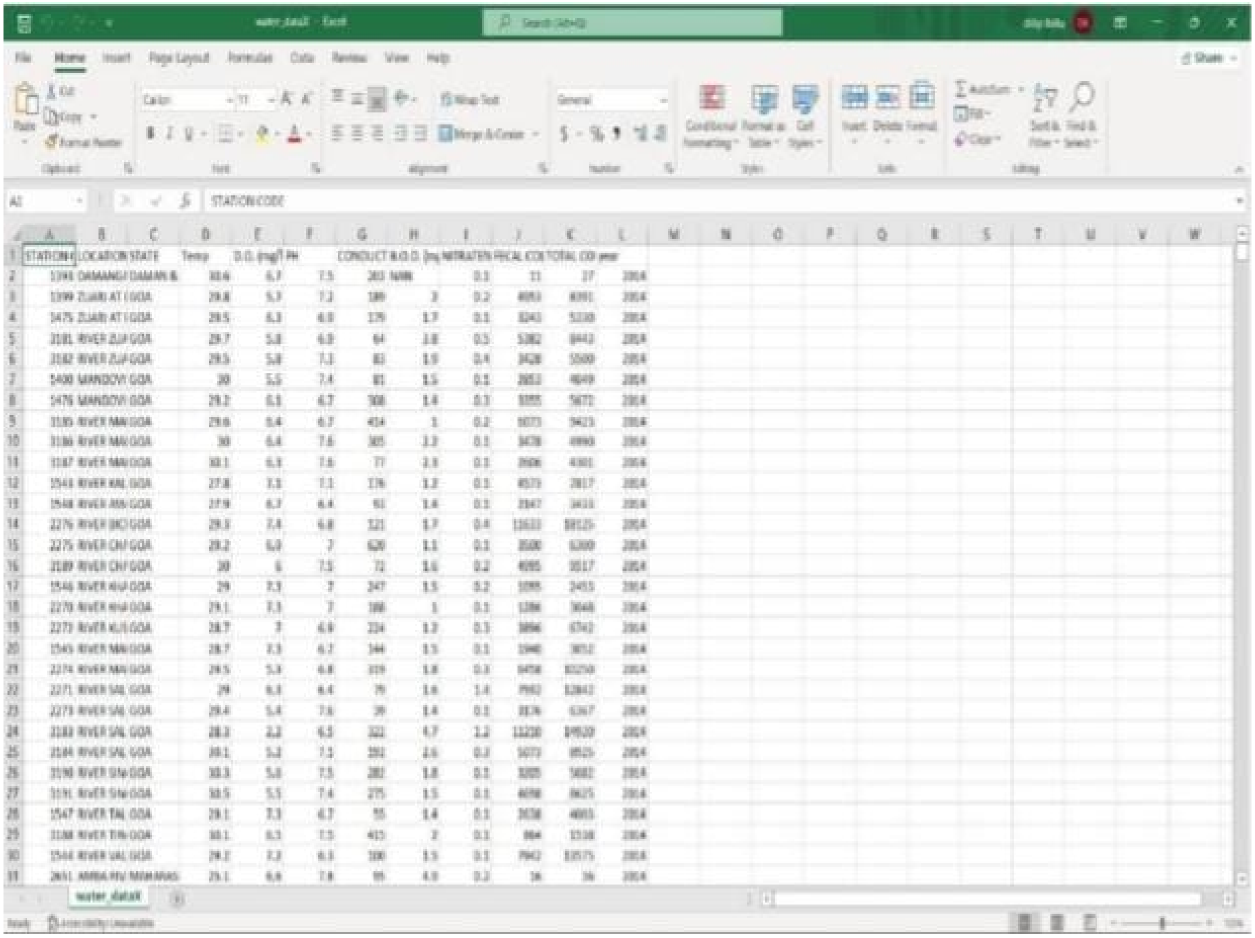
average velocity=20/2=10

7. CODING & SOLUTIONING

7.1 FEATURE 1

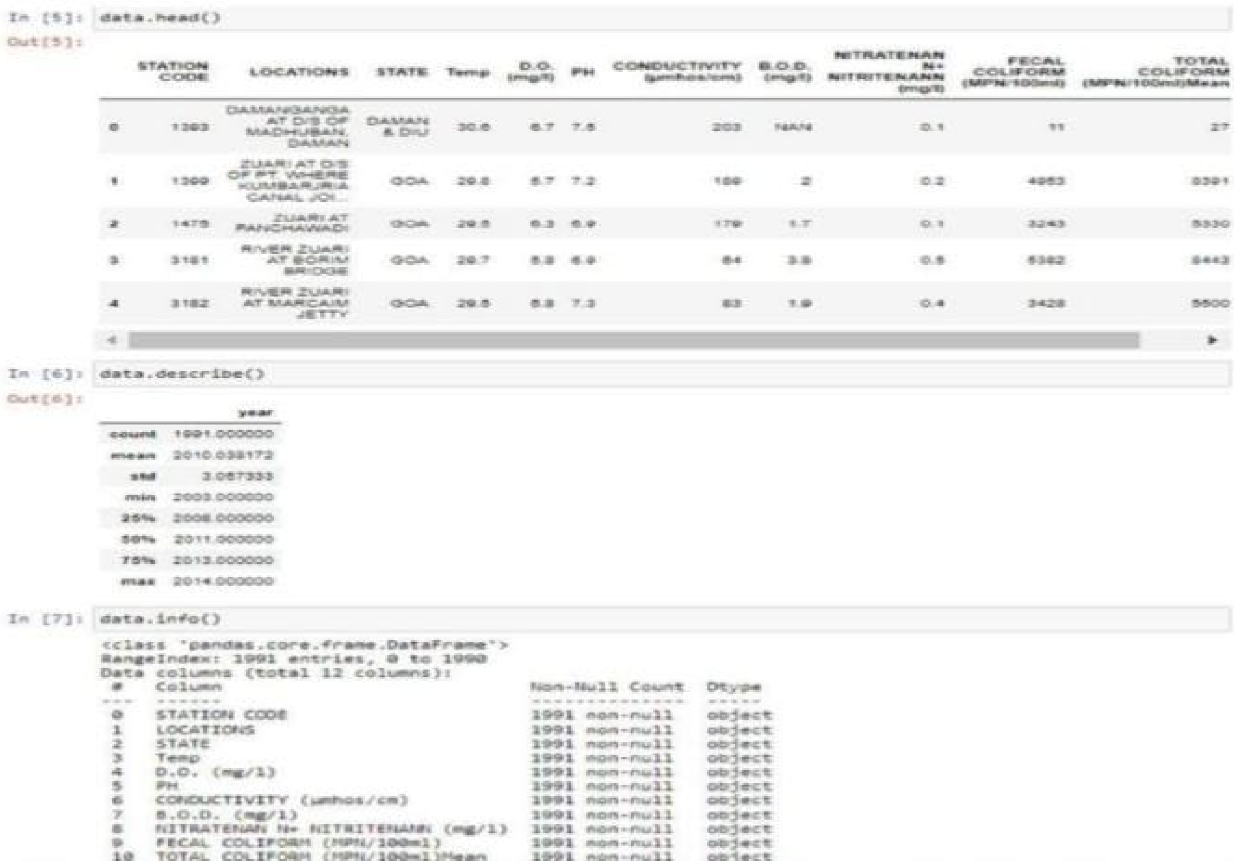
Data collection and creation:

Data mining techniques require domain knowledge in order to generate predictions. For water quality applications, it is vital to understand how various water quality parameters influnces wate.Thisinformationcancomefromadomainexpertorhistoricaldatacollections.Forthe forecasting task, two types of data sets were used: a carefully created huge synthetic data set and an available real dataset



7.2 FEATURE 2

Performance Measures Results True Positives (TP) are when the model predicts the positive class properly. True Negatives (TN) is one of the components of a confusion matrix designed to demonstrate how classification algorithms work. Positive outcomes that the model predicted incorrectly are known as False Positives (FP). False Negatives (FN) are negative outcomes that the model predicts negative class. Accuracy is the most basic and intuitive performance metric, consisting of the ratio of successfully predicted observations to total observations. Accuracy=TP+TN/(TP+FP+FN+TN)



```
In [5]: data.head()
Out[5]:
```

