PROJECT-BASED LEARNING (NAAN MUDHALVAN)

ON

PROFESSIONAL READINESS FOR INNOVATION, EMPLOYABILITY AND ENTREPRENEURSHIP

PROJECT REPORT

Team ID	**NM2023TMID01930**
Project Title	**Assessing the safety of Municipal Drinking Water**

S.No	Team Mates	Name	Register Number
1.	Team Leader	Muthuraj.M	420420104032
2.	Team Member-01	Nedunchezhian.M	420420104035
3.	Team Member-02	Gokula Krishnan.S	420420104012
4.	Team Member-03	Ashwin Kumar.K	420420104003

in partial fulfilment for the award of the degree of

in

BACHELOR OF ENGINEERING

COMPUTER SCIENCE AND ENGINEERING

ADHIPARASAKTHI ENGINEERING COLLEGE
MELMARUVATHUR-603 319

ANNA UNIVERSITY: CHENNAI 600 025

MAY 2023

ANNA UNIVERSITY: CHENNAI 600 025

BONAFIDE CERTIFICATE

Certified that this project report titled **"ASSESSING SAFETY OF MUNICIPAL DRINKING WATER"** is the bonafide work of **MUTHURAJ.M (420420104032), NEDUNCHEZHIAN.M (420420104035), GOKULAKRISHNAN.S(420420104012),ASHWINKUMAR.S (420420104003)** who carried out the work under my supervision.

SIGNATURE
Dr. C. DHAYA, Ph.D.,
HEAD OF THE DEPARTMENT
Department of CSE
Adhiparasakthi Engineering College
Melmaruvathur - 603 319.

SIGNATURE
Mr. K. CHAIRMADURAI, M.E.,
FACULTY MENTOR
Department of CSE
Adhiparasakthi Engineering College
Melmaruvathur - 603 319.

Submitted for the Naan Mudhalvan IBM Project at Adhiparasakthi Engineering College, Melmaruvathur.

TABLE OF CONTENT

1. INTRODUCTION

1.1 Project Overview

This project, titled “Assessing safety of municipal drinking water”, aims at providing a platform for predicting the safety of municipal drinking water by giving the required parameters of the water like PH value, hardness, etc. This can be achieved by training a machine learning model in the colab using various algorithms and after the model developed, it integrated with front end using flask in the Visual Studio IDE.

1.2 Purpose

The people who consume the municipal drinking water prone to illness due to pollution in the drinking water for that we are going to assess the safety of municipal drinking water using ML models is to enhance public health protection by providing accurate, timely, and proactive assessments of water quality. By leveraging historical data, real-time monitoring, and predictive analytics, ML models contribute to improved decision-making, resource optimization, and effective management of water systems, ultimately ensuring the provision of safe drinking water to communities.

