EFFICIENT WATER QUALITY ANALYSIS AND PREDICTION USING MACHINE LEARNING

TEAM ID: PNT2022TMID39608

TEAM MEMBERS:

1. **S.SUDHARSAN**
2. **N.SIDDHARTH**
3. **F.SUHAIL**
4. **M.VIGNESH**

PROBLEM STATEMENT

- Water is considered as a vital resource that affects the various aspects of human health and lives. The quality of water is a major concern for people living in the urban areas.
- Quality of water serves as a powerful environmental determinant and a foundation for the prevention and control of water borne 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 water usage patterns, temperature, pH, total amount of Dissolved solids, etc.
- So it has created a necessity to determine the quality of water samples so as to determine and detect the contaminants present in advance.

PURPOSE

The main purpose of this project is :

- To create a safe and healthy environment by preventing the water borne diseases among normal public by testing the Quality of the sources of water they consume.
- Develop a Machine Learning model that predicts the Water Quality Index of any kind of water samples.
- Create a model such that it should trust and belief among the customers who are using the system to predict the water sample they utilize for their needs.

PROJECT OBJECTIVES

- To predict the Water Quality Index (WQI) for any kind of water samples using the Machine Learning Regression Algorithms based on the important parameters like temperature, pH, total amount of Dissolved and Suspended solids, etc.
- To provide any information regarding the purification technique recommended based on the impurities present in the water sample tested.
- To predict the Water Quality Classification (WQC) using Machine Learning Classification algorithms.

SCOPE OF THE PROJECT

Create a model in such a way that it can be used by all kinds of people like:

- Water Quality Testing Agencies
- Private and Public Laboratories
- Various industries like Textile industry, Cotton industry, etc.
- General Public for testing the water they consume for household purposes.

- It should also possess the ability to sustain the changes by testing any kind of water samples and also to have broader scope of integrating the model with Future Technologies also.

LITERATURE REVIEW

PREDICTING OF WATER QUALITY USING MULTIVARIATE LINEAR REGRESSION :

- This model is used to measure the Biological Oxygen Demand (BOD), Chemical Oxygen Demand (COD) based on the following four parameters: Temperature, pH, total suspended solids and dissolved solids.
- This approach proposed that the deterministic and multivariate linear regression models were used to speed up the process of predicting the water quality.
- As the dataset is considered as time based series, so it is likely to have non linear relationship. So, the performance of this approach is expected to be poor, with large prediction error.

APPLICATION ADAPTIVE NEURO FIZZY INFERENCE SYSTEM (ANFIS):

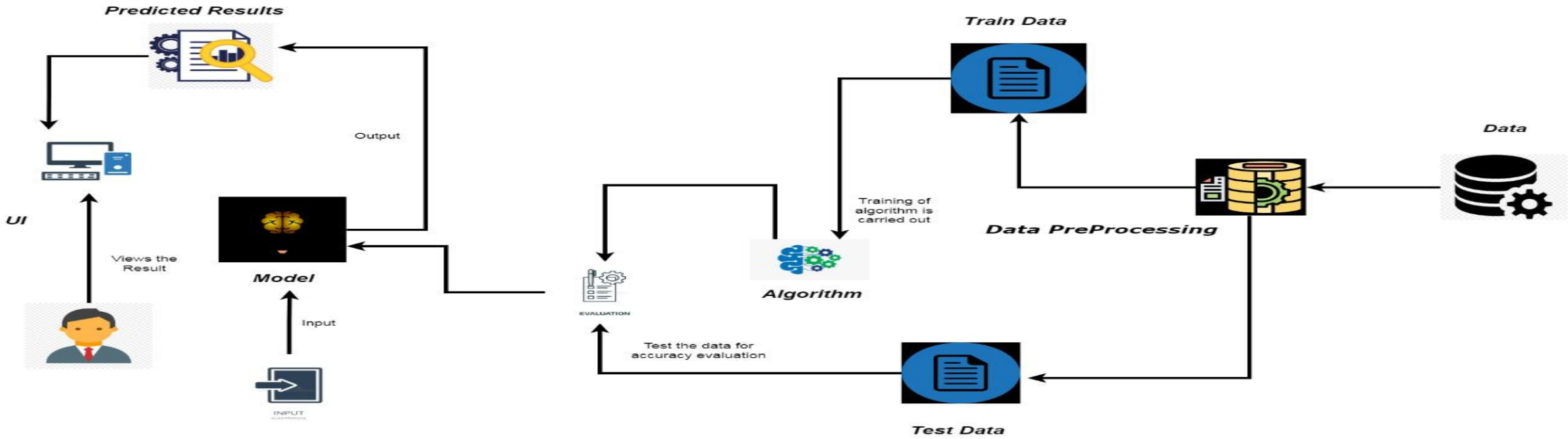
- Many studies have proven that ANFIS, which can integrate linear and non linear relationships hidden in the dataset, is a better option in predicting the effluent water quality and also shows that the ANFIS model works better than the ANN (Artificial Neural Network) model in predicting the Dissolved Oxygen content in the water sample to be tested.
- The model predicts the water quality based on Eight Parameters to predict the total phosphorous and total nitrogen content in the water sample. The model could accurately formulate the hidden relationships and correlation analysis can improve the prediction accuracy.
- The disadvantage is that this proposed model approach requires that the size of training dataset should not be less than the number of training parameters and if the correlation between the data in the dataset are weak then it generates out of range errors.

SOLUTION ARCHITECTURE

Project Design Phase - I

Solution Architecture

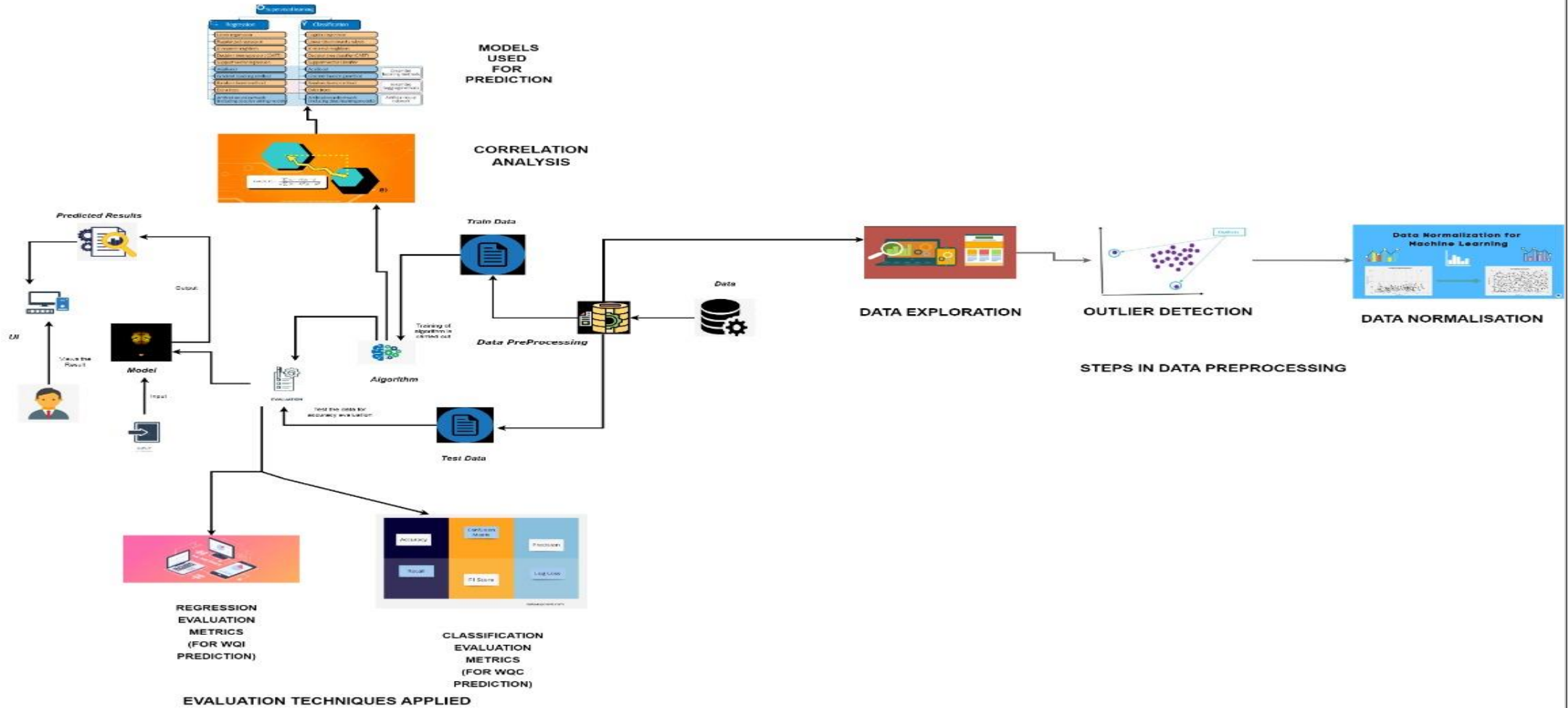
Date	7 October 2022
Team ID	PNT2022TMID39608
Project Name	Efficient Water Quality Analysis and Prediction using Machine Learning
Maximum Marks	4 Marks