	STATION CODE	LOCATIONS	STATE	Temp	D.O. (mg/l)	PH	CONDUCTIVITY (µmhos/cm)	B.O.D. (mg/l)	NITRATENAN N+ NITRITENANN (mg/l)	FECAL COLIFORM (MPN/100ml)	TOTAL COLIFORM (MPN/100ml)Mean
0	1393	DAMANGANGA AT D/S OF MADHUBAN, DAMAN	DAMAN & DIU	30.6	6.7	7.5	203	NAN	0.1	11	27
1	1399	ZUARI AT D/S OF PT. WHERE KUMBARJRIA CANAL JOI...	GOA	29.8	5.7	7.2	189	2	0.2	4953	8391
2	1475	ZUARI AT PANCHAWADI	GOA	29.5	6.3	6.9	179	1.7	0.1	3243	5330
3	3181	RIVER ZUARI AT BORIM BRIDGE	GOA	29.7	5.8	6.9	64	3.8	0.5	5382	8443
4	3182	RIVER ZUARI AT MARCAIM JETTY	GOA	29.5	5.8	7.3	83	1.9	0.4	3428	5500

```
In [6]: data.describe()
Out[6]:
```

	year
count	1991.000000
mean	2010.038172
std	3.057333
min	2003.000000
25%	2008.000000
50%	2011.000000
75%	2013.000000
max	2014.000000

```
In [7]: data.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1991 entries, 0 to 1990
Data columns (total 12 columns):
 #   Column                               Non-Null Count  Dtype
---  ------                               --------------  -----
 0   STATION CODE                         1991 non-null   object
 1   LOCATIONS                            1991 non-null   object
 2   STATE                                1991 non-null   object
 3   Temp                                 1991 non-null   object
 4   D.O. (mg/l)                          1991 non-null   object
 5   PH                                   1991 non-null   object
 6   CONDUCTIVITY (µmhos/cm)              1991 non-null   object
 7   B.O.D. (mg/l)                        1991 non-null   object
 8   NITRATENAN N+ NITRITENANN (mg/l)     1991 non-null   object
 9   FECAL COLIFORM (MPN/100ml)           1991 non-null   object
 10  TOTAL COLIFORM (MPN/100ml)Mean       1991 non-null   object
```

8. TESTING

8.1.TEST CASE 1

8.2 TEST CASE 2

8.2 User Acceptance Testing

1. Purpose of Document:

The purpose of this report 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 d

Resolution	Severity 1	Severity 2	Severity 3	Severity 4	Subtotal
By Design	10	4	2	3	20
Duplicate	1	0	3	0	4
External	2	3	0	1	6
Fixed	11	2	4	20	37
Not Reproduced	0	0	1	0	1
Skipped	0	0	1	1	2
Won't Fix	0	5	2	1	8
Totals	24	14	13	26	77

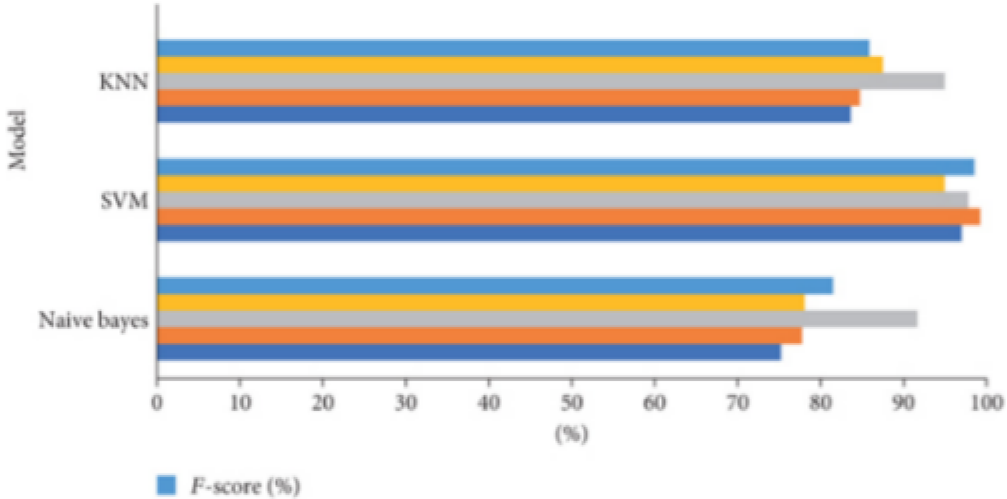
3. Test Case Analysis: This report shows the number of test cases that have passed ,failed, and untested.

