Predicting And Analyzing Urban Water Quality Using IBM Watson Machine Learning Service

1. INTRODUCTION

1.1 Overview

Aim of the project is to predict the quality of the Urban Water according to the sample data given.

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. 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.

1.2 Purpose

Using IBM Watson Studio we train the data set using Random Forest Regression algorithm that help to train the model with the help of machine learning services provided by the IBM. Using the dataset which have the existing sample data of the quality determining experimented values, Machine learn and study the variation according to the values of year, ph content, dissolved oxygen(DO),conductivity, biological oxygen demand(BOD),Nitrate and Total Coliform. So according to

these factors urban water quality can be predicted machine will learn about it using the algorithm, The predicted water quality can be used to understand the purity of the water and it can be used drinking purpose according to the past data.

2. LITERATURE SURVEY

2.1 Existing problem

Using modern chemistry, we can detect thousands of chemicals in water, even at extremely low concentrations. The ever-growing list of tests that are available can feel overwhelming, and the vast majority of methods require state-of-the art lab facilities. Fortunately, we don't need to test for everything, a much smaller and more practical set of tests can provide a good sense of chemical water quality for monitoring purposes. There are different types of test to detect the each chemical content in the water by getting each test values we note each values and calculate the water quality index using the formulas therefore this take times to get the result and the cost is high and also faces Technical Hurdles issues, manual process issue, transportation issue, lack of technologies.

2.2 Proposed solution

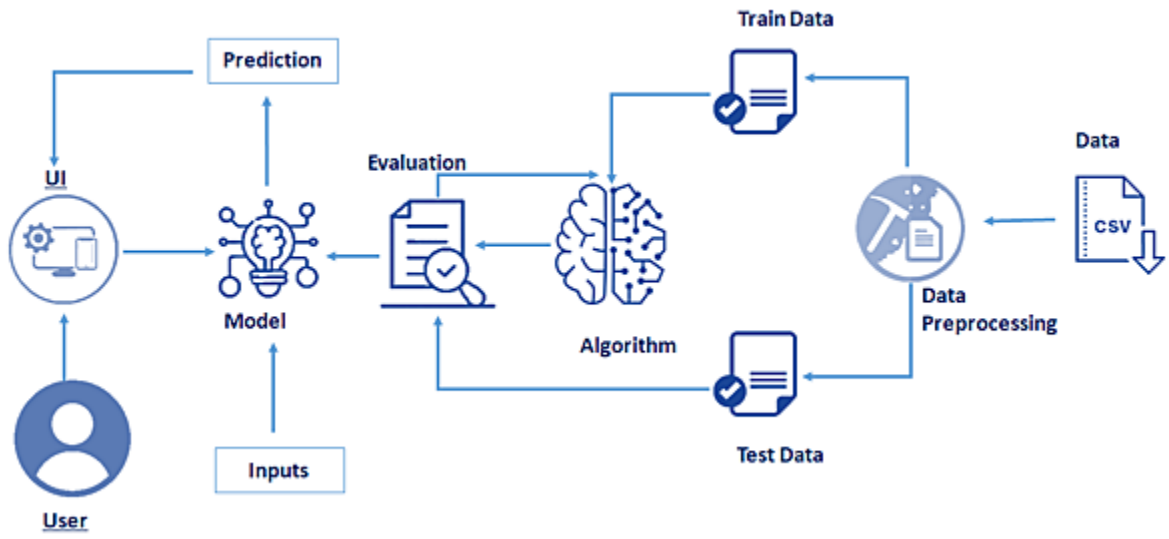
To eliminate all these difficulties and problems we proposed a system that predict the quality using Machine Learning services that we train the data from the past data that helps to predict the result very fast and accurate by using the random forest regression algorithm

Water is a valuable resource that has an impact on many elements of human health and life. For those who live in cities, water quality is a serious concern. Water quality is an important environmental variable as

well as a foundation for preventing and controlling waterborne illnesses. However, because water quality varies non-linearly in urban spaces and is dependent on multiple factors such as meteorology, water usage patterns, and land uses, predicting urban water quality is a difficult task. As a result, this project aims to develop a Machine Learning (ML) model to predict water quality by taking into account all water quality standard indicators.