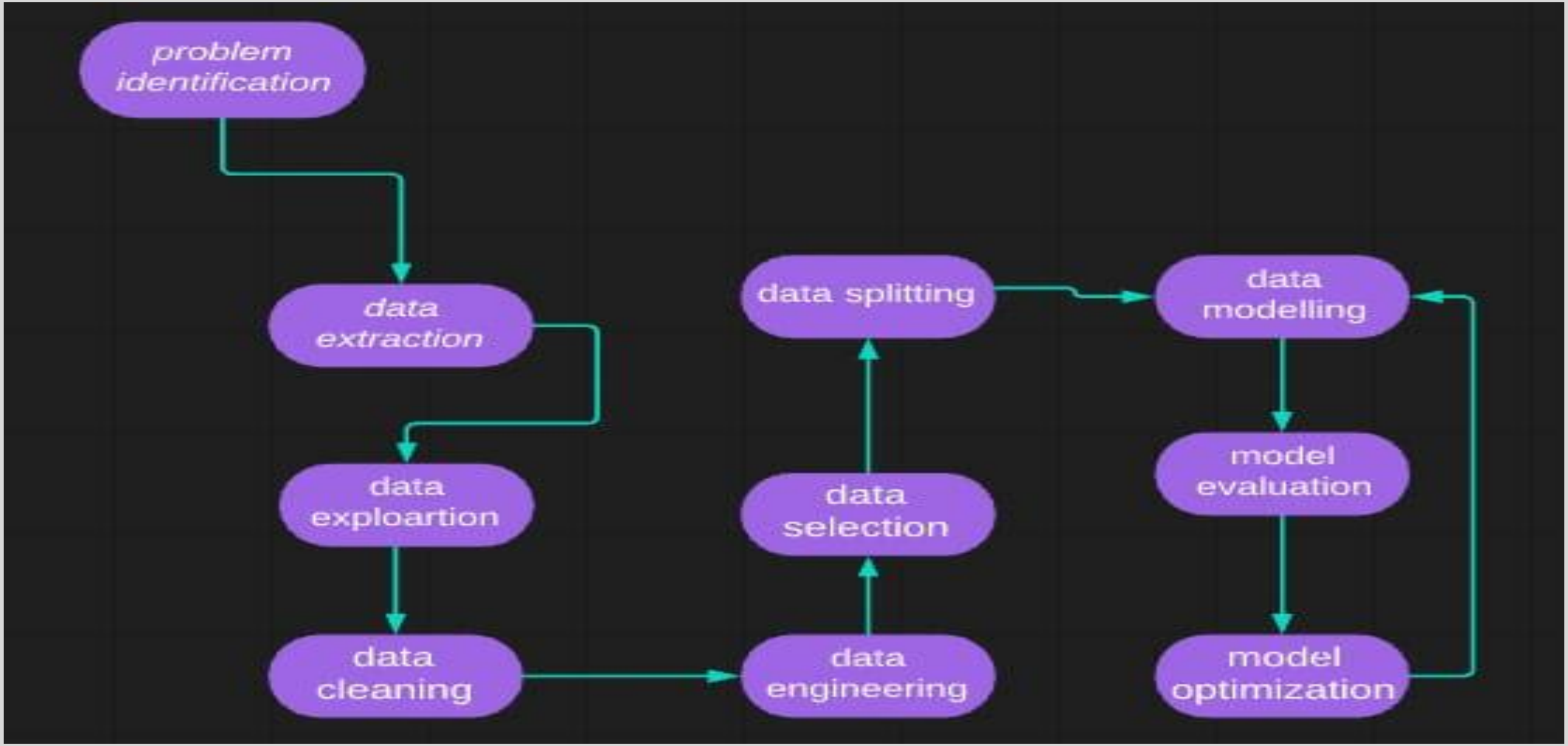
TECHNOLOGY ARCHITECTURE

TECHNICAL ARCHITECTURE

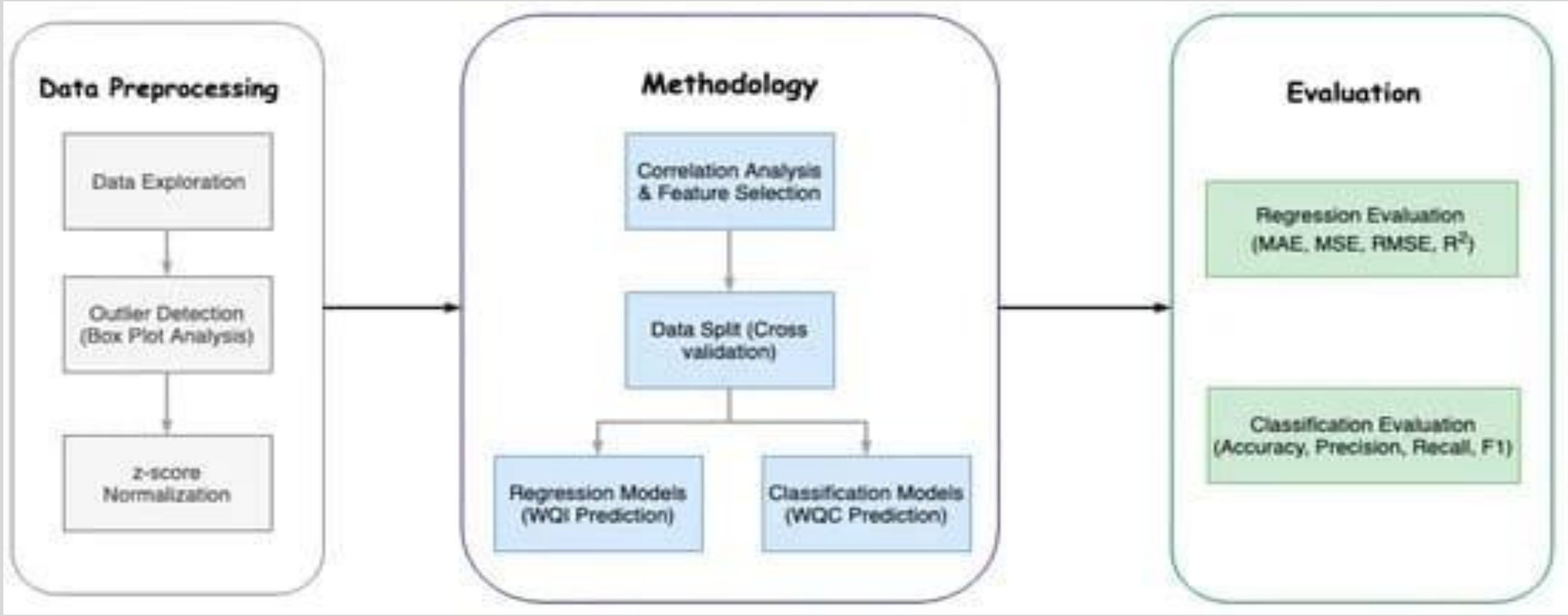
METHODOLOGY APPLIED



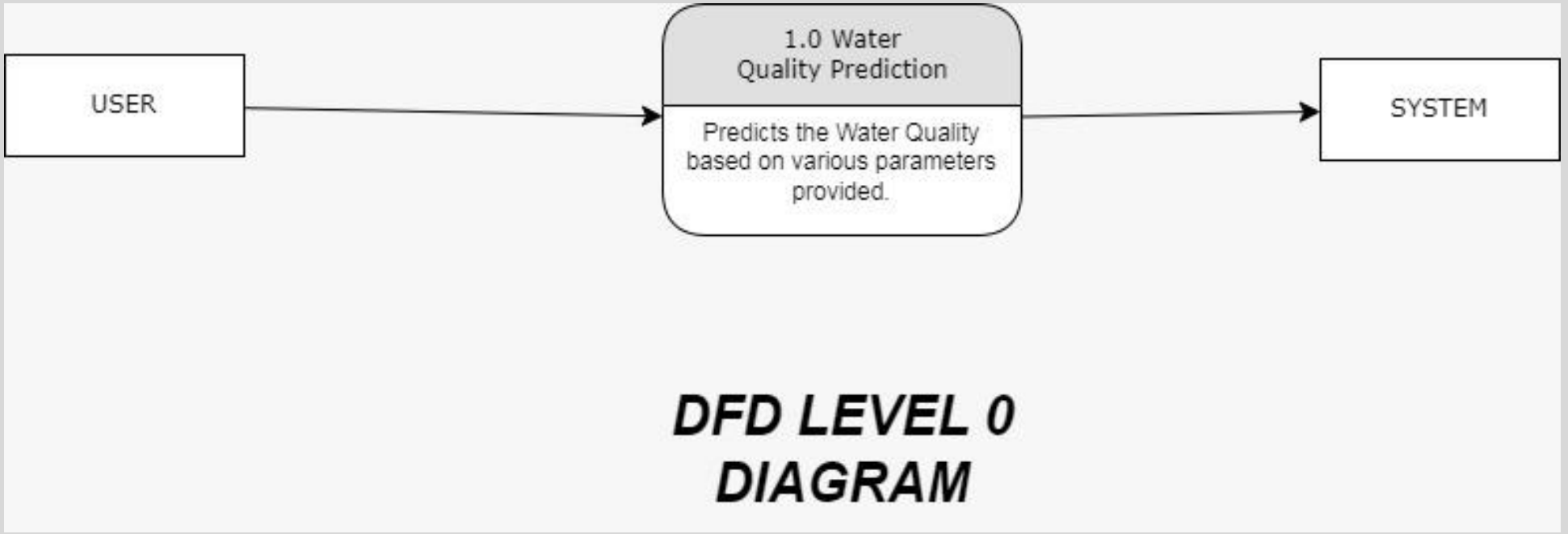
PROCESS INVOLVED



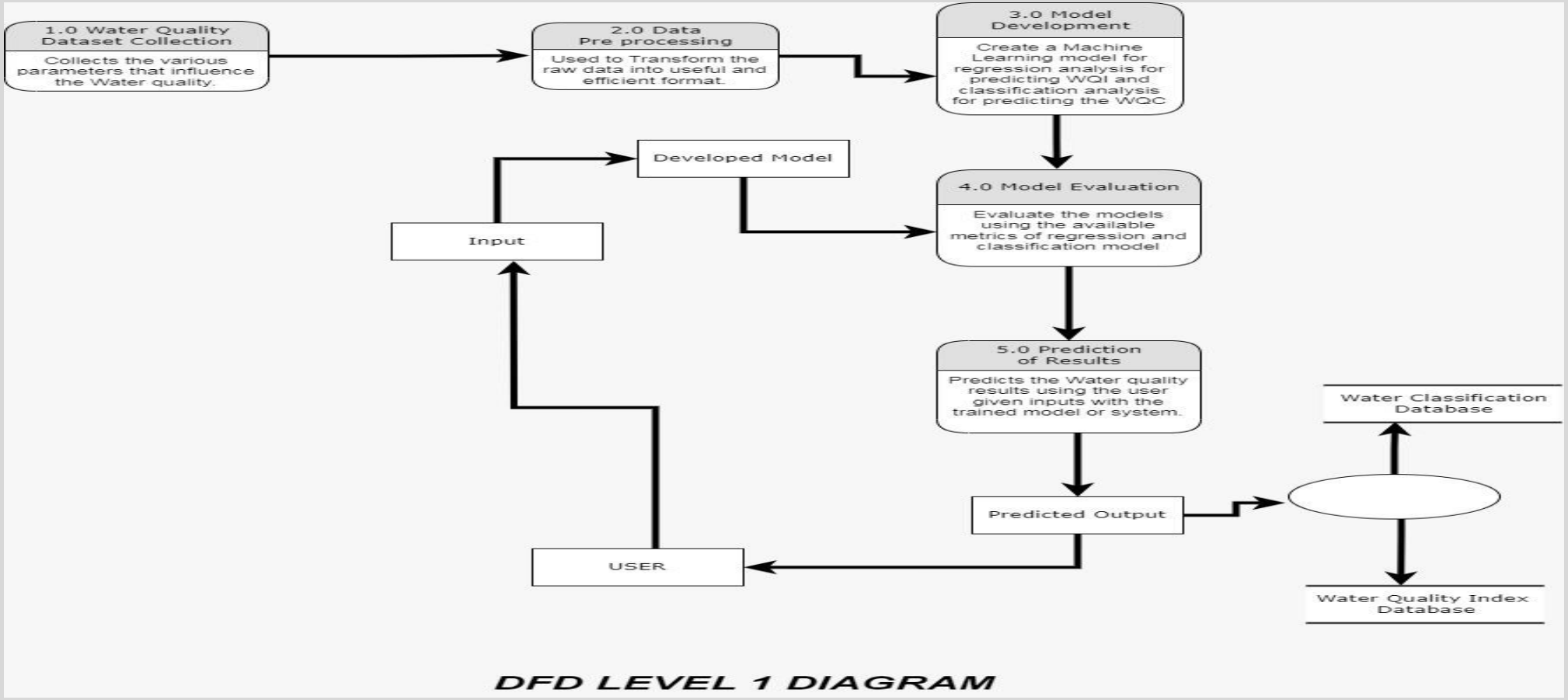
METHODOLOGIES APPLIED IN EACH PHASES



DFD LEVEL 0 DIAGRAM



DFD LEVEL 1 DIAGRAM



DFD LEVEL 1 DIAGRAM

PROJECT FLOW

Dataset

Collect the Dataset or create the Dataset.

