TEAM ID: PNT2022TMID29328

PROJECT TITLE: Efficient Water Quality Analysis & Prediction using Machine Learning

Project Report

1. INTRODUCTION

1.1 Project Overview

Water is considered as a vital resource that affects various aspects of human health and lives. The quality of water is a major concern for people living in urban areas. The quality of water serves as a powerful environmental determinant and a foundation for the prevention and control of waterborne diseases. However, predicting the urban water quality is a challenging task since the water quality varies in urban spaces non-linearly and depends on multiple factors, such as meteorology, water usage patterns, and land uses.

1.2 Purpose

This project aims at building a Machine Learning (ML) model to Predict Water Quality by considering all water quality standard indicators.Using ML techniques (Regression models) to predict the quality of water instead of using physical measurements or sensors to obtain the quality of water. ML techniques improves the accuracy of measurement over existing chemical and physical techniques as it is infeasible to obtain all the required features to predict the water quality.

2. LITERATURE SURVEY

2.1 Existing problem

The proposed system is intended to determine portability. It is divided into two phases, one for training and the other for testing. The following procedures are carried out in both sections. The data set was chosen as follows: The collection of essential parameters that affect water quality, identification of the number of data samples, and definition of the class labels for each data sample present in the data are all factors that go into selecting the water quality data set, which is a prerequisite to model construction. Ten indicator parameters make up the data sets used in this study. pH value and hardness are examples of these factors. The proposed approach, however, is not constrained by the number of parameters or the selection of parameters. A k-fold cross-validation technique is employed to set the learning and testing framework in this study, corresponding to each data sample in the data set. Using this technique, the dataset is separated into k-disjoint sets of equal size, each with roughly the same class distribution. In turn, this division's subsets are utilized as the test set, with the remaining subsets serving as the training set. These are the Decision Tree (DT) and K-Nearest Neighbour (KNN) methods. Each strategy takes a different approach in terms of the underlying relational structure between the indicator parameters and the class label. As a result, each technique's performance for the same data set is likely to differ. Validating the performance of different classifiers on an unknown data set: Data mining provides several metrics for validating the performance of different classifiers on an unknown data set. A repeated cross-validation procedure in the Matlab caret package created the learning and testing environment. The following procedure was used to apply the classification algorithm:

1. The data set was split into training (80%) and testing (20%). (20 percent).
2. The training set was subjected to repeated cross-validation, with the number of iterations fixed to Classifiers being trained in this manner.
3. The model's optimal parameter configuration was selected, resulting in maximum accuracy.
4. The model was scrutinized.