3. THEORITICAL ANALYSIS

3.1 Block diagram



3.2 Hardware / Software designing

IBM Watson Studio - IBM Watson Studio helps data scientists and analysts prepare data and build models at scale across any cloud.

IBM Watson Machine Learning - IBM Watson Machine Learning helps data scientists and developers accelerate AI and

machine-learning deployment.
IBM Cloud Object Storage - IBM Cloud Object Storage makes it possible to store practically limitless amounts of data, simply and cost effectively.
Machine Learning Services - Machine learning as service is an umbrella term for collection of various cloud-based platforms that use machine learning tools to provide solutions that can help ML teams with: out-of-the box predictive analysis for various use cases, data pre-processing, model training and tuning.

4. EXPERIMENTAL INVESTIGATIONS

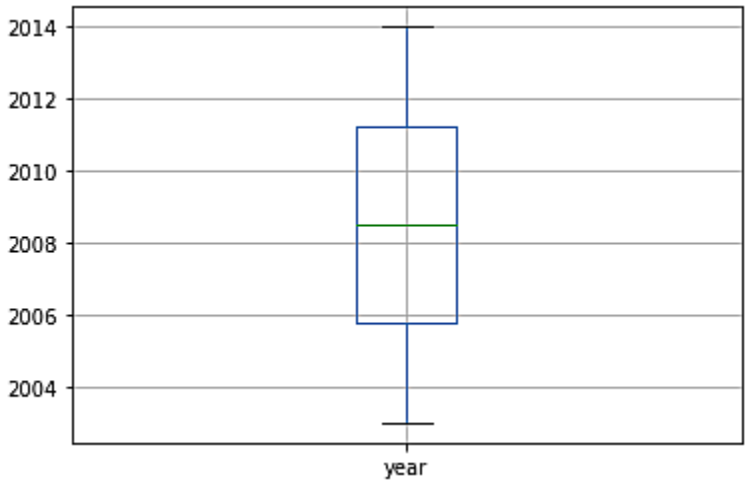
Here we are going to build a machine learning model that predicts the Urban Water Quality based on the following parameters

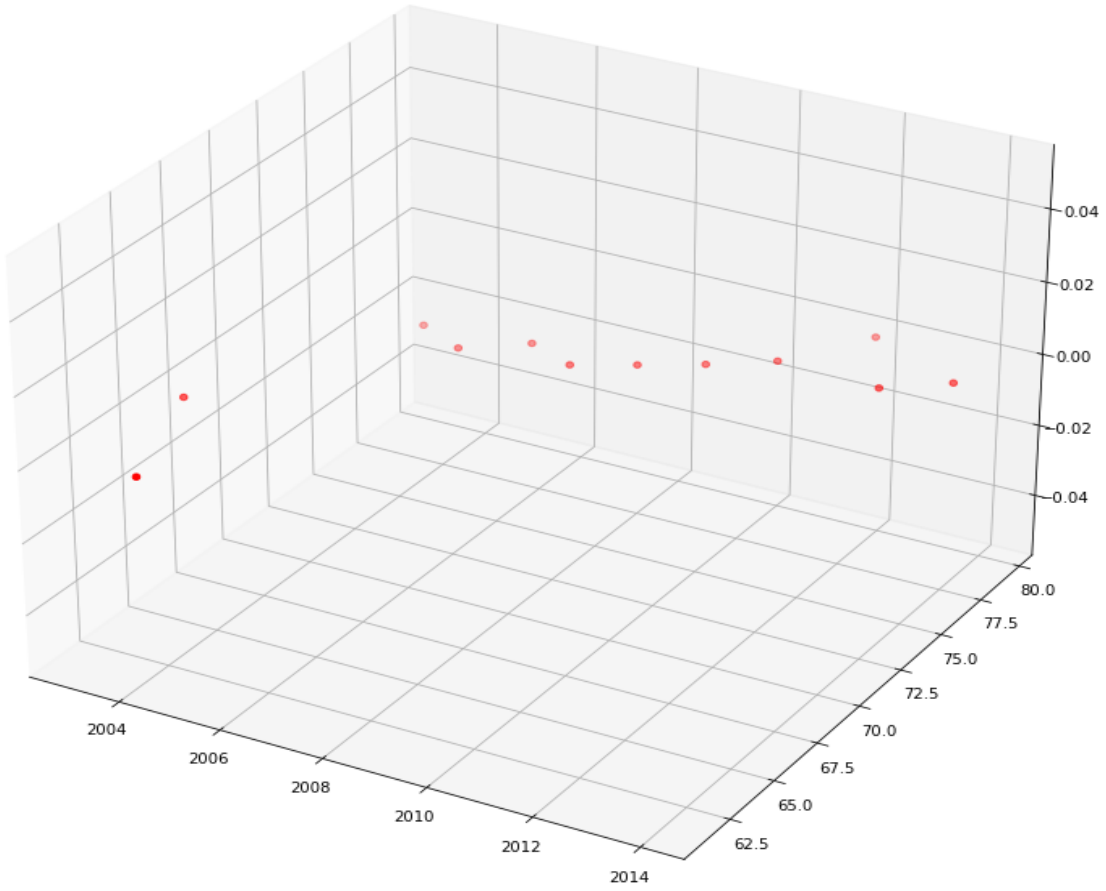
- Year
- PH
- DO (Dissolved Oxygen)
- BOD (Biological Oxygen Demand)
- Nitrate
- Conductivity
- Total Coliform