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.RESULT

9.1 PERFORMANCE METRICS

For validating the developed model, the dataset has been divided into 70% training and 30%testing subsets. While the ANN and LSTM models were used to predict the WQI, the SVM, KNN, and Naive Bayes were utilized for the water quality classification prediction



SO,WE ARE GOING
Performance Measures Results True Positives (TP) are when the model predicts the positive clasPerformance Measures Results True Positives (TP) are when the model predicts the positive class properly. True Negatives (TN) is one of the components of a confusion matrix designed to demonstrate how classification algorithms work. Positive outcomes that the model predicted incorrectly are known as False Positives (FP). False Negatives (FN) are negative outcomes that the model predicts negative class. Accuracy is the most basic and intuitive performance metric, consisting of the ratio of successfully predicted observations to tota lobservations .Accuracy=TP+TN/(TP+FP+FN+T)

ADVANTAGE

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

DISADVANTAGES

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

CONCLUSION

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

12. SOURCE CODE

Machinelearninghasbeenwidelyusedasapowerfultooltosolveproblemsin the water environment because it can be applied to predict water quality, optimize water resource allocation, manage water resource shortages, etc. Despite this, several challenges remain in fully applying machinelearning approaches in this field to evaluate water quality:

(1) Machine learning is usually dependent on large amounts of high-quality data. Obtaining

sufficient data with high accuracy in water treatment and management systems is often difficult owing to the cost or technology limitations.

(2) As the conditions in real water treatment and management systems can be extremely complex, the current algorithms may only be applied to specific systems, which hinders the wide application of machinelearning approaches.

(3) The implementation of machine learning algorithms in practical applications requires researchers to have certain professional background knowledge.

. 13. APPENDIX

REQUIREMENT.TX

```
FLASK==2.2.2
JOBLIB==1.2.0
NUMPY==1.23.4
PANDAS==1.5.1
SCIKIT-LEARN==1.1.3
.XGBOOST==1.7.1
GUNICORN==20.1.0
SEABORN==0.12.1
GEVENT
REQUESTS
FLASK-CORS==3.0.10
```

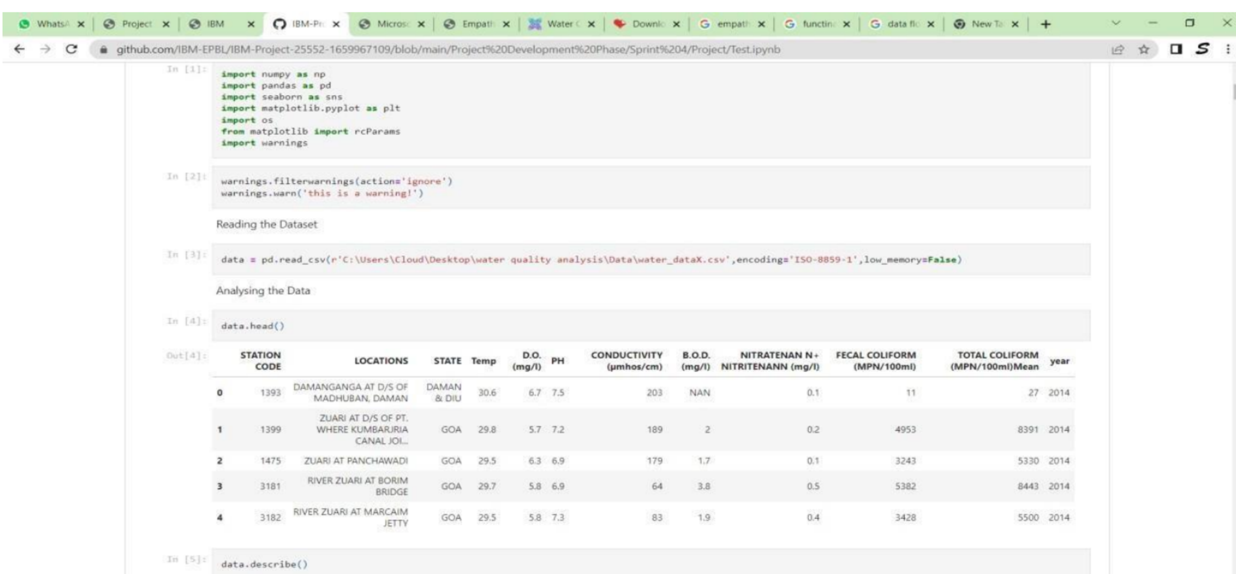
APP.PY

```
import numpy as np
tron Flask sport klass, render template, request
 Import pickle
model pickle.load(open('ul.pk1","rh"}}
 def home():
year = request.form["year"]
ph request.form["ph"

co request.form["co"]

na request.form["na"]
request.forni
total- [[int(year), Fist (do), #Inat (ph), inat(en), Inet (nd), Inas (na), Int(tr)}]
y pred nadel.predict(total)
y.predy pred[0]
Fly pred 35 y pred 100)
return render_template("Index_fitel, showcase "Excellent, The Predicted Value
is"estr(y_pred))
fly pred 80 and y pred ced)
```