2.2 References

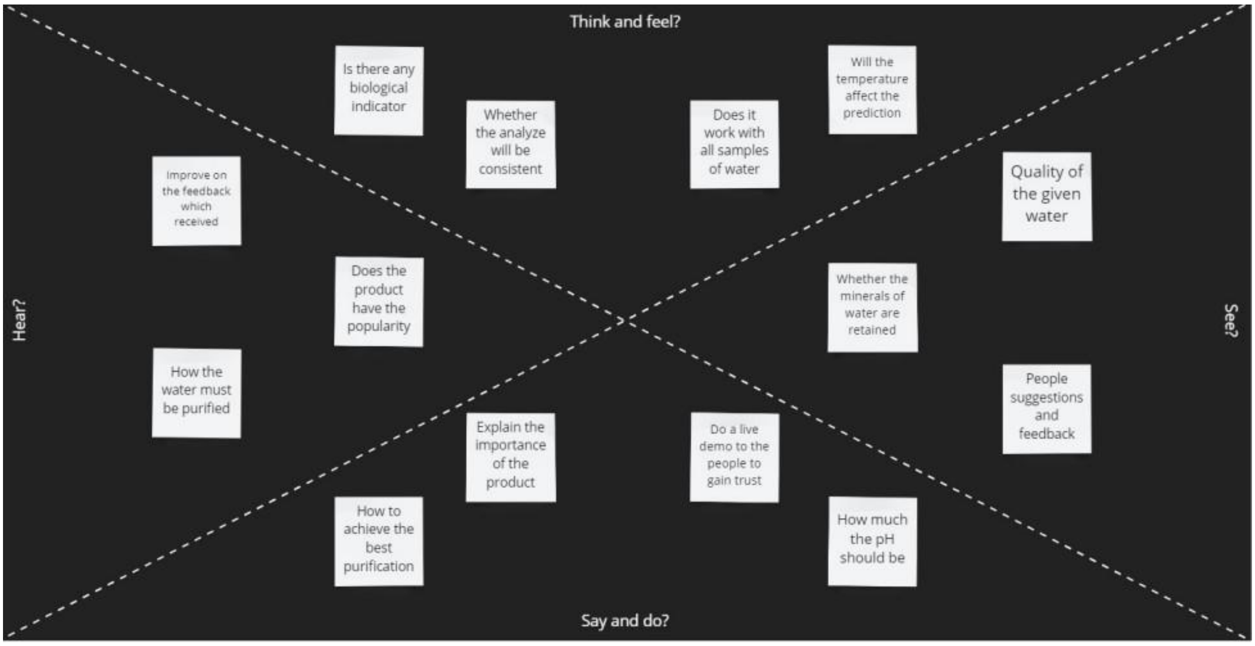
- PCRWR. National Water Quality Monitoring Programme, Fifth Monitoring Report (2005–2006); Pakistan Council of Research in Water Resources Islamabad: Islamabad, Pakistan, 2007.
- Ling, J.K.B. Water Quality Study and Its Relationship with High Tide and Low Tide at Kuantan River. Bachelor’s Thesis, Universiti Malaysia Pahang, Gambang, Malaysia, 2010.

2.3 Problem Statement Definition

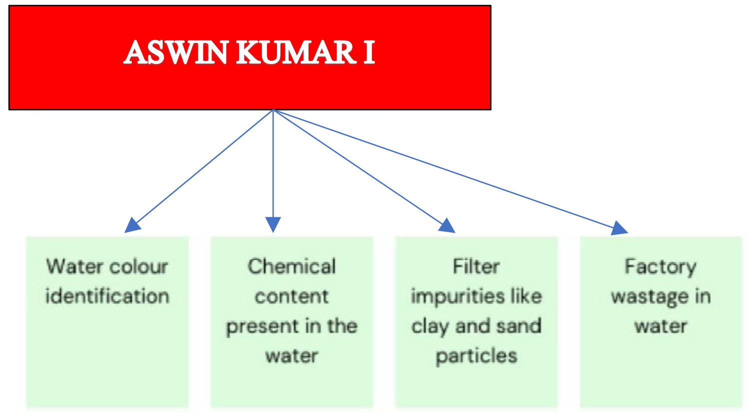
The main aim of the project is to predict the quality of the water. We are building a web app to predict the quality of the water. Project aims at building a Machine Learning (ML) model to Predict Water Quality by considering all water quality standard indicators. WQI is fundamentally calculated by initially multiplying the q value of each parameter by its corresponding weight, adding them all up and then dividing the result by the sum of weights of the employed parameters