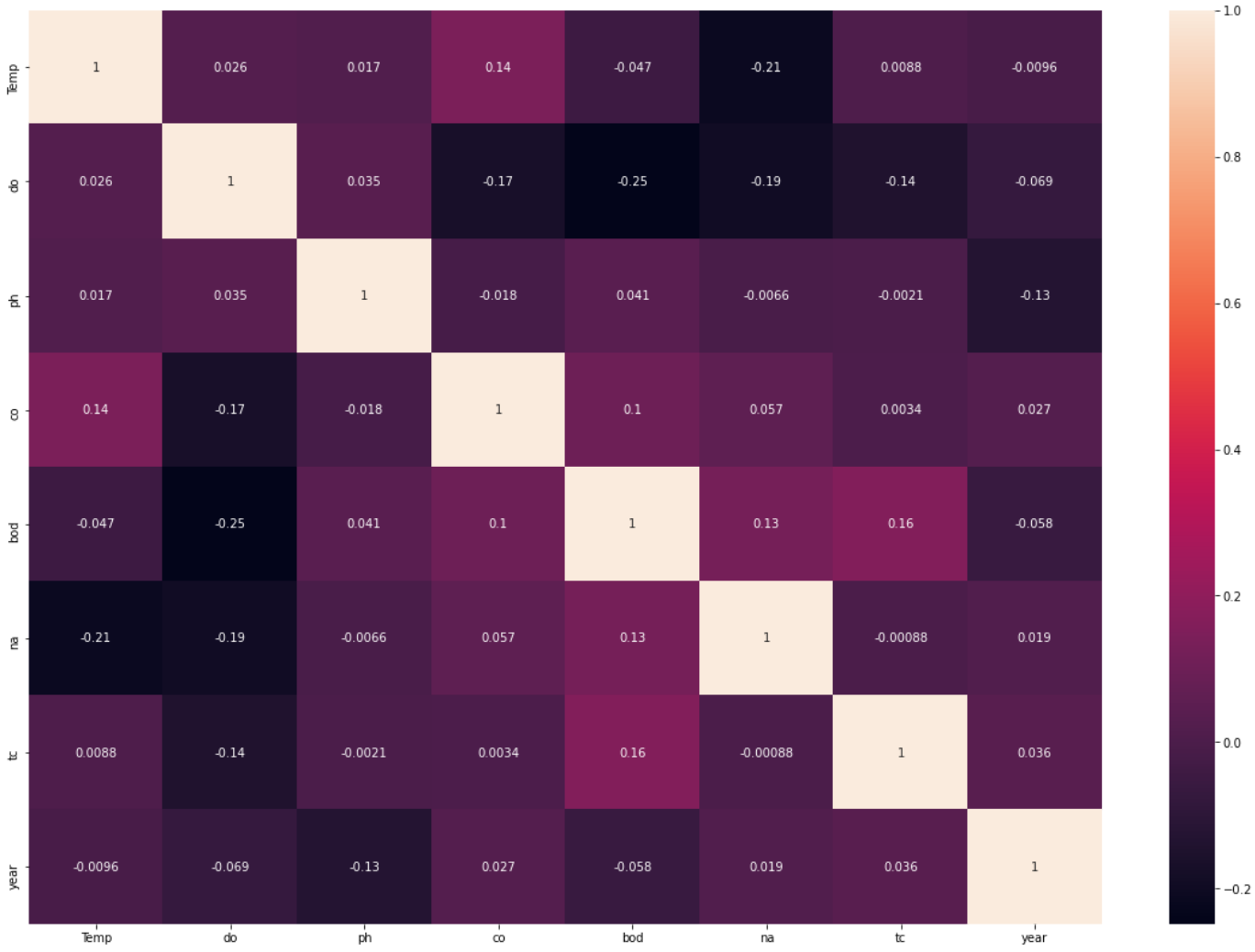
Here there are 7 parameters which is used to detect the quality of the urban water. They are year, ph content, DO, BOD, Nitrate, conductivity and total coliform. All these values are the factors which predict the quality of the water.