Data Preprocessing.

Importing the libraries.

Importing the Dataset.

Checking for Null Values.

Data Visualization.

Taking care of Missing Data.

Label Encoding.

Splitting the Data into Train and Test.

Model Building

Training and Testing the Model.

Evaluation of Model.

Application Building

Create an HTML file.

Build a Python code.

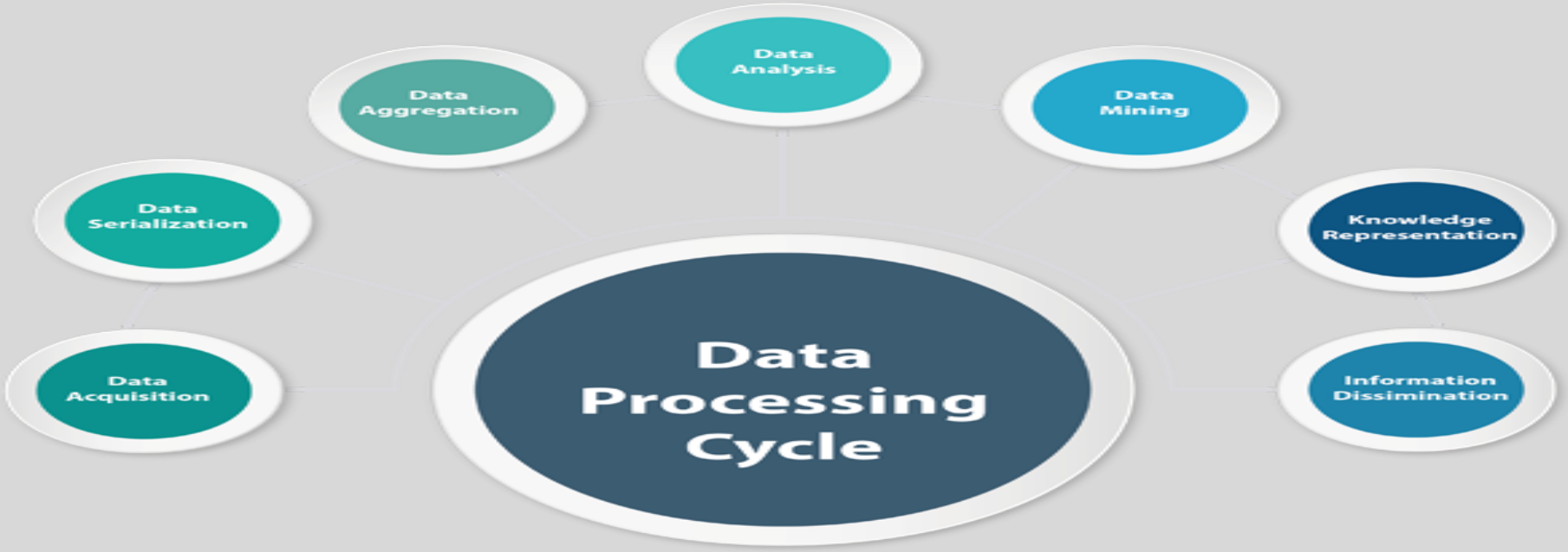
DATASET

- It contains the different samples from different Indian states and has 7 significant parameters namely

1. **Dissolved Oxygen**
2. **pH**
3. **Conductivity**
4. **Biological Oxygen Demand**
5. **Nitrate**
6. **Fecal Coliform**
7. **Total Coliform**

DATA PREPROCESSING

- It is an important phase in the Data analysis to improve the data quality.
- The WQI has been calculated from the most significant parameters of the dataset.
- Then the water samples have been classified on the basis of the WQI values.



MODEL BUILDING

- Building an ML model requires splitting of Data into two sets such as

1. Training Set
2. Testing Set

A set of supervised (for labelled) and unsupervised (for unlabeled data) algorithms are available to be chosen depending on the Nature of Data and business outcome to predict.

We have chosen "Random Forest Regression" Algorithm as the model as we are predicting a numeric data WQI value based on the various numeric basic Water Quality Parameters.

APPLICATION BUILDING

- We are using various Application Frameworks for creating an Interactive and User Friendly application so that the users can predict the Water Quality Index (WQI) value based on the inputs given by the user on the interface.
- Application Framework:
- It is a software library that provides a fundamental structure to

support the development of applications for a specific environment.

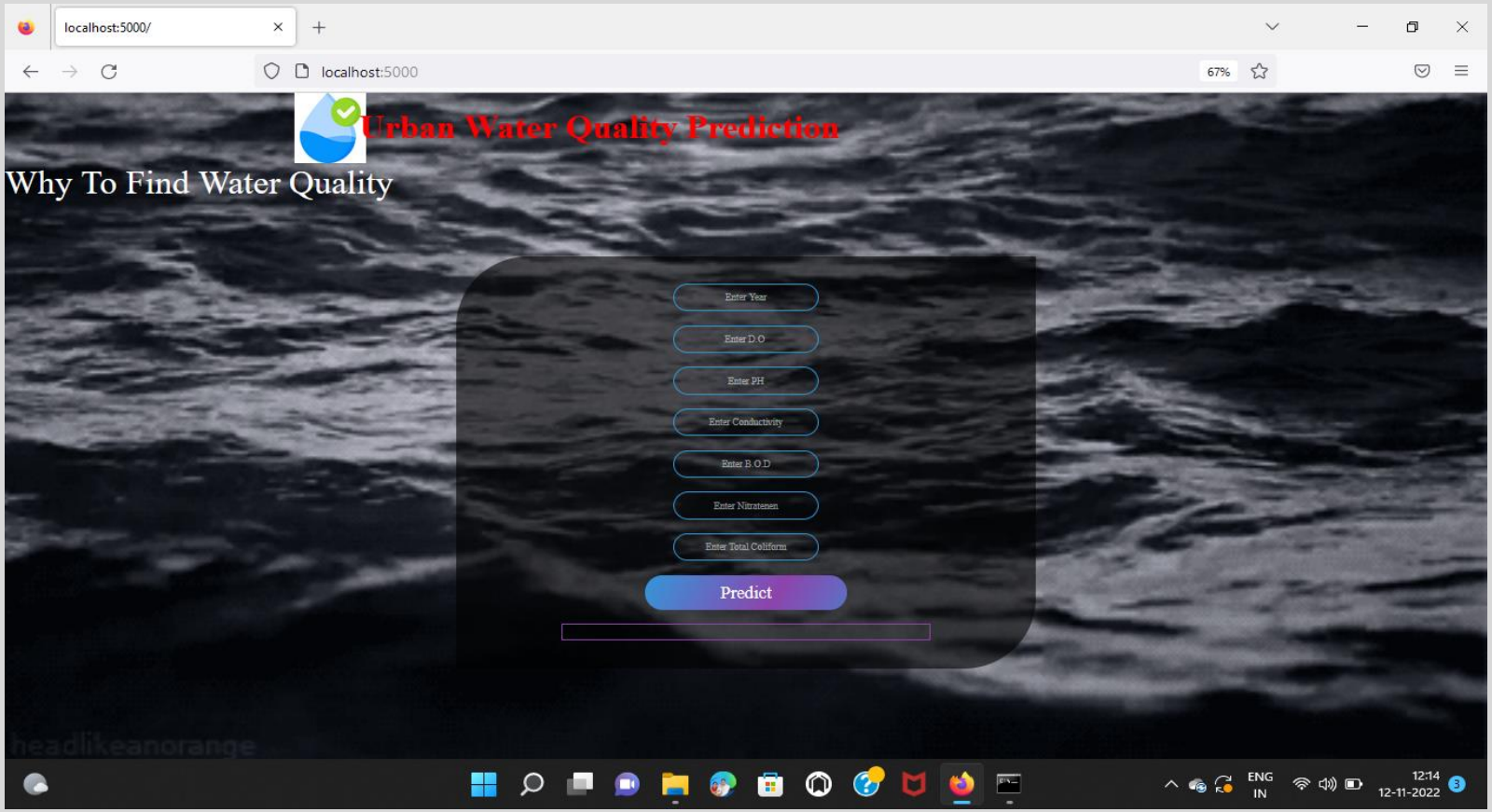
- Application Frameworks Used:

1. Flask (Light Weight WSGI web application framework.
2. Anaconda Navigator

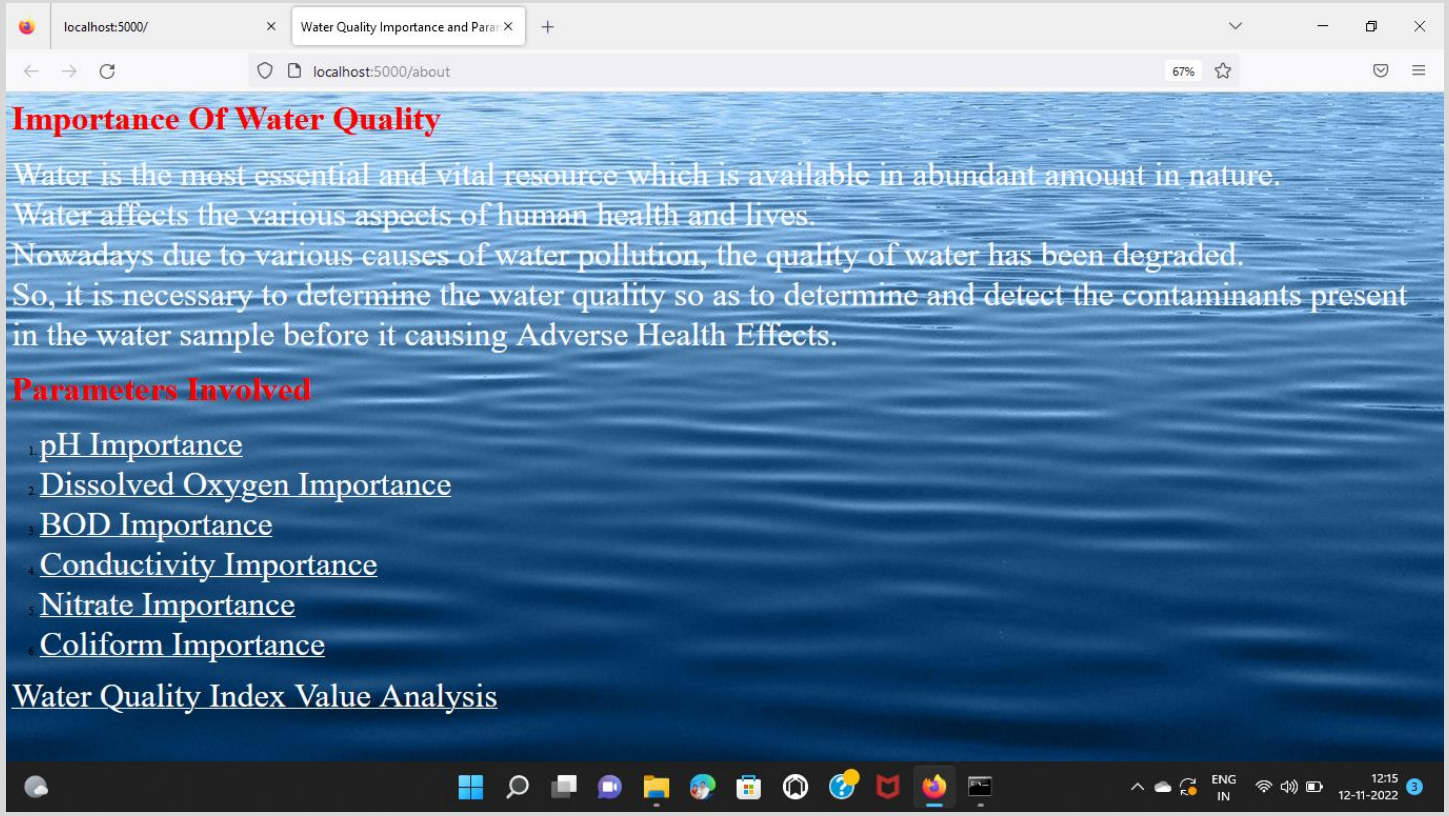
*(Distribution of R and Python Programming languages for scientific computing.