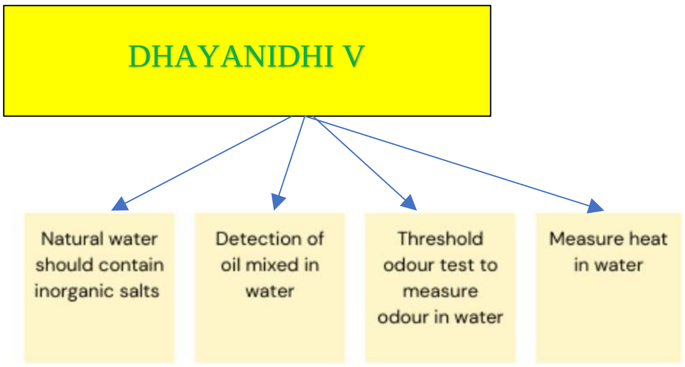
3. IDEATION & PROPOSED SOLUTION

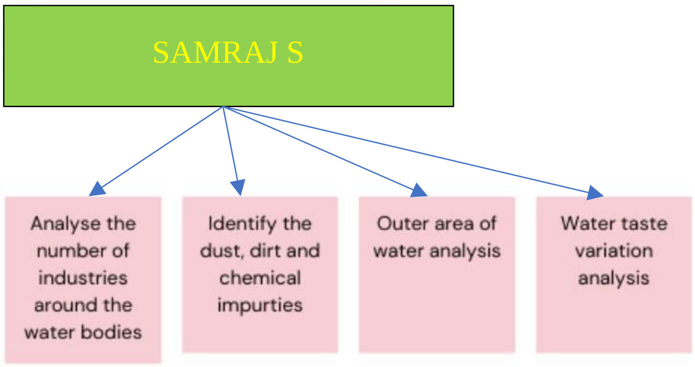
3.1 Empathy Map Canvas

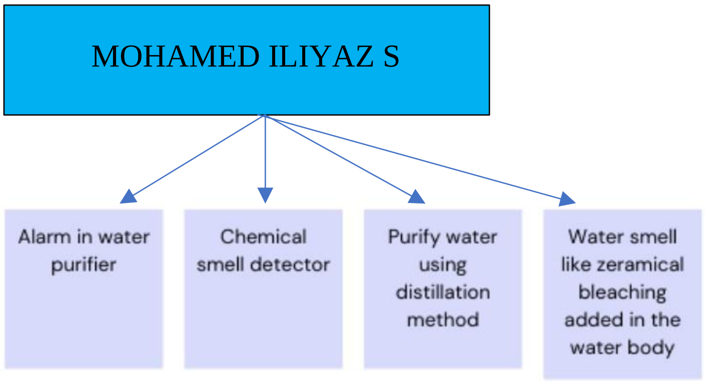

Think and feel?
Is there any biological indicator
Whether the analyze will be consistent
Does it work with all samples of water
Will the temperature affect the prediction
Improve on the feedback which received
Quality of the given water
Does the product have the popularity
Whether the minerals of water are retained
Hear?
See?
How the water must be purified
People suggestions and feedback
Explain the importance of the product
Do a live demo to the people to gain trust
How to achieve the best purification
How much the pH should be
Say and do?

3.2 Ideation & Brainstorming


ASWIN KUMAR I
Water colour identification
Chemical content present in the water
Filter impurities like clay and sand particles
Factory wastage in water


DHAYANIDHI V
Natural water should contain inorganic salts
Detection of oil mixed in water
Threshold odour test to measure odour in water
Measure heat in water


SAMRAJ S
Analyse the number of industries around the water bodies
Identify the dust, dirt and chemical impurties
Outer area of water analysis
Water taste variation analysis


MOHAMED ILIYAZ S
Alarm in water purifier
Chemical smell detector
Purify water using distillation method
Water smell like zeramical bleaching added in the water body

3.3 Proposed Solution

S.No	Parameter	Description
1.	Problem Statement	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.
2.	Idea / Solution description	The solution is derived from the data sets by comparing the accuracy rate with the previous data set and the current data set.
3.	Novelty / Uniqueness	Using ML techniques (Regression models) to predict the quality of water instead of using physical measurements or sensors to obtain the quality of water. ML techniques improves the accuracy of measurement over existing chemical and physical techniques as it is infeasible to obtain all the required features to predict the water quality. Physical and chemical measurements may lead to the usage of expensive instruments and also take a lot of time. ML techniques make the process easier, feasible and faster.
4.	Social Impact /Customer Satisfaction	Our intended audience consists of people who are concerned about the quality of water they drink. Water's health is more important which should be considered as many water-borne diseases are more widely known. The proposed solution will help in identifying water pollution and helps the customer to drink healthy water.
5.	Business Model (Revenue Model)	Industries that provide sanitation facilities and products (like water purifiers, quality testers etc.) can deploy this solution to provide more waste water treatment plants, better insights in health concerns and there may also be an increase in awareness and demand for better water quality testing and