4.1 EXPERIMENTAL ANALYSIS

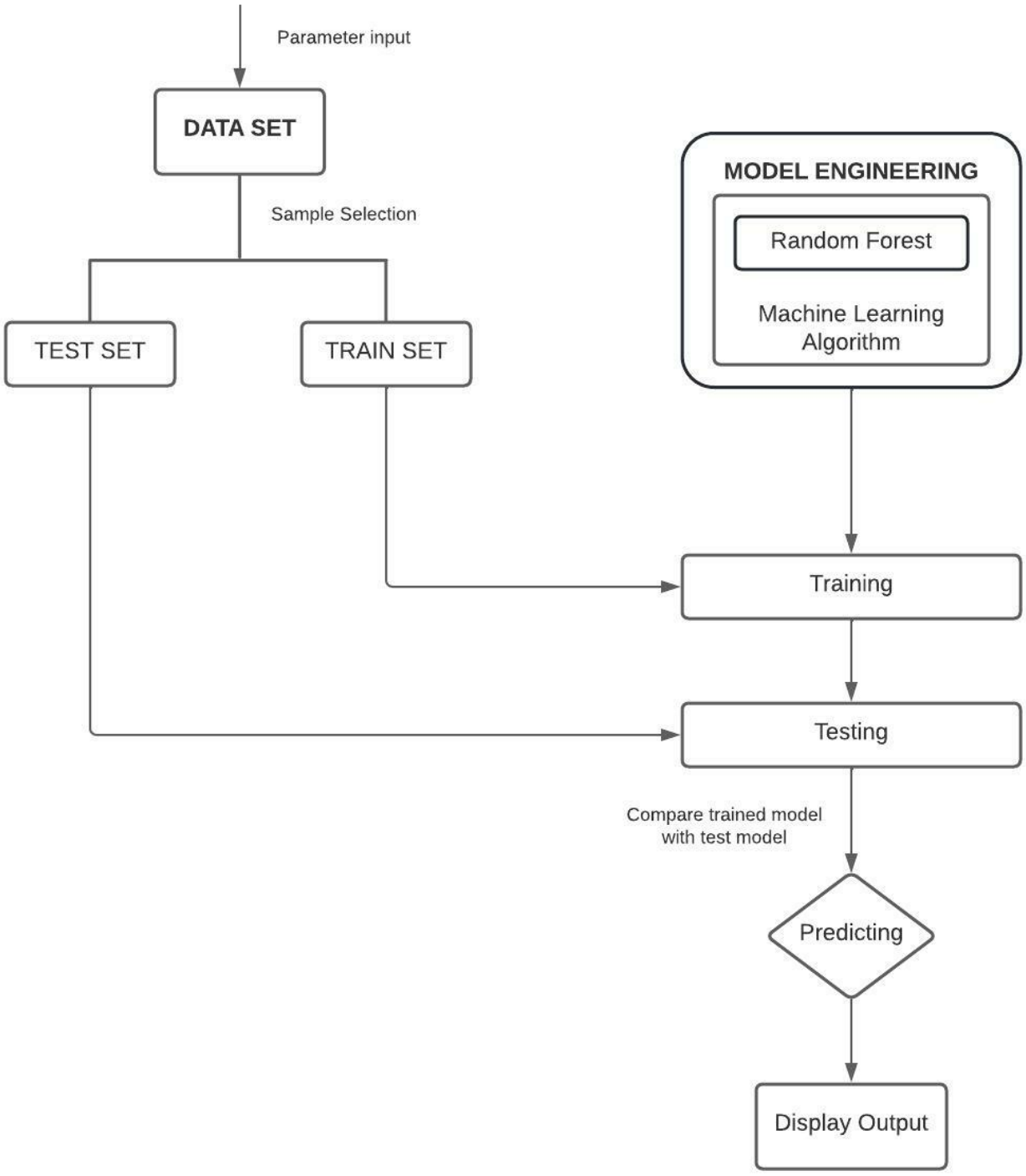
- Visualization



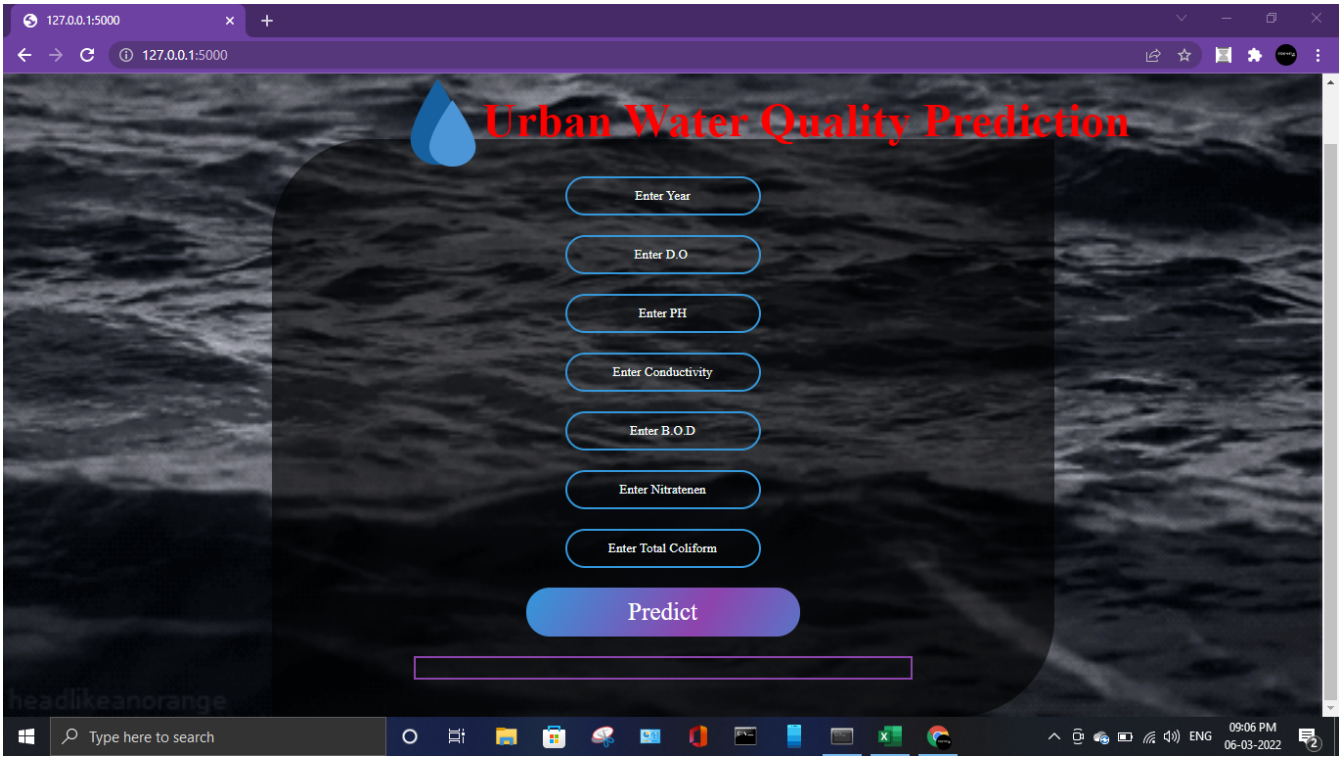



Temp
do
ph
co
bod
na
tc
year
1
0.026
0.017
0.14
-0.047
-0.21
0.0088
-0.0096
0.026
1
0.035
-0.17
-0.25
-0.19
-0.14
-0.069
0.017
0.035
1
-0.018
0.041
-0.0066
-0.0021
-0.13
0.14
-0.17
-0.018
1
0.1
0.057
0.0034
0.027
-0.047
-0.25
0.041
0.1
1
0.13
0.16
-0.058
-0.21
-0.19
-0.0066
0.057
0.13
1
-0.00088
0.019
0.0088
-0.14
-0.0021
0.0034
0.16
-0.00088
1
0.036
-0.0096
-0.069
-0.13
0.027
-0.058
0.019
0.036
1
1.0
0.8
0.6
0.4
0.2
0.0
−0.2

5. FLOWCHART


Parameter input
DATA SET
Sample Selection
TEST SET
TRAIN SET
MODEL ENGINEERING
Random Forest
Machine Learning Algorithm
Training
Testing
Compare trained model with test model
Predicting
Display Output