OUTPUT

- **User Interface Page**

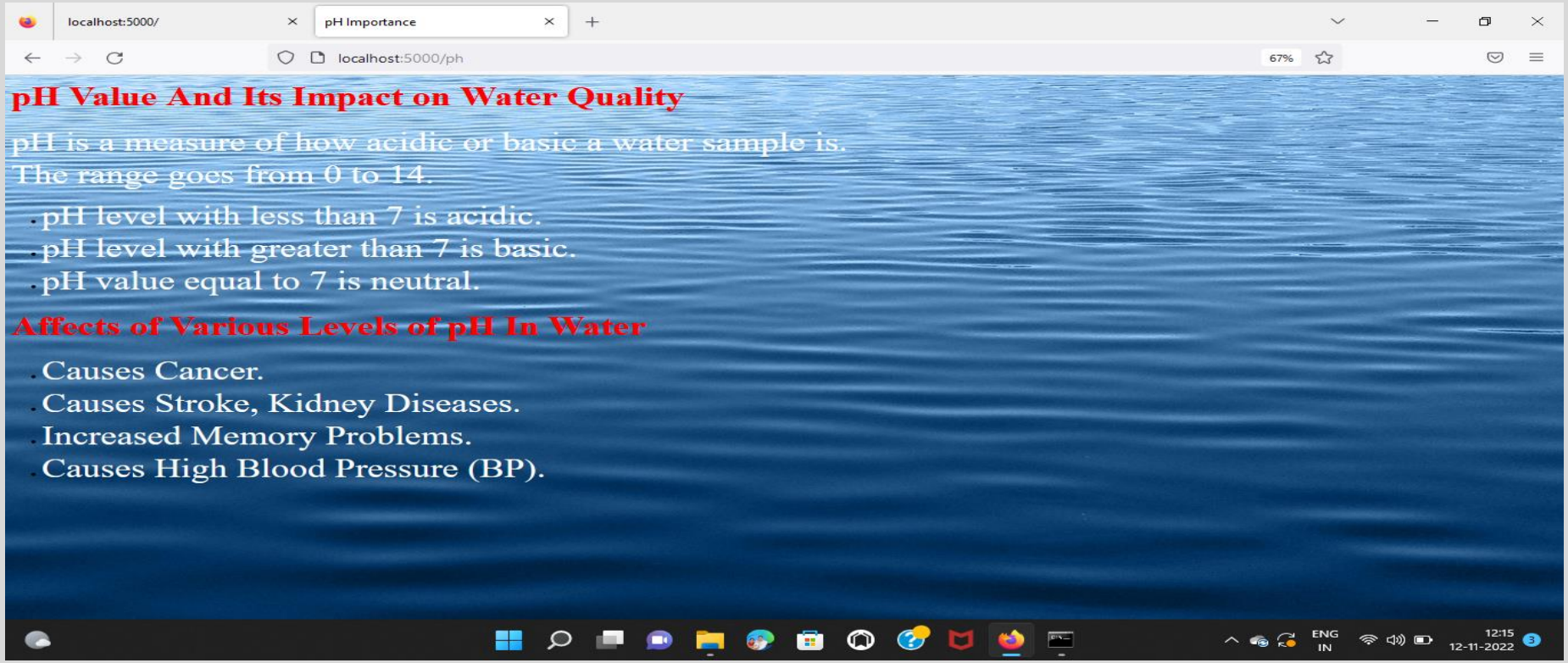


WHY TO FIND WATER QUALITY PAGE

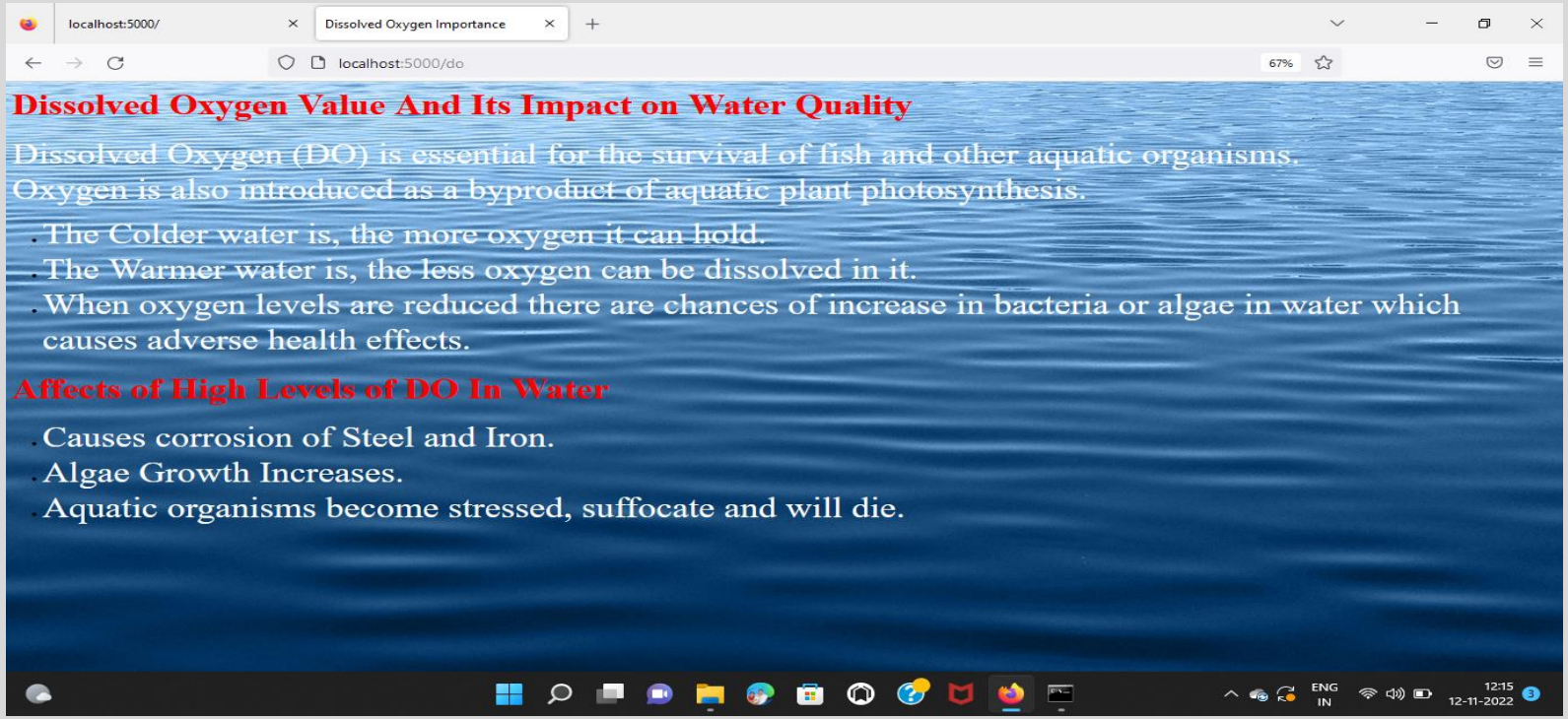


WATER QUALITY DETERMINING PARAMETERS

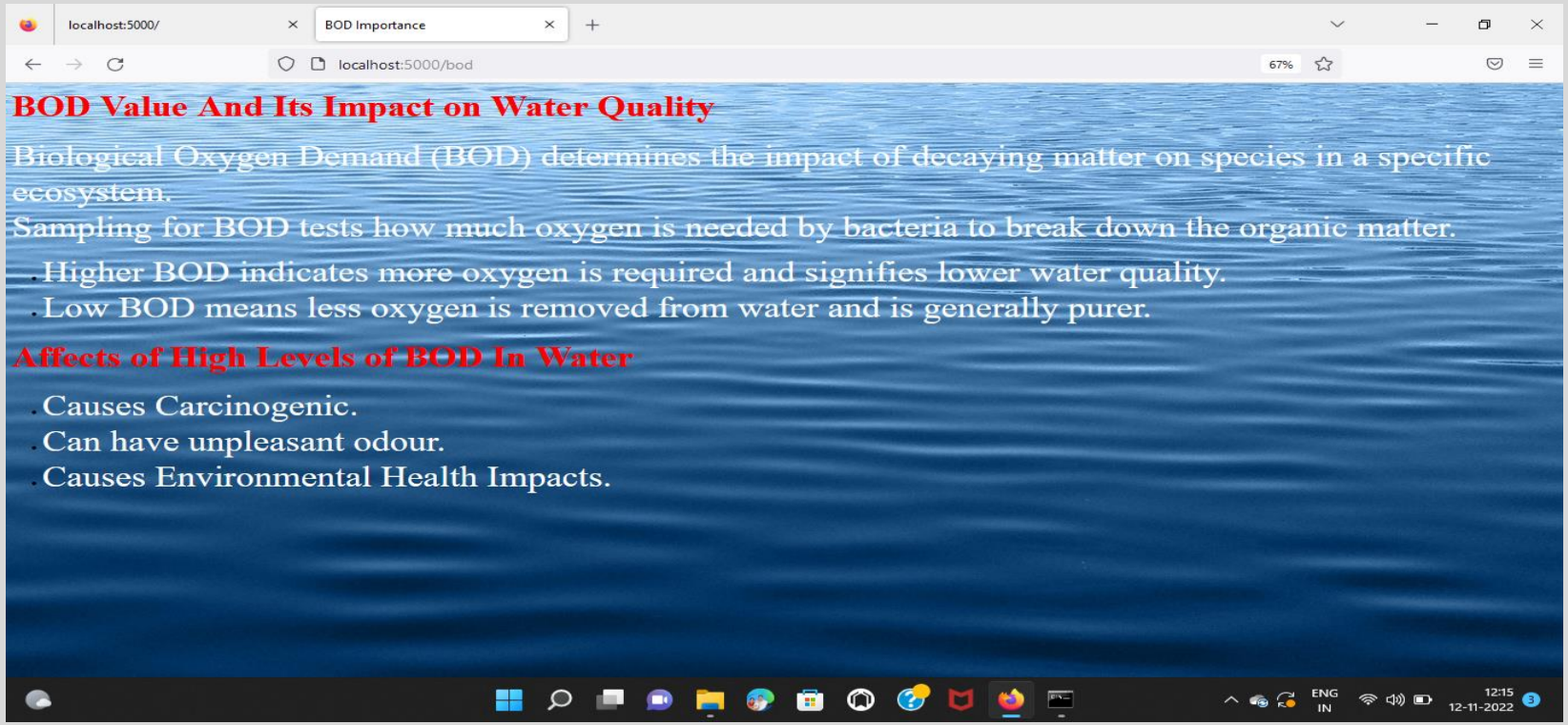
- **pH**



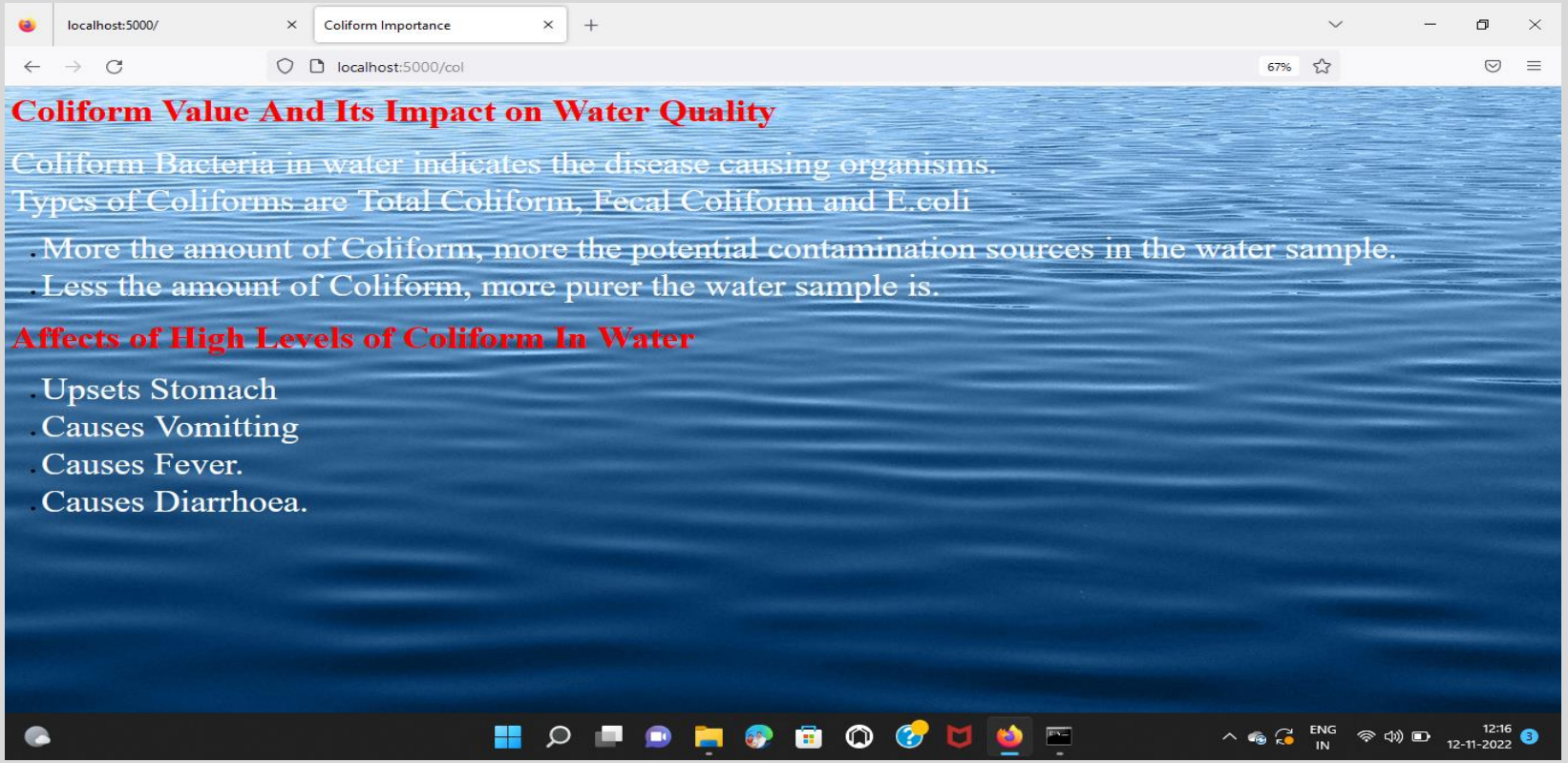
- **Dissolved Oxygen**



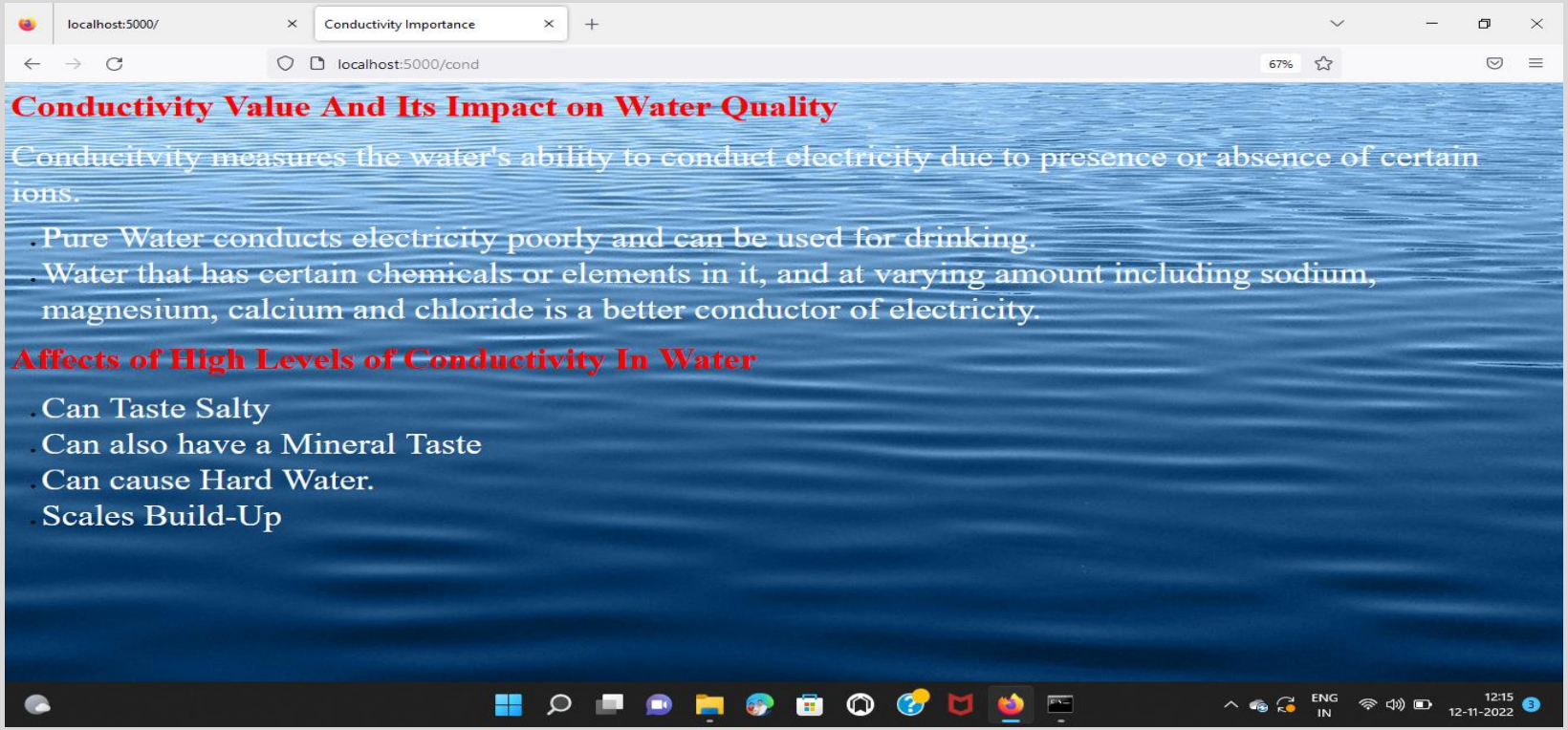
- **Biological Oxygen Demand:**



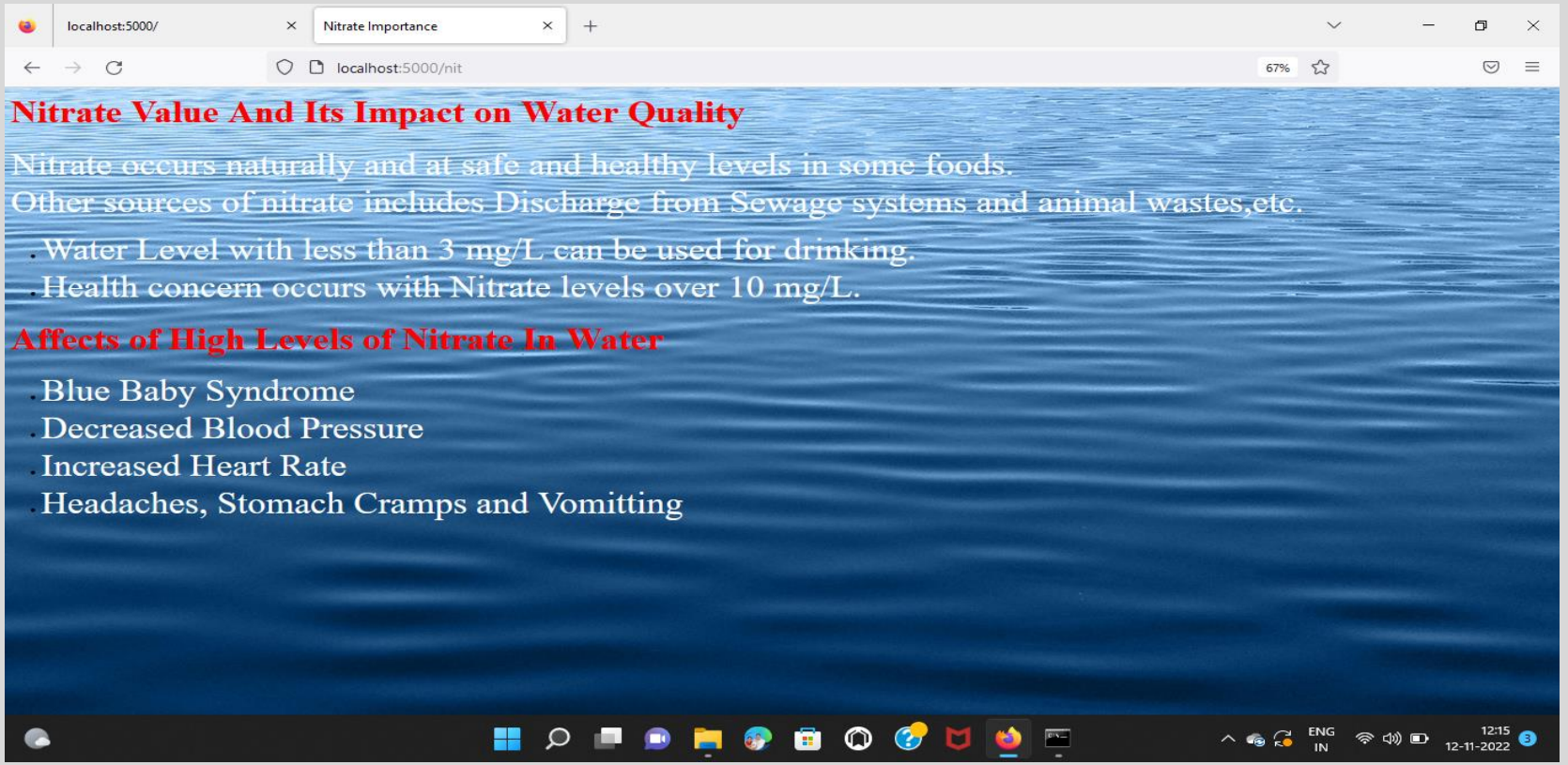
- **Coliform:**


localhost:5000/
Coliform Importance
localhost:5000/col
67%
Coliform Value And Its Impact on Water Quality
Coliform Bacteria in water indicates the disease causing organisms.
Types of Coliforms are Total Coliform, Fecal Coliform and E.coli
.More the amount of Coliform, more the potential contamination sources in the water sample.
.Less the amount of Coliform, more purer the water sample is.
Affects of High Levels of Coliform In Water
.Upsets Stomach
.Causes Vomitting
.Causes Fever.
.Causes Diarrhoea.
ENG
IN
12:16
12-11-2022