		availability. People will start looking for treatments related to water-borne diseases as the awareness increases
6.	Scalability of the Solution	The solution proposed will be deployed as a web application. So, it is easily accessible by anyone who has internet services and has no specific software and hardware specifications

3.4 Problem Solution fit:

The Problem-Solution Fit simply means that you have found a problem with your customer and that the solution you have realized for it actually solves the customer's problem. It helps entrepreneurs, marketers and corporate innovators identify behavioral patterns.

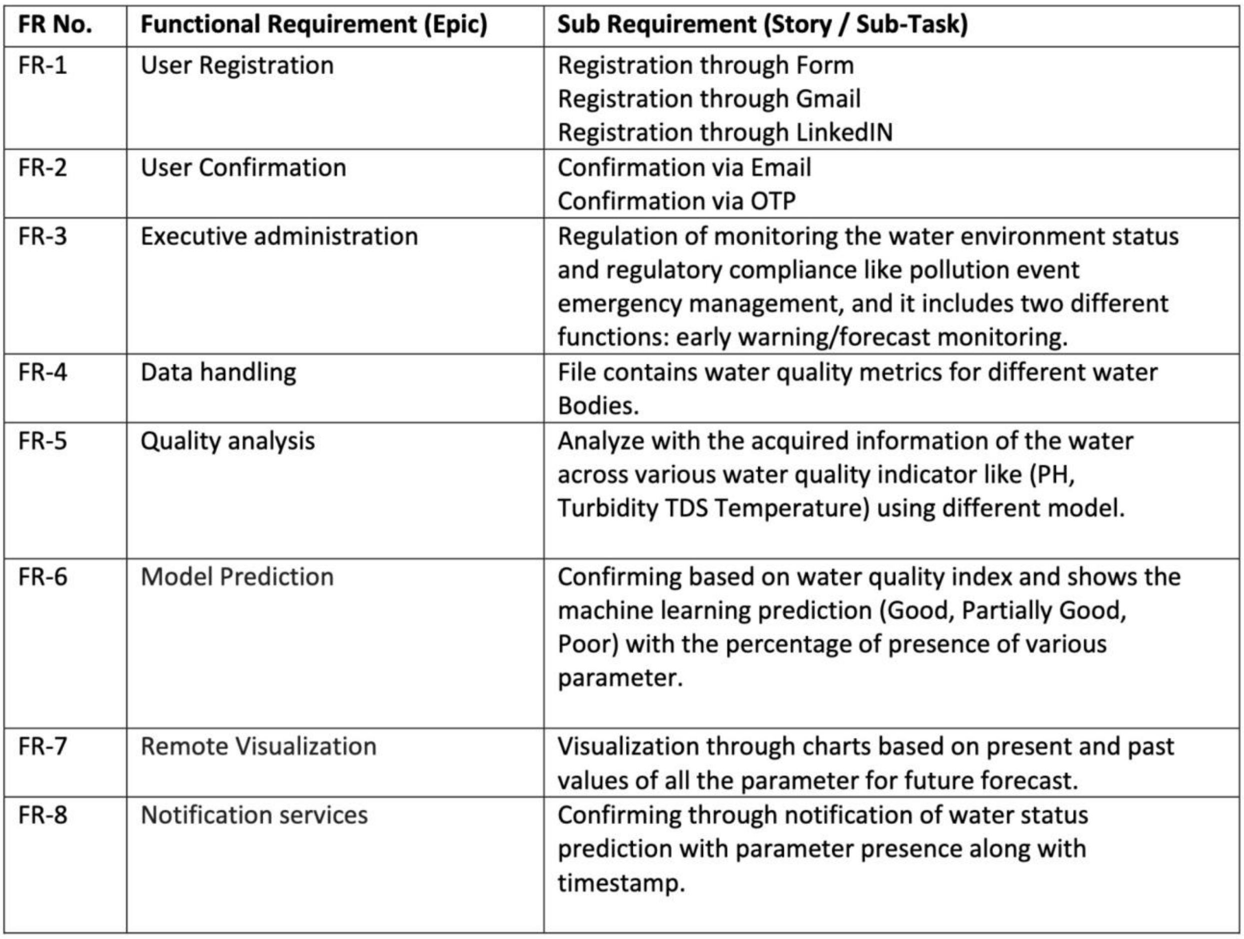
Purpose:

- Customer needs to know about water's parameters such as pH, nitrate content so that it can be given to the ML model to predict the quality of water.
- User uses various experimental techniques like analyzing the quantity of chemical present and also analyzes physical properties of the water.
- Solve complex problems in a way that fits the state of your customers.
- Succeed faster and increase your solution adoption by tapping into existing mediums and channels of behavior.
- Sharpen your communication and marketing strategy with the right triggers and messaging.

4. REQUIREMENT ANALYSIS

4.1 Functional requirement

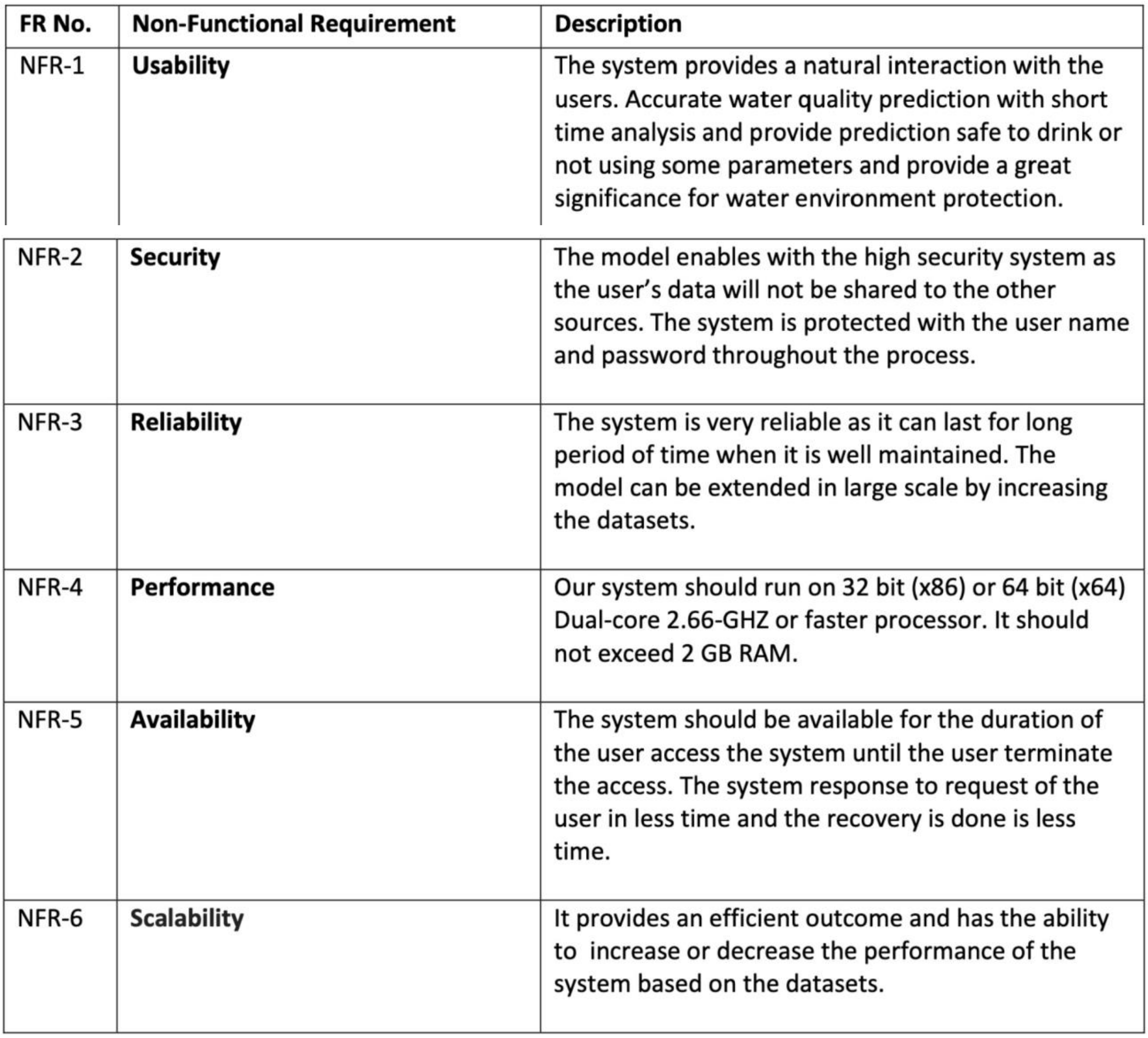
Following are the functional requirements of the proposed solution