TEST.IPYNB


github.com/IBM-EPBL/IBM-Project-25552-1659967109/blob/main/Project%20Development%20Phase/Sprint%204/Project/Test.ipynb

In [1]:
```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import os
from matplotlib import rcParams
import warnings
```

In [2]:
```
warnings.filterwarnings(action='ignore')
warnings.warn('this is a warning!')
```

Reading the Dataset

In [3]:
```
data = pd.read_csv(r'C:\Users\Cloud\Desktop\water quality analysis\Data\water_dataX.csv',encoding='ISO-8859-1',low_memory=False)
```

Analysing the Data

In [4]:
```
data.head()
```

Out[4]:

	STATION CODE	LOCATIONS	STATE	Temp	D.O. (mg/l)	PH	CONDUCTIVITY (µmhos/cm)	B.O.D. (mg/l)	NITRATENAN N+ NITRITENANN (mg/l)	FECAL COLIFORM (MPN/100ml)	TOTAL COLIFORM (MPN/100ml)Mean	year
0	1393	DAMANGANGA AT D/S OF MADHUBAN, DAMAN	DAMAN & DIU	30.6	6.7	7.5	203	NAN	0.1	11	27	2014
1	1399	ZUARI AT D/S OF PT. WHERE KUMBARJRIA CANAL JOI...	GOA	29.8	5.7	7.2	189	2	0.2	4953	8391	2014
2	1475	ZUARI AT PANCHAWADI	GOA	29.5	6.3	6.9	179	1.7	0.1	3243	5330	2014
3	3181	RIVER ZUARI AT BORIM BRIDGE	GOA	29.7	5.8	6.9	64	3.8	0.5	5382	8443	2014
4	3182	RIVER ZUARI AT MARCAIM JETTY	GOA	29.5	5.8	7.3	83	1.9	0.4	3428	5500	2014

In [5]:
```
data.describe()
```

INDEX.HTML

```
<!DOCTYPE
html>
            <html lang="en">
            <head>
                <meta charset="UTF-8">
                <meta http-equiv="X-UA-Compatible" content="IE=edge">
                <meta name="viewport" content="width=device-width, initial-scale=1.0">
                <title>Water Quality Analysis</title>
                <link rel="stylesheet" href="../static/index.css">
            </head>
            <body style="background-image: linear-gradient(cyan,white);">
            <center><div class="header1"><span><img src="/static/logo.png" width="20px"></span
            Analysis</font></div></center>
                <br><br><br>
                <form class="fill" action="/login" method="post">
                    <br>
                    <center>
                    <label>Enter Year</label>
                <input type="text" name="year" placeholder="Enter Year" id="" required><br><br
                    <label>Enter D.O</label>
                <input type="text" name="do" placeholder="Enter D.O" id="" required><br><br>
                    <label>Enter PH</label>
            <input type="text" name="ph" placeholder="Enter PH" id="" required><br><br>
            <label> Enter B.O.D</label>

            <inputtype="text" name="bod" placeholder="Enter B.O.D" id="" required><br><br>

            <label>Enter Nitratenen</label>

            <input type="text" name="na" placeholder="Enter Nitratenen" id="" required><br><br

            <label>Enter Total Coliform</label>

            <input type="text" name="tc" placeholder="Enter Total Coliform" id="" required><br

        <input type="submit" class="logbtn" value="Predict" style="width: 10%;">
         </center>
        <div class="bor"><center><b><font color="black"
           size=5>{{showcase}}</font></b></center></div>
      </form>
     </body>
</html>
```

LINKS:

GITHUB- https://github.com/IBM-EPBL/IBM-Project-53536-1661414817/blob/main/DESIGN%20AND%20PLANNING/IDEATION/Empathy%20map.pdf

DEMOLINK

https://drive.google.com/file/d/1DBRB6RwlyfBx989eLSieViEQXm2BSQci/view?usp=drivesd