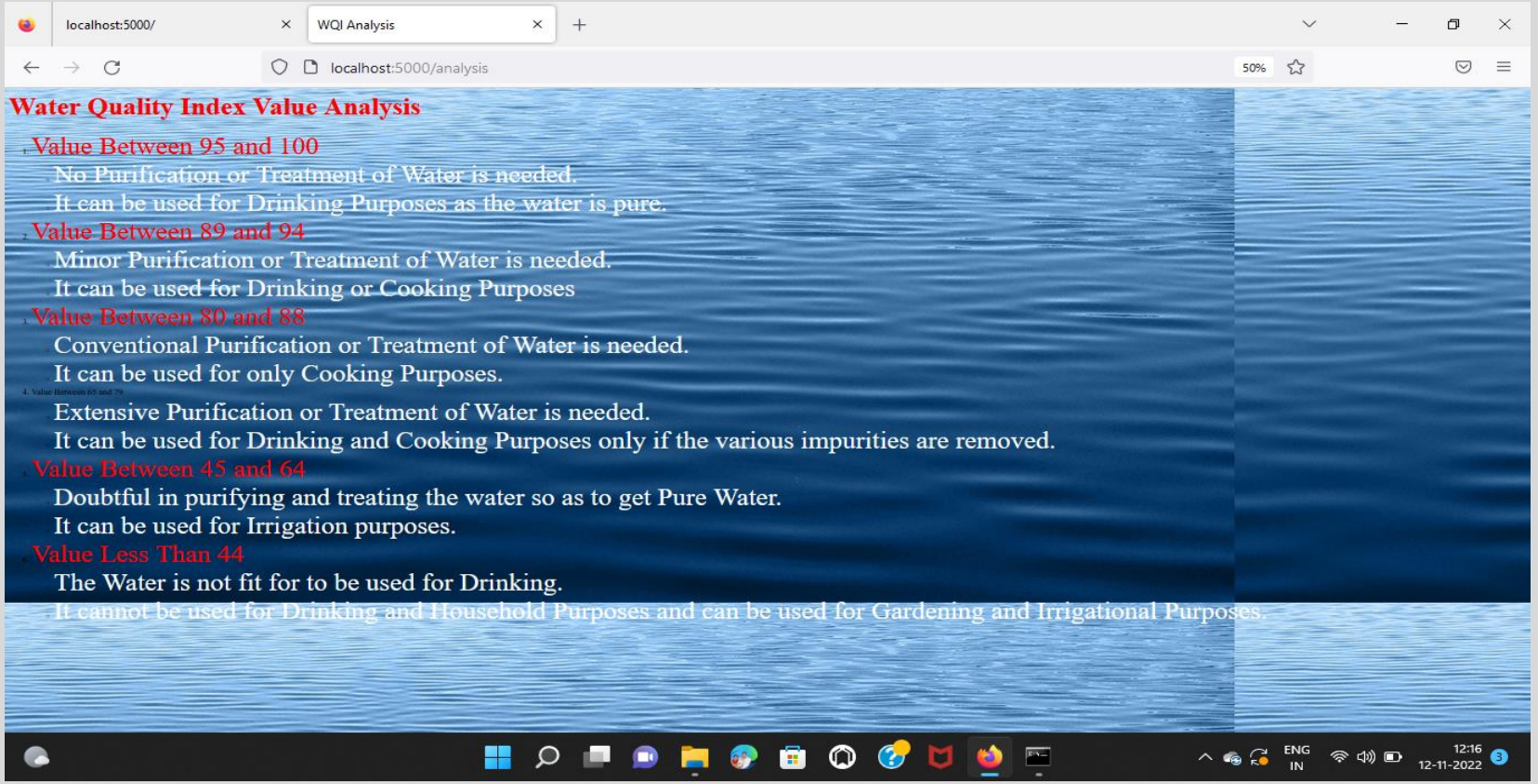
- **Conductivity:**



- **Nitrate:**

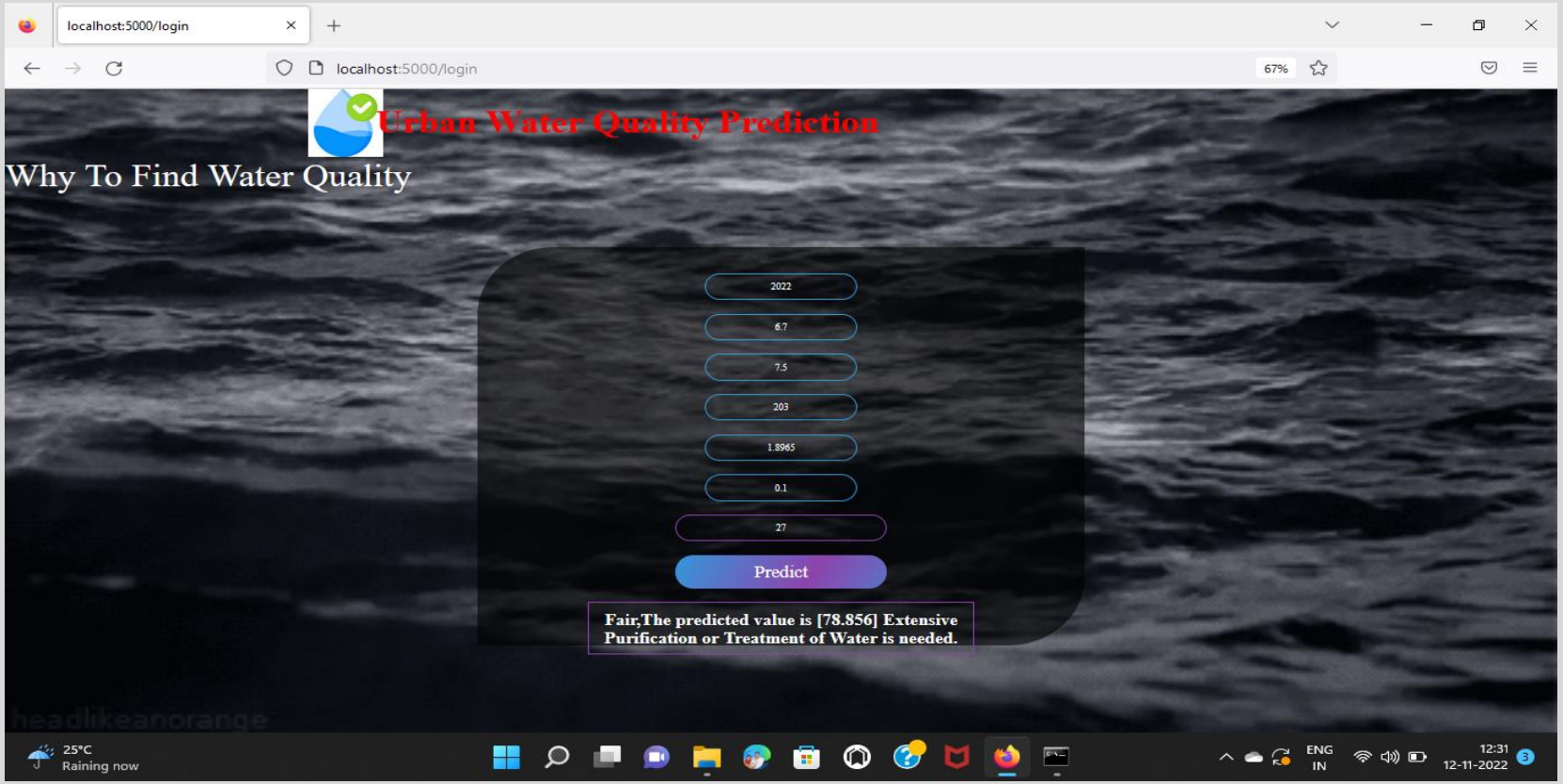

localhost:5000/
Nitrate Importance
localhost:5000/nit
67%
Nitrate Value And Its Impact on Water Quality
Nitrate occurs naturally and at safe and healthy levels in some foods.
Other sources of nitrate includes Discharge from Sewage systems and animal wastes,etc.
.Water Level with less than 3 mg/L can be used for drinking.
.Health concern occurs with Nitrate levels over 10 mg/L.
Affects of High Levels of Nitrate In Water
.Blue Baby Syndrome
.Decreased Blood Pressure
.Increased Heart Rate
.Headaches, Stomach Cramps and Vomitting
ENG
IN
12:15
12-11-2022

WQI INDEX VALUE ANALYSIS TO CREATE AN AWARENESS



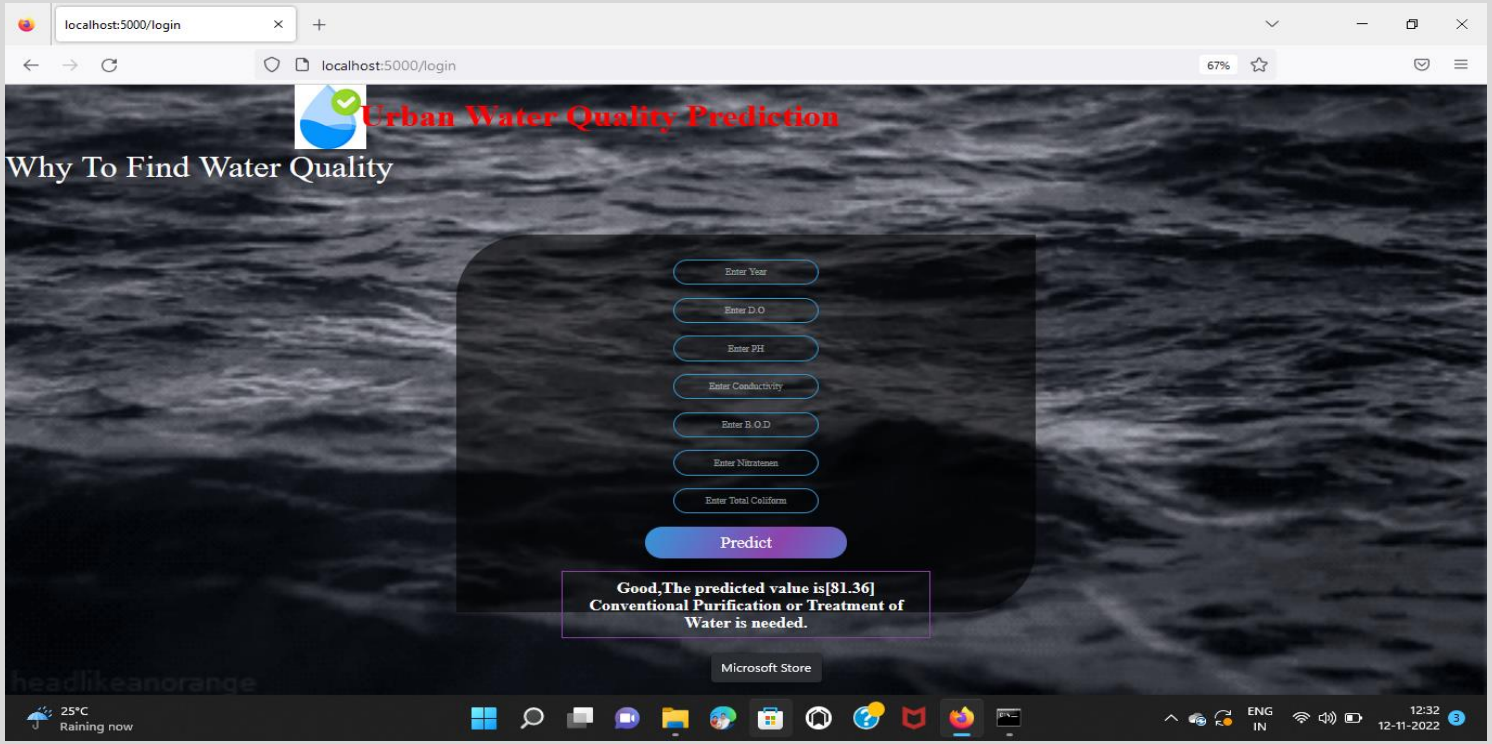
GOOD WATER QUALITY SAMPLE DATA AND PREDICTION

- **INPUT :**



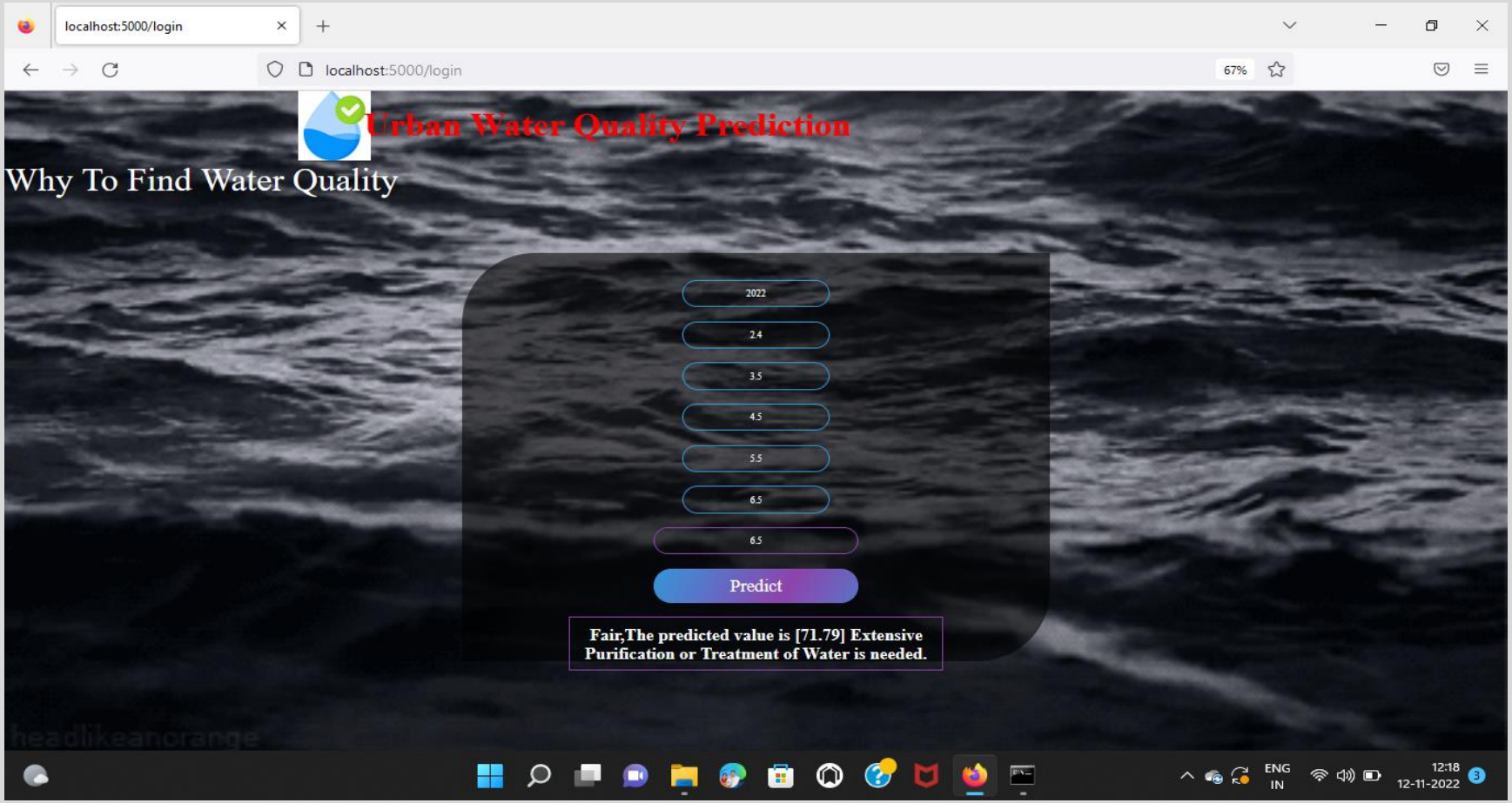
GOOD WATER QUALITY SAMPLE DATA AND PREDICTION