FR No.	Functional Requirement (Epic)	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 Confirmation via OTP
FR-3	Executive administration	Regulation of monitoring the water environment status and regulatory compliance like pollution event emergency management, and it includes two different functions: early warning/forecast monitoring.
FR-4	Data handling	File contains water quality metrics for different water Bodies.
FR-5	Quality analysis	Analyze with the acquired information of the water across various water quality indicator like (PH, Turbidity TDS Temperature) using different model.
FR-6	Model Prediction	Confirming based on water quality index and shows the machine learning prediction (Good, Partially Good, Poor) with the percentage of presence of various parameter.
FR-7	Remote Visualization	Visualization through charts based on present and past values of all the parameter for future forecast.
FR-8	Notification services	Confirming through notification of water status prediction with parameter presence along with timestamp.

4.2 Non-Functional requirements

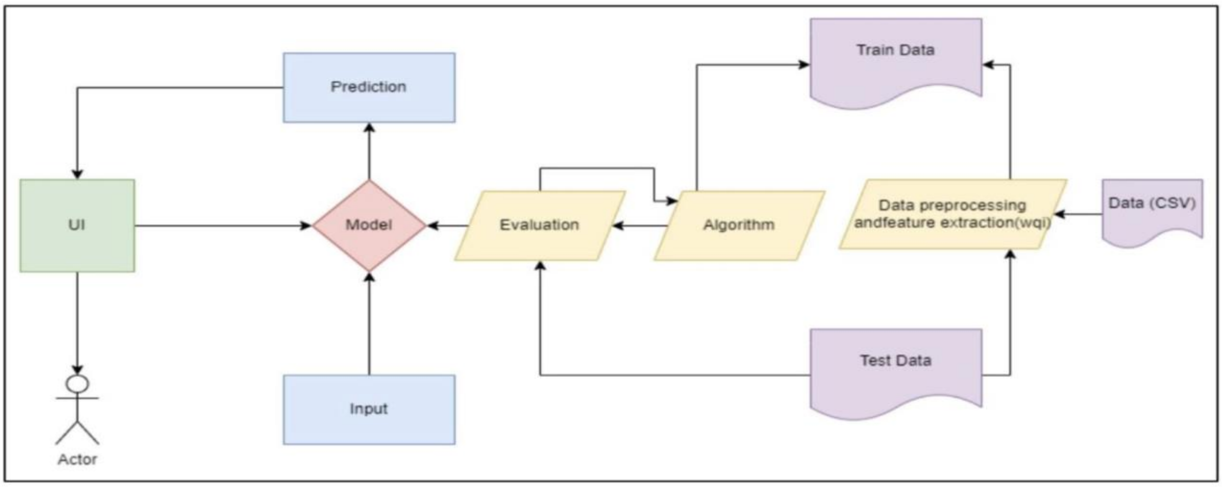
Following are the non-functional requirements of the proposed solution.