2. IDEATION & PROPOSED SOLUTION

2.1 Problem Statement Definition

Figure 2.1: POV of Users

2.2 Empathy Map Canvas

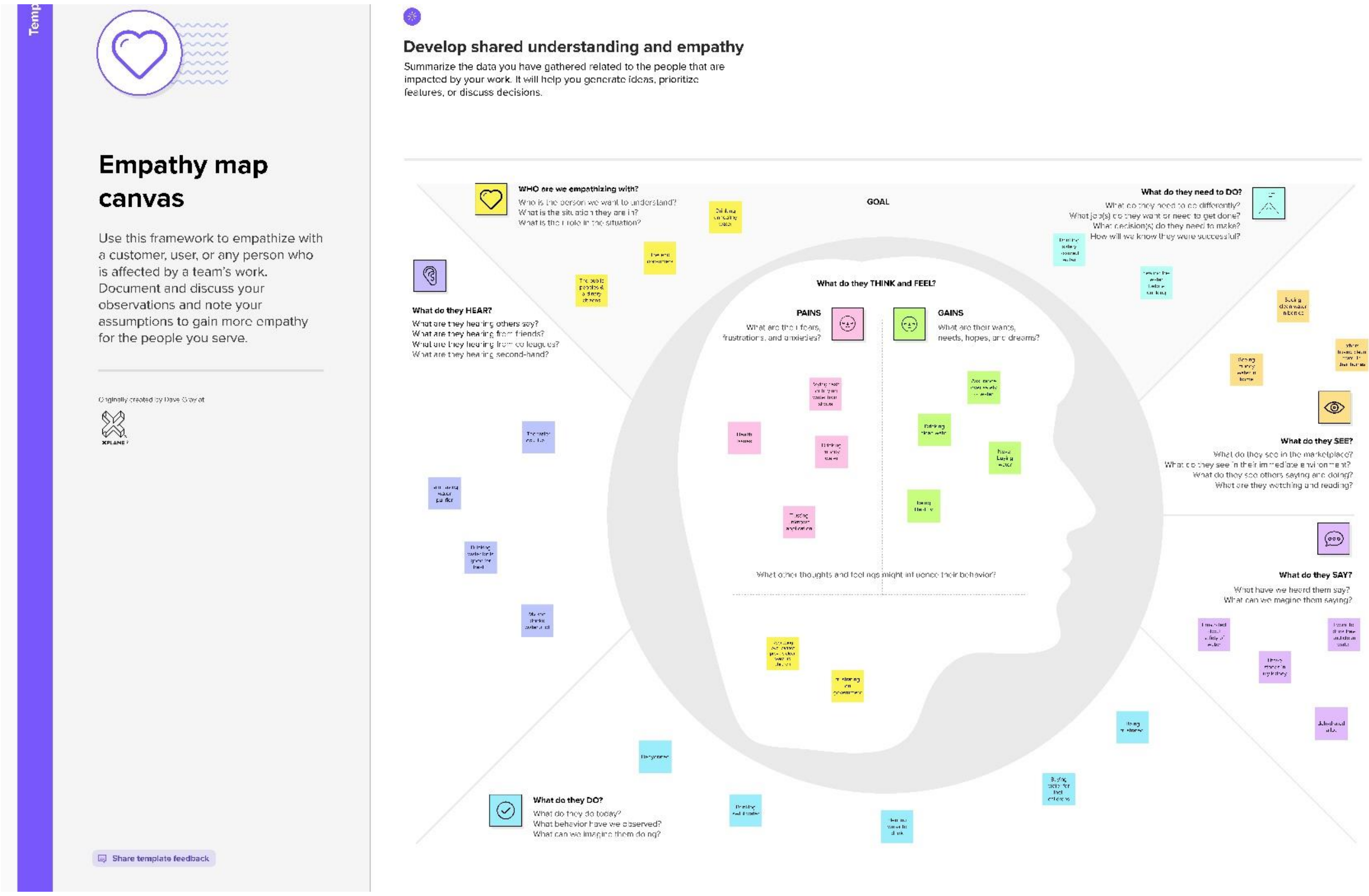


Figure 2.2: Empathy Map

2.3 Ideation & Brainstorming

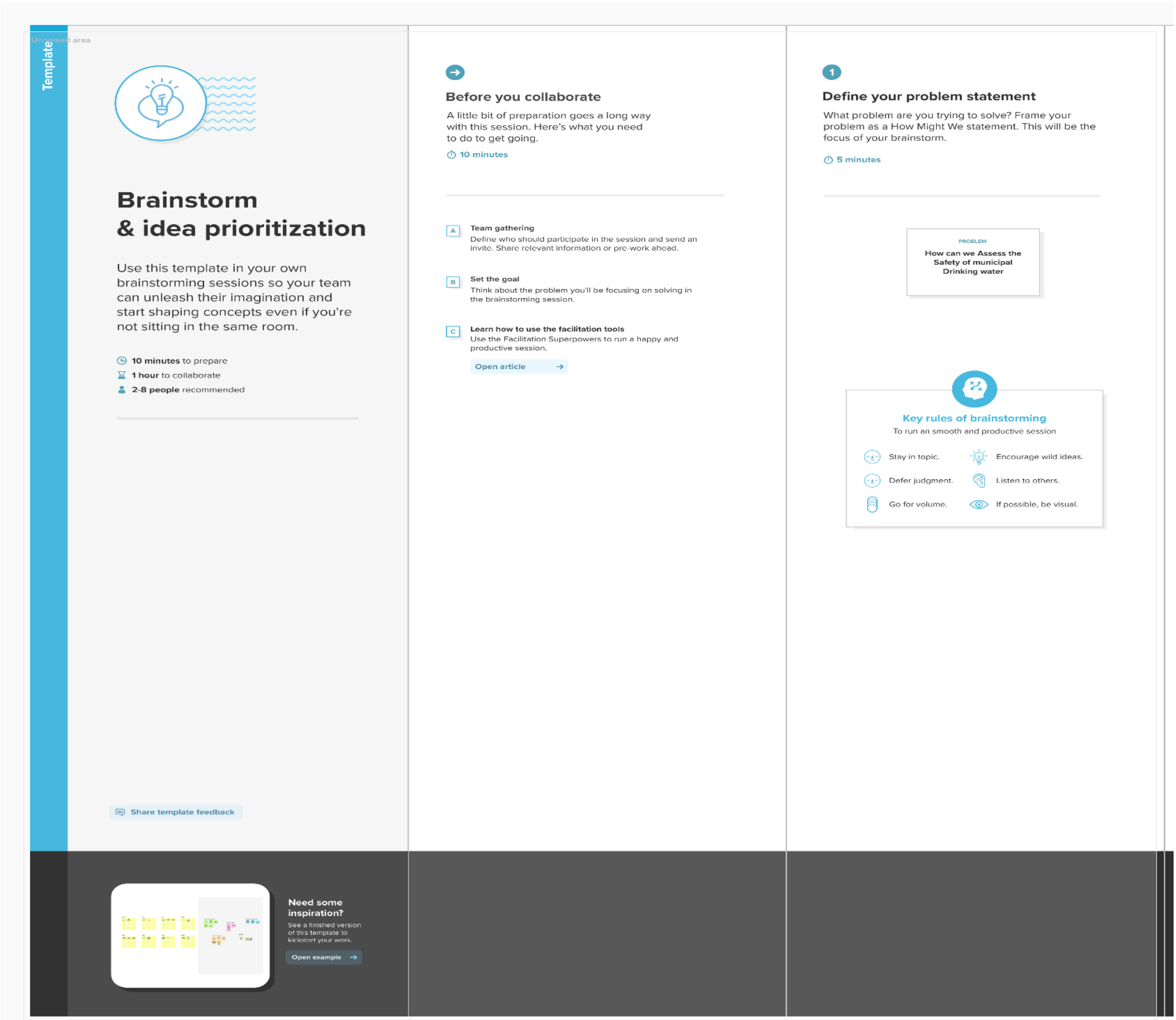


Figure 2.3: Brainstorming

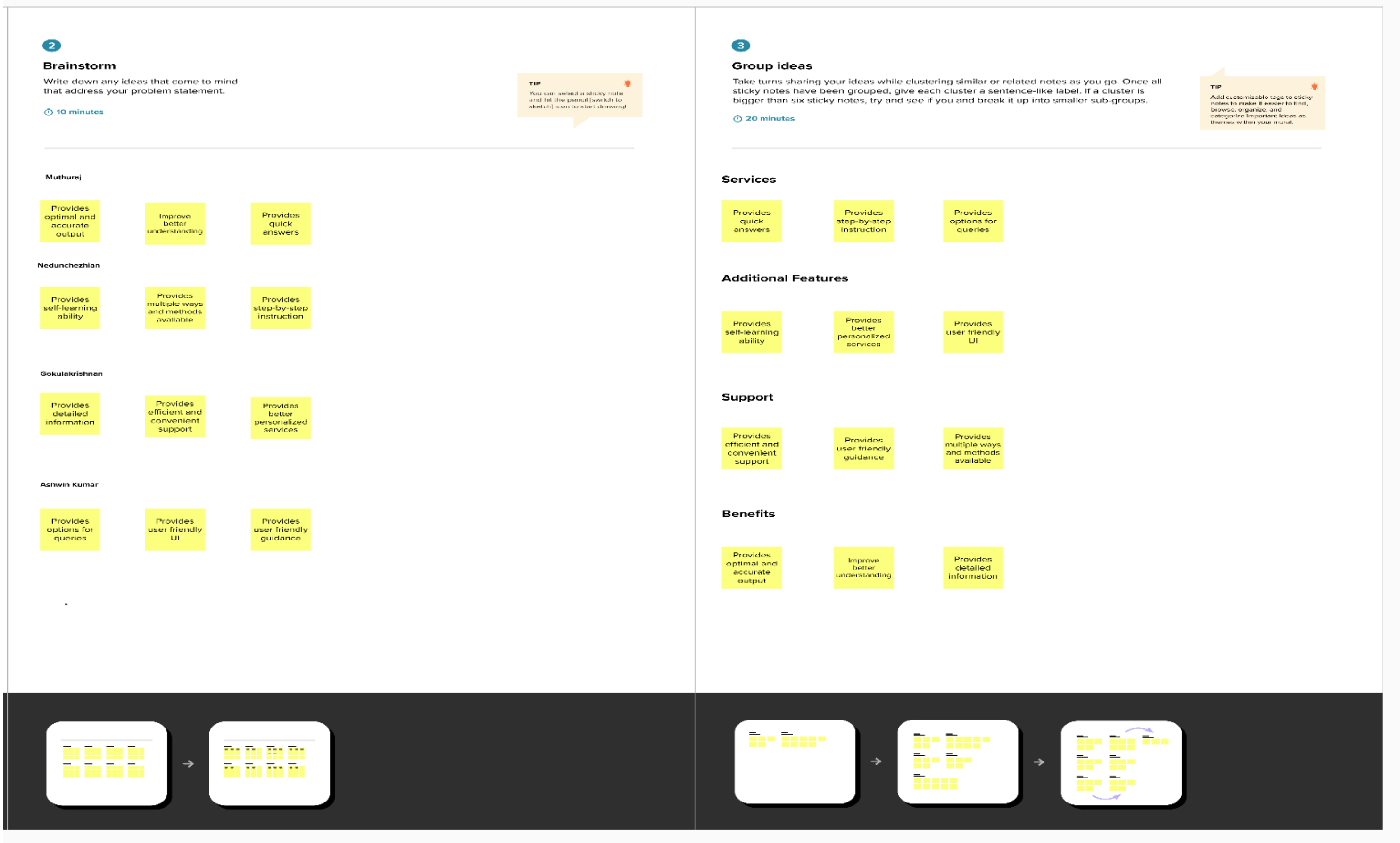


Figure 2.4: Brainstorm and Group ideas

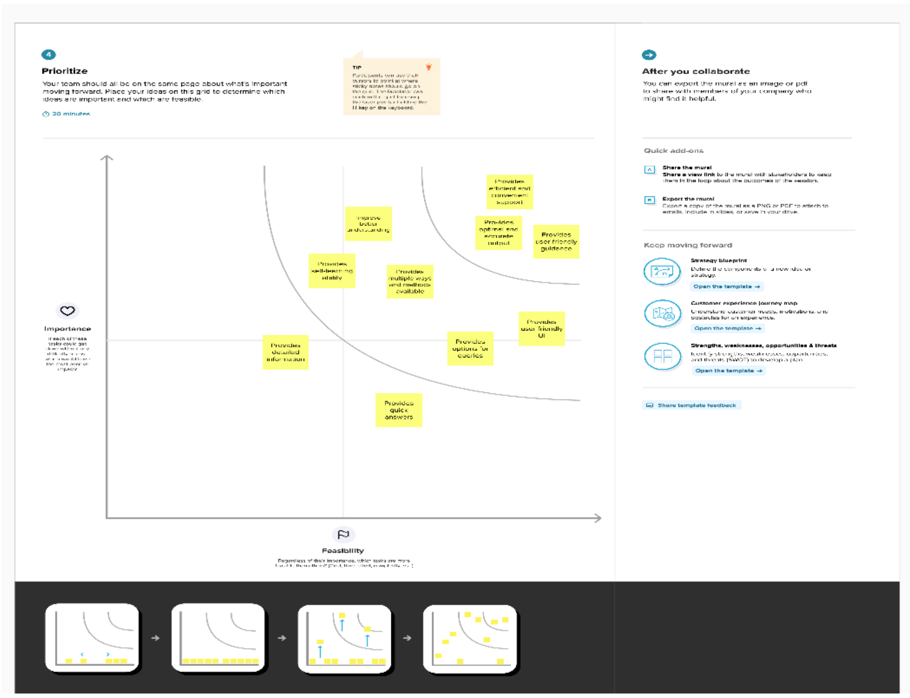


Figure 2.5: Idea Prioritization

3.3 Proposed Solution

S.No	Parameter	Description
1.	Problem Statement	• The safety of municipal drinking water is a critical public health issue, as contaminated water can lead to a variety of illnesses and diseases. The problem of contaminated water arises when harmful substances, such as bacteria, viruses, parasites, chemicals, and minerals, are present in the water supply at levels that can cause harm to human health.
2.	Idea / Solution description	• Regular water quality testing is an essential component of assessing the safety of municipal drinking water. Samples of water from various points in the distribution system should be analysed to detect the presence of contaminants such as bacteria, viruses, pesticides, heavy metals, and other harmful substances.
3.	Novelty / Uniqueness	• The use of innovative methods or technologies, such as remote sensing, machine learning, or advanced water treatment techniques, could also make a project assessing the safety of municipal drinking water unique.
4.	Social Impact / Customer Satisfaction	• Ensuring access to safe drinking water is essential for promoting equity and social justice. • Communities that lack access to safe drinking water are often those that are already marginalized and vulnerable, such as low-income communities, communities of colour, and rural communities. • Assessing the safety of municipal drinking water can help identify these disparities and take action to address them.
5.	Business Model (Revenue Model)	• Municipalities and local government bodies responsible for providing drinking water to their residents. • Service fees charged to municipalities and private organizations for assessing the safety of their drinking water supply.
6.	Scalability of the Solution	• Technology and infrastructure • Cost and resources • Regulatory compliance • Community engagement