- **PREDICTION:**



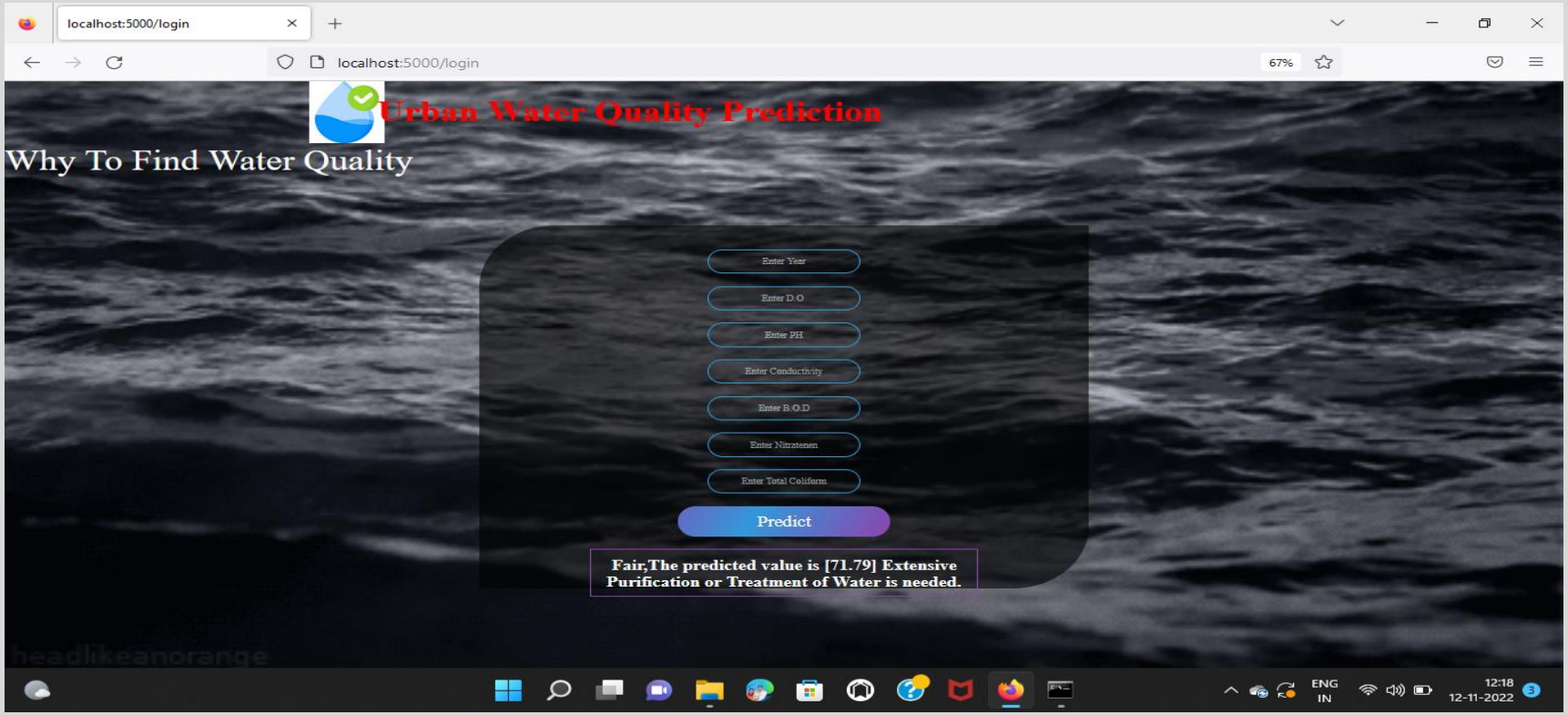
FAIR WATER QUALITY SAMPLE DATA AND PREDICTION

- **INPUT:**



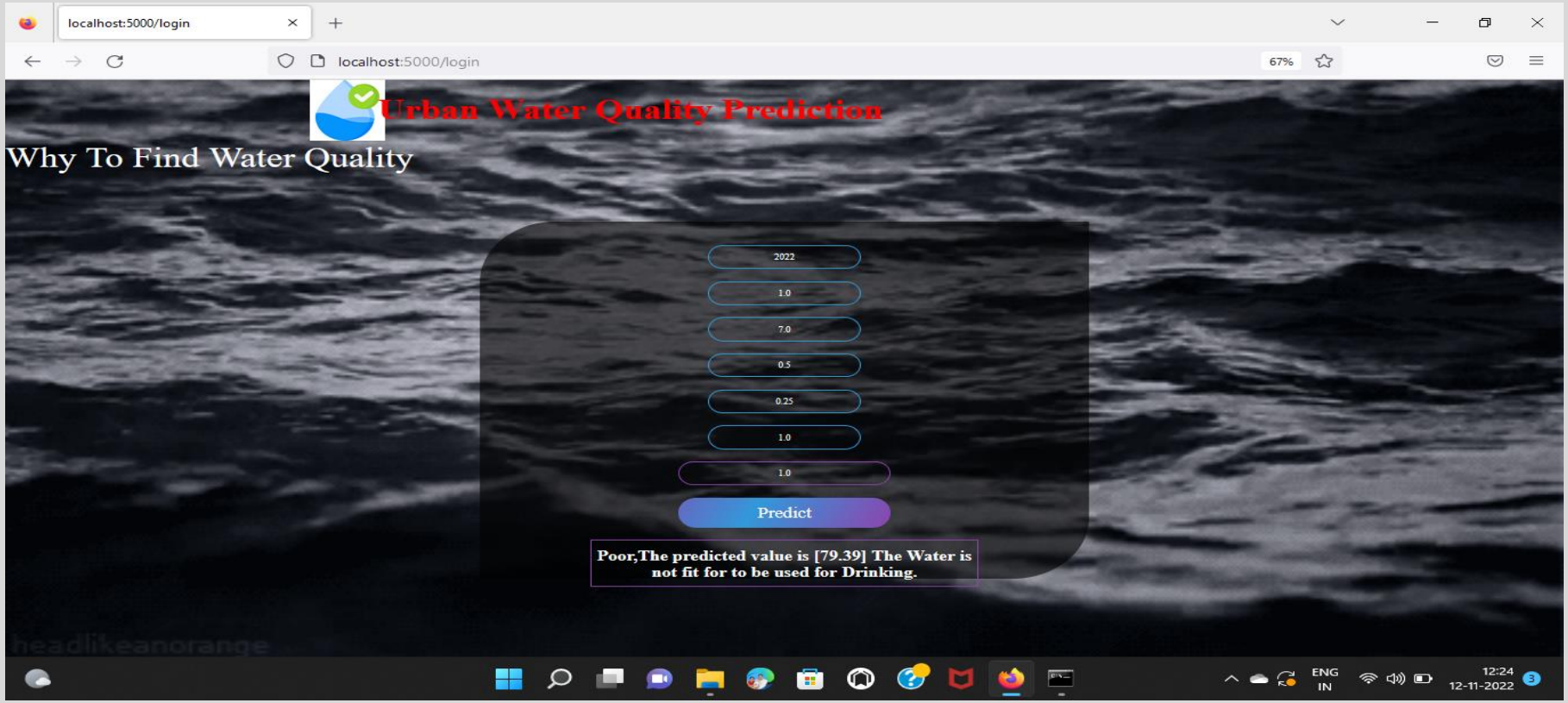
FAIR WATER QUALITY SAMPLE DATA AND PREDICTION

- **PREDICTION:**



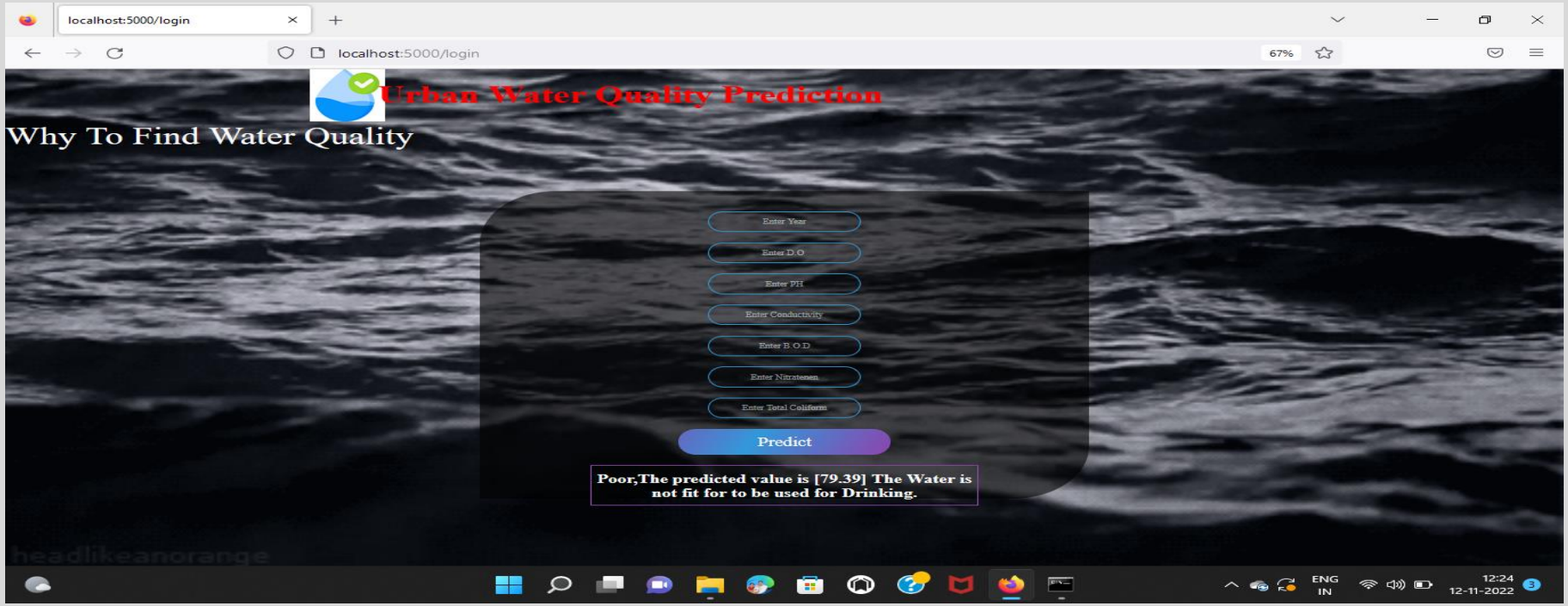
POOR WATER QUALITY SAMPLE DATA AND PREDICTION

- **INPUT:**



POOR WATER QUALITY SAMPLE DATA AND PREDICTION

- **PREDICTION:**



THANK YOU