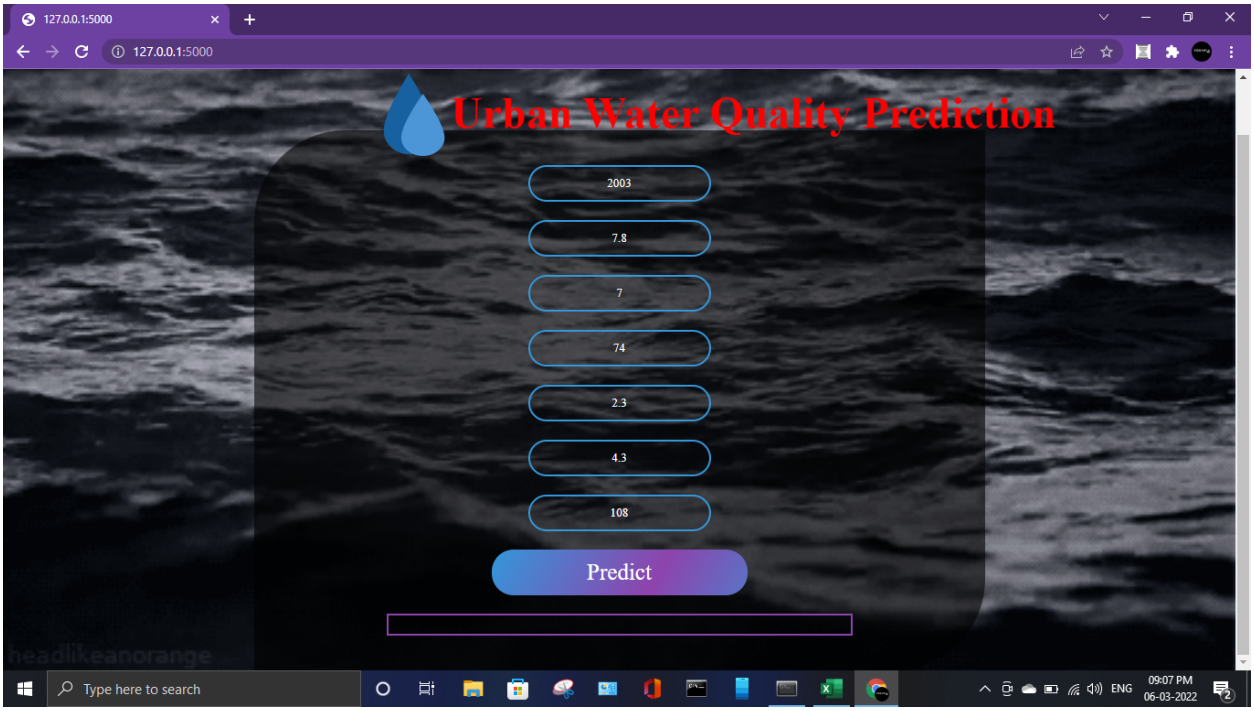
6. RESULT

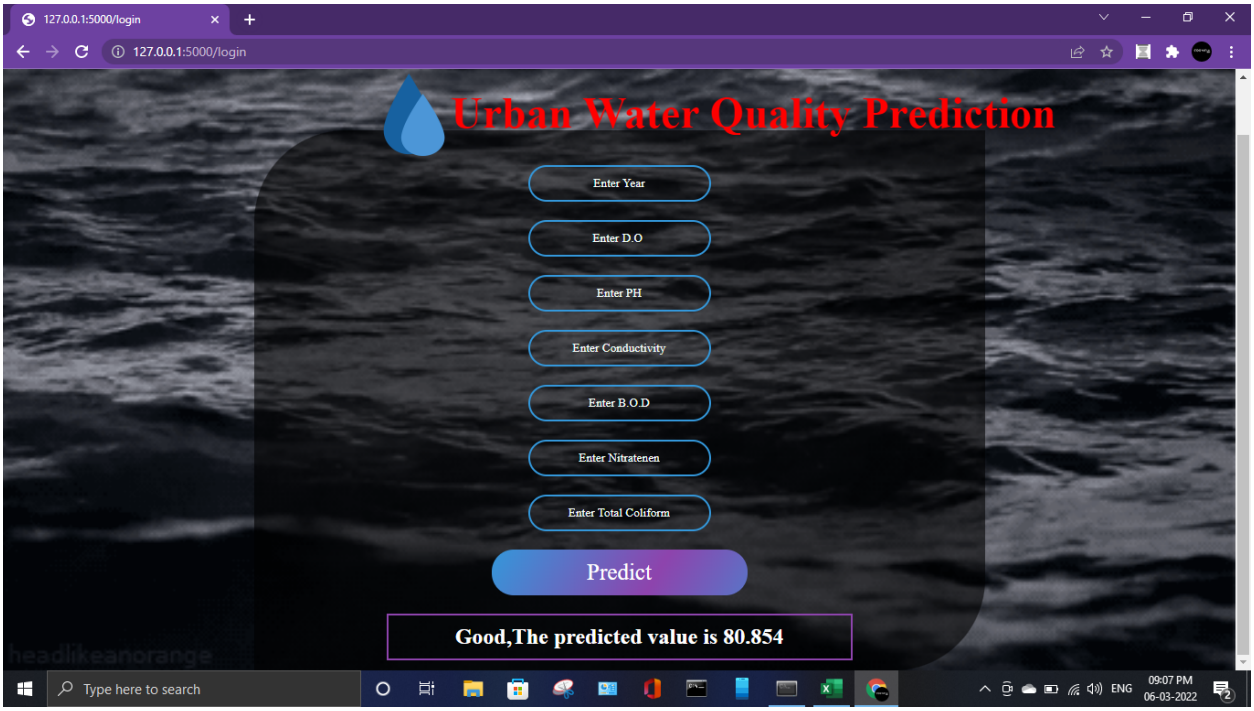


When we input the values in each input fields after finding the ph value, nitrate, DO, BOD, year, conductivity and total coliform

Input as year: 2003, DO:7.8, ph: 7, conductivity:74, B O D: 2.3, Nitrate: 4.3, Total coliform: 108

The Result Given Below:-





Here we got the prediction of urban water quality as 80.854 and Good.

7. ADVANTAGES & DISADVANTAGES

7.1 Advantages

Faster Save time:

We can predict the output very fast compared to other system and it save time due to machine learning so we can predict the quality of the water using the machine learning techniques

Cost Efficiency:

Due to automation of everything therefore the cost is reduced compared to the manual process.

1. Automation of Everything

Machine Learning is responsible for cutting the workload and time. By automating things we let the algorithm do the hard work for us. Automation is now being done almost everywhere. The reason is that it is very reliable. Also, it helps us to think more creatively.

Due to ML, we are now designing more advanced computers. These computers can handle various Machine Learning models and algorithms efficiently. Even though automation is spreading fast, we still don’t completely rely on it. ML is slowly transforming the industry with its automation.