3. REQUIREMENT ANALYSIS

3.1 Functional requirement

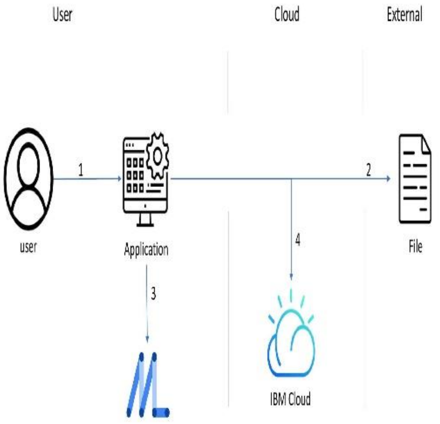
FR NO.	Functional Requirement	Sub Requirement (story/sub task)
FR-1	User Registration	• Registration through Form. • Registration through Gmail. • Registration through LinkedIn.
FR-2	User Confirmation	• Confirmation via Email.
FR-3	Assessing water safety by machine learning algorithms	• System can determine the safety of the water by • Provided factors of the water like turbidity etc., • System can continuously learn from historical data to improve accuracy. • System can generate alerts for users. • System can provide detailed reports.
FR-4	user-friendly interface for users	• Users can easily access the data history. • Users can filter and search for relevant data and activities. • Users can customize alerts and reports.
FR-5	web-based interface for users	• System is deployed on IBM Cloud platform to ensure scalability, availability, and security. • System can integrate with other systems, such as accounting and financial systems, for seamless data exchange.
FR-6	Performance System can quickly and accurately	• System can provide real-time alerts and reports to users without delay or latency. • System can handle high volumes of data for processing without sacrificing performance or accuracy.

3.2 Non-Functional requirements

NFR No.	Non-Functional Requirement	Description
NFR-1	Usability	• User-friendly interface with clear instructions and intuitive navigation.
NFR-2	Security	• Ensure data confidentiality, authorization, encryption, regular audits, and vulnerability assessments.
NFR-3	Reliability	• Consistent and error-free operation, quick recovery from errors, and avoidance of single points of failure.
NFR-4	Performance	• Timely analysis of large data volumes, fast response times, and ability to handle high loads without sacrificing performance.
NFR-5	Availability	• Minimal downtime or maintenance, avoidance of scheduled maintenance, and disaster recovery plan in place.
NFR-6	Scalability	• Designed to handle growth in data volume and user numbers, ability to scale up or down as needed, and handle peak loads without additional resources.

4. PROJECT DESIGN

4.1 Data Flow Diagrams



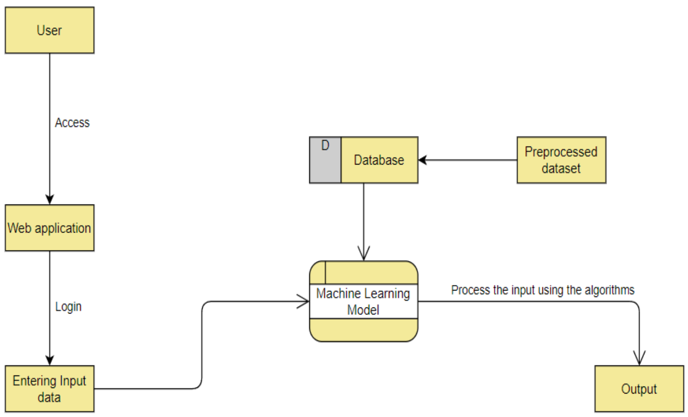


Figure 4.1: System Architecture **Figure 4.2: Data flow diagrams**

1. The user login into the webpage.

2. The User select and load the data needed to be processed and identified.

3. The input data is given to the machine learning model, and it process them with given algorithms and give the output data.

4. Dataset needed is stored in the IBM Cloud storage.

4.2 Solution & Technical Architecture

Solution Architecture:

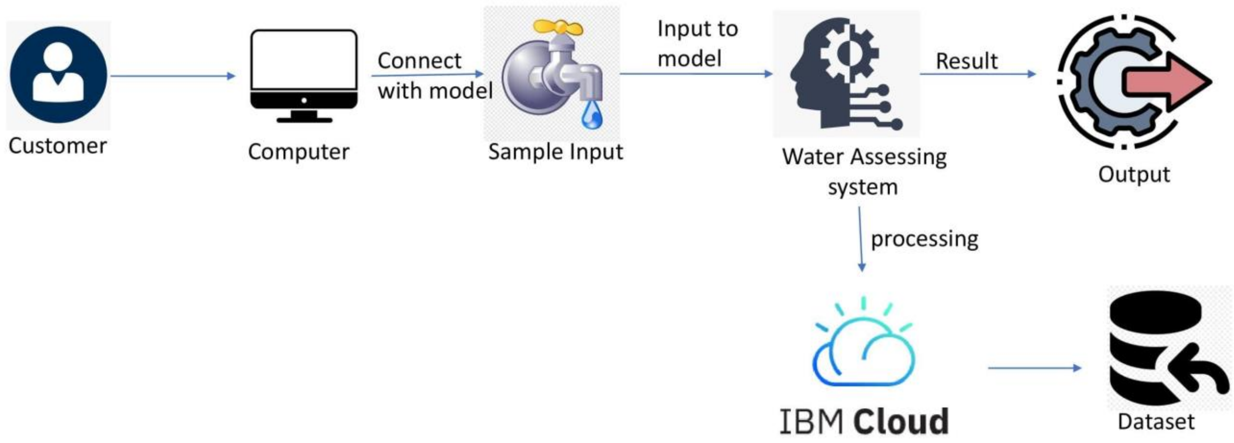


Figure 4.3: Solution Architecture for Prediction

Technical Architecture:

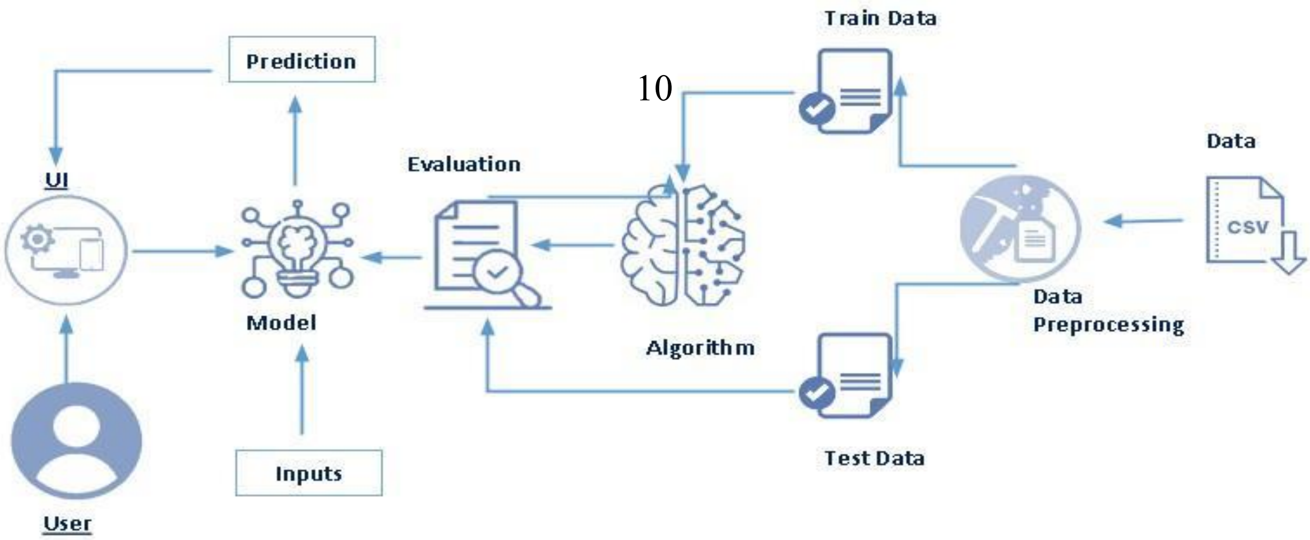


Figure 4.4: Technical Architecture for Prediction

4.3 User Stories

User Type	Functional Requirement	User story number	User story/task	Acceptance criteria	Priority	Team Member
Water treatment plant operators	Data collection and analysis	USN-1	I need to ensure that the water is properly treated and meets all the regulatory requirements for safe drinking water.	The System determines the quality of water by provided data.	High	Muthuraj
Water quality inspectors	Real time water quality analysis	USN-2	I will check the quality of the water standards which needed to be satisfied.	The expert checks output data of the water with standard requirement data.	High	Nedunchezhian
Public health officials	Monitoring Public health	USN-3	As an expert, I need to monitor the public's health who consumes the water.	The expert checks the medical records of the area.	High	Gokula Krishnan
Water consumers	User Friendly interface	USN-4	As a consumer, I need to check the quality of the water I consume.	The user can check the safety of the water by using the model.	Medium	Ashwin Kumar

5. CODING & SOLUTIONING

7.1 Feature 1:

Python Flask

Python Flask is mainly used to render and integrate the web application with the machine learning model in the browser. By running the python application, the suitable server domain link is obtained and run in the browser.

HTML

The HTML and CSS is used to design the overall web application. HTML is used to add UI components and CSS is used to add style to those components. These two are integrated with the model and opened in web browsers.

Build PYTHON FLASK Code:

App.py

```
from flask import Flask, request, render_template
import joblib
import numpy as np

app = Flask(__name__)

model = joblib.load("water_quality.joblib")
@app.route("/")
def f():
   return render_template("index.html")
@app.route("/inspect")
def inspect():
```

```
    return render_template("inspect.html")
@app.route("/home", methods=["GET", "POST"])
def home():
    if request.method == 'POST':
        var1 = float(request.form["pH_Value"])
        var2 = float(request.form["Hardness"])
        var3 = float(request.form["Solids"])
        var4 = float(request.form["Chloramines"])
        var5 = float(request.form["Sulfate"])
        var6 = float(request.form["Conductivity"])
        var7 = float(request.form["Organic_carbon"])
        var8 = float(request.form["Trihalomethanes"])
        var9 = float(request.form["Turbidity"])

        input_data = np.array([[var1, var2, var3, var4, var5, var6, var7, var8,
var9]])
        prediction = model.predict(input_data)

        if prediction == 0:
            return render_template('output.html', predict="Portable")
        else:
            return render_template('output.html', predict="Not Portable")
        return render_template('index.html')
if __name__ == "__main__":
    app.run(debug=True)
```

and 3 html files are added with the model using this flask code in the visual studio IDE.

7.2 Feature 2:

Web Application:

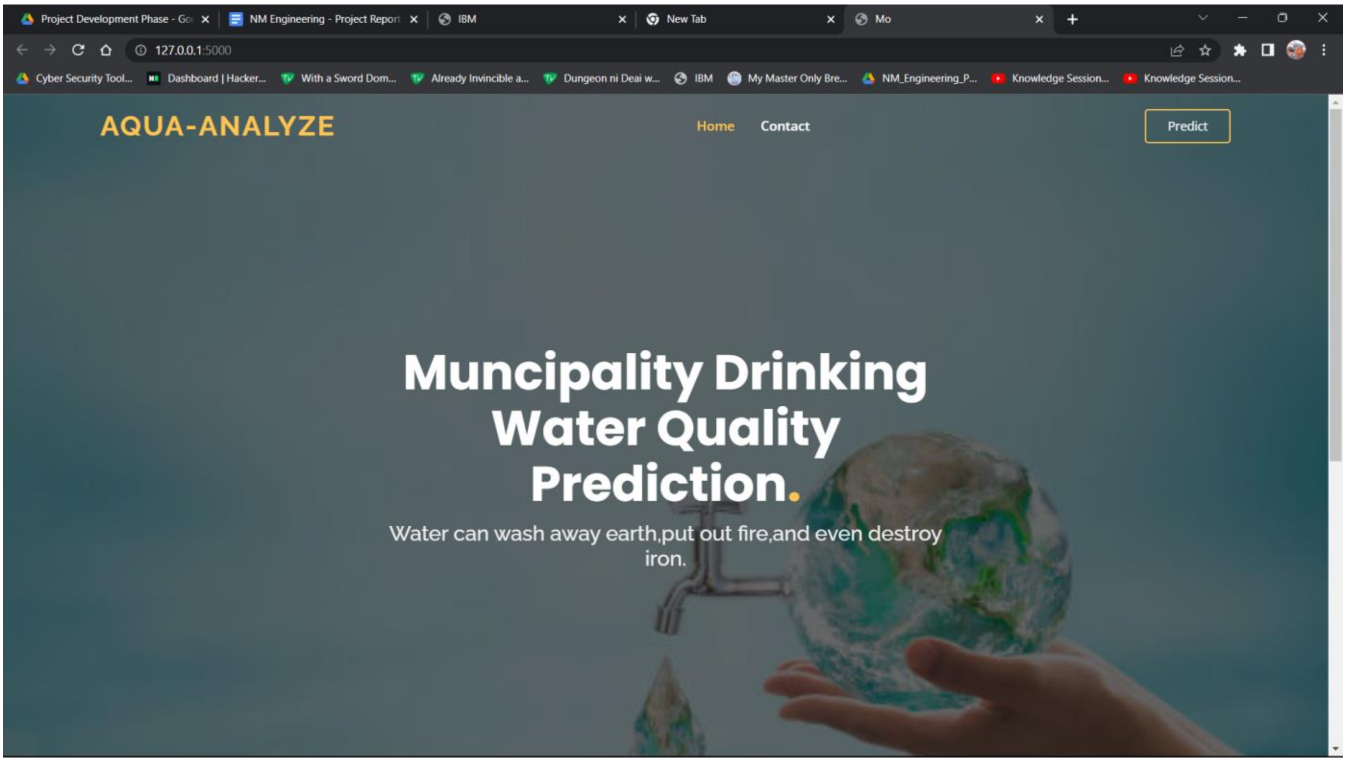


Figure 7.1: Index Page of the Application

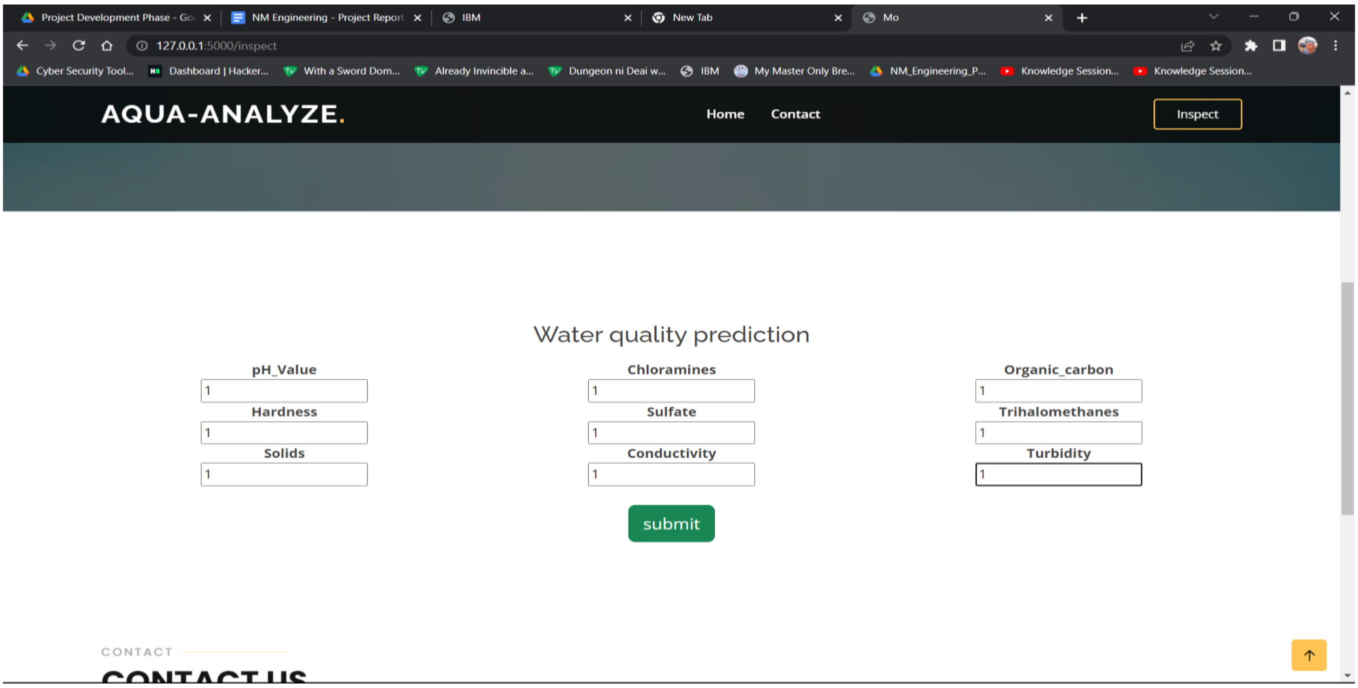


Figure 7.2: Inspect Page of the Application

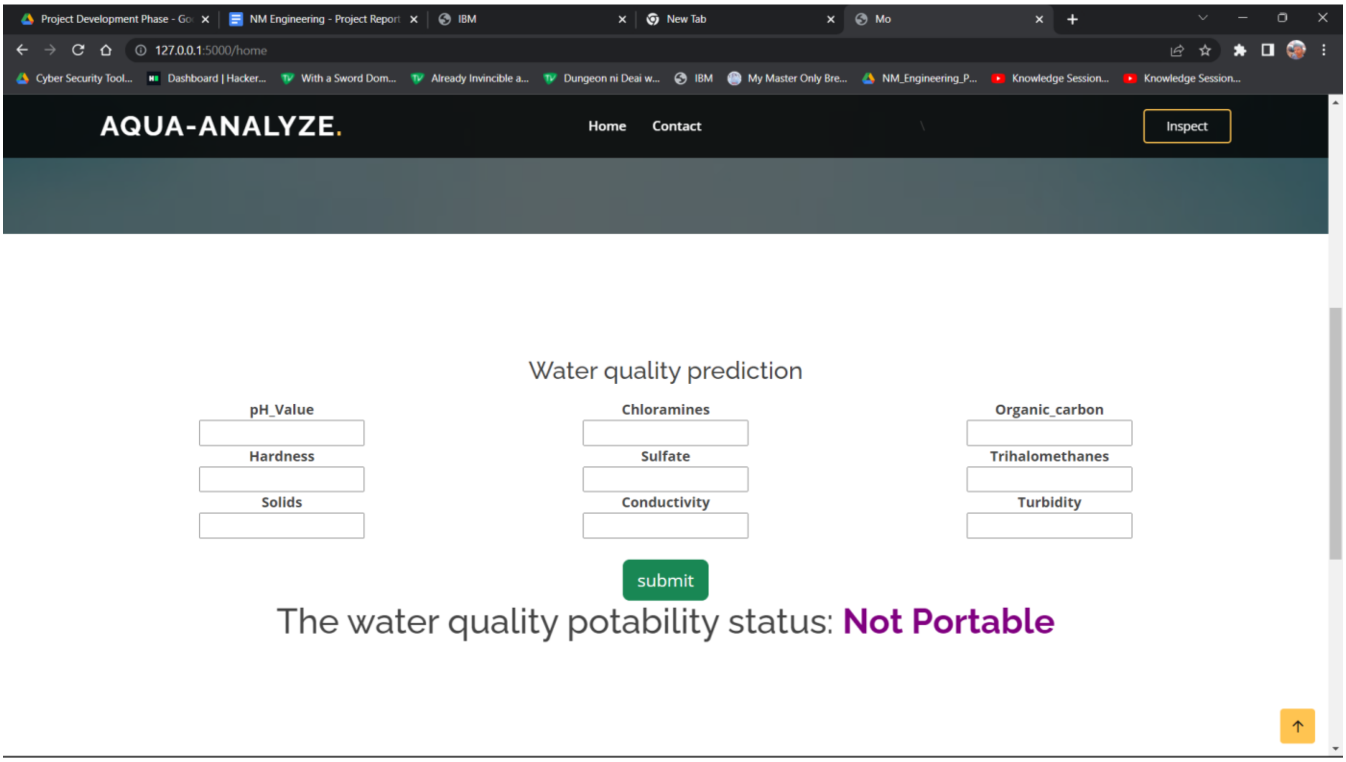
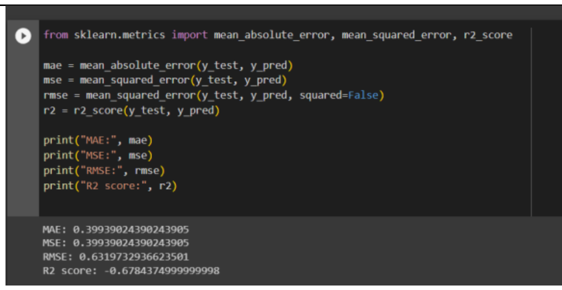


Figure 7.3: Output Page of the Application

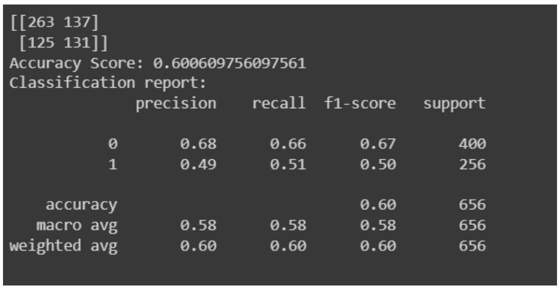
6. RESULTS

6.1 Performance Metrics:

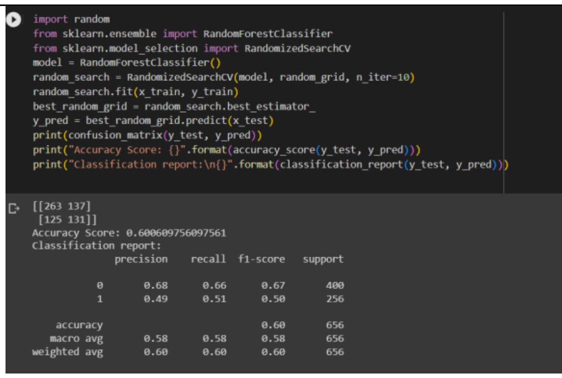
S.No	Parameter	Values	Screenshot
1.	Metrics	Regression Model: MAE -, MSE - , RMSE - , R2 score - Classification Model: Confusion Matrix -, Accuracy Score- & Classification Report	(see below)
2.	Tune the Model	Hyperparameter Tuning - Validation Method -	



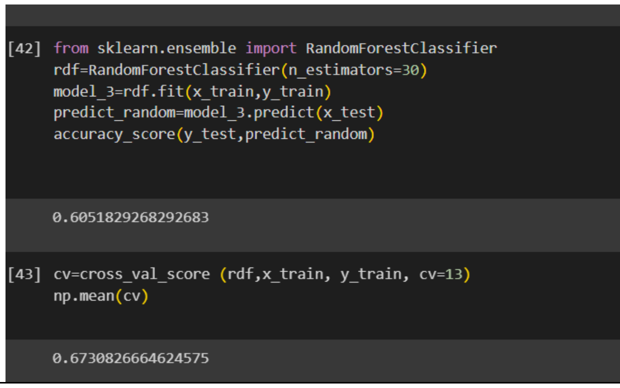
```
from sklearn.metrics import mean_absolute_error, mean_squared_error, r2_score

mae = mean_absolute_error(y_test, y_pred)
mse = mean_squared_error(y_test, y_pred)
rmse = mean_squared_error(y_test, y_pred, squared=False)
r2 = r2_score(y_test, y_pred)

print("MAE:", mae)
print("MSE:", mse)
print("RMSE:", rmse)
print("R2 score:", r2)

MAE: 0.39939024390243905
MSE: 0.39939024390243905
RMSE: 0.6319732936623501
R2 score: -0.6784374999999998
```



```
[[263 137]
 [125 131]]
Accuracy Score: 0.600609756097561
Classification report:
              precision    recall  f1-score   support

           0       0.68      0.66      0.67       400
           1       0.49      0.51      0.50       256

    accuracy                           0.60       656