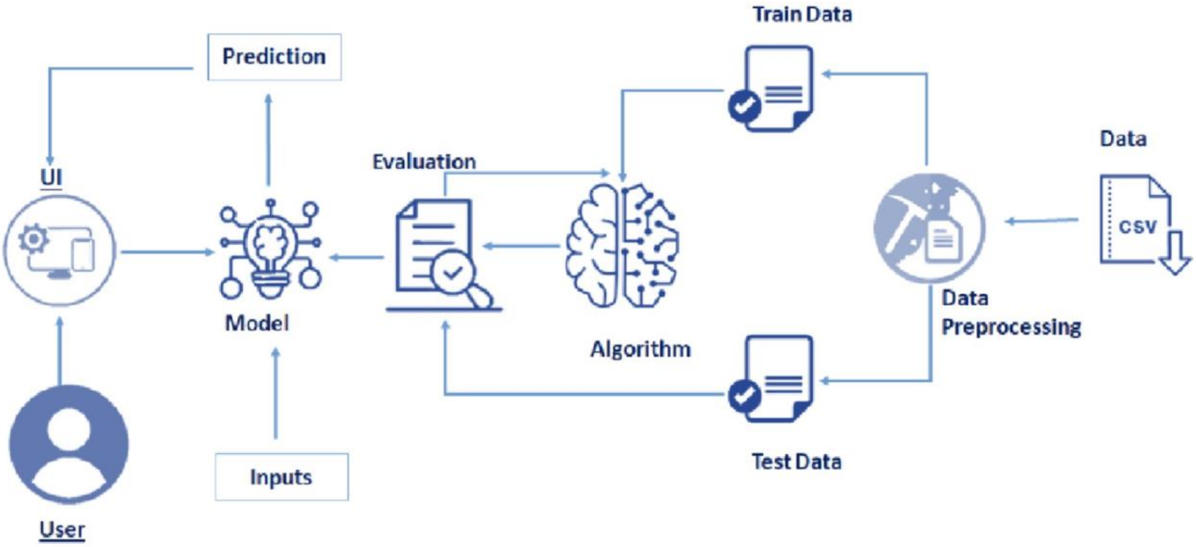
FR No.	Non-Functional Requirement	Description
NFR-1	**Usability**	The system provides a natural interaction with the users. Accurate water quality prediction with short time analysis and provide prediction safe to drink or not using some parameters and provide a great significance for water environment protection.
NFR-2	**Security**	The model enables with the high security system as the user’s data will not be shared to the other sources. The system is protected with the user name and password throughout the process.
NFR-3	**Reliability**	The system is very reliable as it can last for long period of time when it is well maintained. The model can be extended in large scale by increasing the datasets.
NFR-4	**Performance**	Our system should run on 32 bit (x86) or 64 bit (x64) Dual-core 2.66-GHZ or faster processor. It should not exceed 2 GB RAM.
NFR-5	**Availability**	The system should be available for the duration of the user access the system until the user terminate the access. The system response to request of the user in less time and the recovery is done is less time.
NFR-6	**Scalability**	It provides an efficient outcome and has the ability to increase or decrease the performance of the system based on the datasets.

5. PROJECT DESIGN

5.1 Data Flow Diagrams



5.2 Solution & Technical Architecture


Train Data
Prediction
UI
Evaluation
Data
CSV
Model
Algorithm
Data Preprocessing
Test Data
Inputs
User