   macro avg       0.58      0.58      0.58       656
weighted avg       0.60      0.60      0.60       656
```



```
import random
from sklearn.ensemble import RandomForestClassifier
from sklearn.model_selection import RandomizedSearchCV
model = RandomForestClassifier()
random_search = RandomizedSearchCV(model, random_grid, n_iter=10)
random_search.fit(x_train, y_train)
best_random_grid = random_search.best_estimator_
y_pred = best_random_grid.predict(x_test)
print(confusion_matrix(y_test, y_pred))
print("Accuracy Score: {}".format(accuracy_score(y_test, y_pred)))
print("Classification report:\n{}".format(classification_report(y_test, y_pred)))

[[263 137]
 [125 131]]
Accuracy Score: 0.600609756097561
Classification report:
              precision    recall  f1-score   support

           0       0.68      0.66      0.67       400
           1       0.49      0.51      0.50       256

    accuracy                           0.60       656
   macro avg       0.58      0.58      0.58       656
weighted avg       0.60      0.60      0.60       656
```



```
[42] from sklearn.ensemble import RandomForestClassifier
     rdf=RandomForestClassifier(n_estimators=30)
     model_3=rdf.fit(x_train,y_train)
     predict_random=model_3.predict(x_test)
     accuracy_score(y_test,predict_random)

     0.6051829268292683

[43] cv=cross_val_score (rdf,x_train, y_train, cv=13)
     np.mean(cv)

     0.6730826664624575
```

7. ADVANTAGES & DISADVANTAGES

Advantages:

1. **Public Health Protection:** Regular assessment of drinking water safety helps protect public health by identifying potential contaminants and taking appropriate measures to mitigate them. It ensures that the water supply meets the necessary quality standards and regulations, reducing the risk of waterborne diseases and illnesses.

2. **Disease Prevention:** Assessments can help identify and address sources of contamination, such as bacteria, viruses, parasites, and chemical pollutants. By detecting and preventing the presence of harmful substances in the water, the risk of waterborne diseases, including gastrointestinal illnesses, cholera, typhoid, and hepatitis, can be significantly reduced.

3. **Early Detection of Problems:** Regular assessments allow for the early detection of potential issues in the water supply system. By monitoring key indicators and conducting routine testing, any deviations from the acceptable levels can be identified promptly. This enables authorities to investigate the cause of the problem and implement corrective measures before it escalates into a larger health or environmental issue.

Disadvantages:

1. **Limited Data Availability:** Machine learning models rely on large amounts of high-quality data for training and accurate predictions. However, there may be limitations in the availability or quality of data related to drinking water safety. In some cases, historical data may be incomplete or inconsistent, making it challenging to train robust models.

2. **Data Bias and Representativeness:** Machine learning models are only as good as the data they are trained on. If the training data used for the model is biased or not representative of the entire population or geographical area, it can lead to skewed results. For example, if certain areas or communities are underrepresented in the training data, the model may not accurately reflect the unique challenges and risks they face.

3. **Model Interpretability and Transparency:** Machine learning models, particularly complex ones like deep learning models, can lack interpretability and transparency. It may be challenging to understand how the model arrives at its predictions or assess the factors contributing to those predictions. This can be problematic when it comes to building trust, understanding potential biases, and effectively communicating the results to stakeholders.

8. CONCLUSION

In conclusion, using machine learning models to assess the safety of municipal drinking water can offer numerous benefits but also comes with its challenges. While machine learning models have the potential to improve the accuracy and efficiency of water safety assessments, it is important to consider the limitations and potential disadvantages associated with their implementation. The availability and quality of data, as well as potential biases and representativeness issues, can impact the reliability of machine learning models. Implementing machine learning models for water safety assessments may require substantial investments in resources and infrastructure, which can be a limitation for smaller municipalities or regions with limited budgets. Considering these factors, a holistic approach that combines machine learning models with human expertise, ongoing monitoring, and validation processes can help overcome the challenges and maximize the benefits of using machine learning for assessing the safety of municipal drinking water. By addressing data limitations, promoting transparency, and incorporating domain knowledge, machine learning can be a valuable tool in improving the overall effectiveness of water safety assessments and safeguarding public health.

9. FUTURE SCOPE

1. **Enhanced Data Collection and Integration:** Improvements in data collection methods, such as sensor technologies and IoT devices, can provide real-time data on water quality parameters. Machine learning models can integrate and analyse these diverse data sources to provide more accurate and timely assessments of water safety.
2. **Predictive Modelling for Early Warning Systems**: Machine learning models can be trained to detect patterns and predict potential water quality issues in advance. By analysing historical data, weather conditions, and other relevant factors, these models can help create early warning systems to identify and mitigate potential risks before they escalate.
3. **Adaptive and Dynamic Models:** Future models can be designed to adapt to the dynamic nature of water quality by incorporating real-time monitoring and feedback mechanisms. These adaptive models can continuously learn and update their predictions based on new data, enabling more accurate and up-to-date assessments of water safety.
4. **Integration of Multiple Data Sources:** Machine learning models can leverage data from various sources, such as satellite imagery, social media, and health records, to improve water safety assessments. By incorporating additional contextual data, these models can provide a more comprehensive understanding of the factors influencing water quality and health outcomes.
5. **Decision Support Systems:** Machine learning models can be integrated into decision support systems that assist water utility managers and policymakers in making informed decisions. These systems can provide actionable insights and recommendations based on the analysis of large-scale data, helping optimize resource allocation and prioritize interventions.

10. APPENDIX

10.1 REFERENCES:

1. Fu, Zhao, "Water Quality Prediction Based on Machine Learning Techniques" (2020). UNLV Theses, Dissertations, Professional Papers, and Capstones. 3994. http://dx.doi.org/10.34917/22110053

2. ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538 Volume 10 Issue VI June 2022- Available at www.ijraset.com ©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 | 4173. Water Quality Prediction using Machine Learning Nishant Rawat1 , Mangani Daudi Kazembe2 , Pradeep Kumar Mishra3 1, 2, 3Department of Computer Science and Engineering, Sharda University, Greater Noida. https://www.ijraset.com/best-journal/water-quality-prediction-using-machine-learning

3. A Machine Learning-Based Water Potability Prediction Model by Using Synthetic Minority Oversampling Technique and Explainable AI-2022 Jinal Patel,Charmi Amipara,Tariq Ahamed Ahanger, Komal Ladhva, Rajeev Kumar Gupta, Hashem O. Alsaab, Yusuf S. Althobaiti, and Rajnish Ratna. https://doi.org/10.1155/2022/9283293

4. A Water Quality Prediction Model Based on Multi-Task Deep Learning: A Case Study of the Yellow River, China Xijuan Wu, Qiang Zhang, Fei Wen and Ying Qi – 2022 https://www.mdpi.com/2073-4441/14/21/3408/pdf

10.2 GITHUB LINK:

https://github.com/naanmudhalvan-SI/PBL-NT-GP--5700-1680798847

10.3 PROJECT DEMO LINK:

https://www.youtube.com/watch?v=4UpY3k67pzQ