2. Wide Range of Applications

ML has a wide variety of applications. This means that we can apply ML on any of the major fields. ML has its role everywhere from

medical, business, banking to science and tech. This helps to create more opportunities.

3. Scope of Improvement

Machine Learning is the type of technology that keeps on evolving. There is a lot of scope in ML to become the top technology in the future. The reason is, it has a lot of research areas in it. This helps us to improve both hardware and software.

In hardware, we have various laptops and GPUs. These have various ML and Deep Learning networks in them. These help in the faster processing power of the system. When it comes to software we have various UIs and libraries in use. These help in designing more efficient algorithms.

4. Efficient Handling of Data

Machine Learning has many factors that make it reliable. One of them is data handling. ML plays the biggest role when it comes to data at this time. It can handle any type of data.

Machine Learning can be multidimensional or different types of data. It can process and analyze these data that normal systems can’t. Data is the most important part of any Machine Learning model. Also, studying and handling of data is a field in itself.

7.2 Disadvantages

1. Possibility of High Error

In ML, we can choose the algorithms based on accurate results. For that, we have to run the results on every algorithm. The main problem

occurs in the training and testing of data. The data is huge, so sometimes removing errors becomes nearly impossible. These errors can cause a headache to users. Since the data is huge, the errors take a lot of time to resolve.

2. Algorithm Selection

The selection of an algorithm in Machine Learning is still a manual job. We have to run and test our data in all the algorithms. After that only we can decide what algorithm we want. We choose them on the basis of result accuracy. The process is very much time-consuming.

3. Data Acquisition

In ML, we constantly work on data. We take a huge amount of data for training and testing. This process can sometimes cause data inconsistency. The reason is some data constantly keep on updating. So, we have to wait for the new data to arrive. If not, the old and new data might give different results. That is not a good sign for an algorithm.

4. Time and Space

Many ML algorithms might take more time than you think. Even if it’s the best algorithm it might sometimes surprise you. If your data is large and advanced, the system will take time. This may sometimes cause the consumption of more CPU power. Even with GPUs alongside, it sometimes becomes hectic. Also, the data might use more than the allotted space.

8. APPLICATIONS

We use different types of applications to predict and detect the quality of the water

Like Real-time monitoring and water pollution alarming is also type of application that use to conclude the quality of the water. LGC ,Center of Process, Access sensor are some of the top companies that test the quality of the water using this machine learning applications and Auto AI methods and also uses deep learning methods to test the quality of the water.

9. CONCLUSION

Using this machine learning techniques and IBM Watson studio we are able to detect the quality of the urban water, Its only a prediction not an exact value due to automation there will be slight change in prediction but using the algorithm we get the accurate values by cross checking data values using Random Forest Regression and finally we get the quality of the water using the factors such as ph, conductivity, year etc..

In this dissertation, the water quality prediction problems are classified into four to the dataset size. Several machine learning methods based on the ANFIS model are proposed to improve the prediction performance for problems falling into different Scenarios.

First, to eliminate out-of-range errors of the ANFIS model in the testing stage, the stratified sampling strategy is used for mitigating the uneven distribution of training and testing datasets.

A general framework of water quality prediction system based on wavelet de-noised ANFIS model using stratified sampling is proposed.

For problems in Scenario 2, i.e., medium size dataset and strong correlation between parameters, wavelet transform is integrated with ANFIS model to predict wastewater discharge quality parameters related to salinity levels in Chapter 4. The initial dataset is preprocessed with

a statistical stratified sampling strategy. Experimental results show that the proposed model with stratified sampling outperforms MLR, ANNs, ANFIS, EANFIS, and WT-ANFIS models. This demonstrates that the proposed model, with stratified sampling and a general-purpose input parameter selection method, is reliable to model the quality of parameters related to salinity. This model can be applied to reduce the number of parameters monitored to lower the cost associated with monitoring the quality of wastewater discharges.

10. FUTURE SCOPE

-

While machine learning artificial intelligence may be seen as a data-hungry machine, the crucial aspect of a successful AI system that manages a client’s healthcare is its ability to develop efficient reasoning and intuitively read and understand trends.

So in the future we can add more parameters to predict and analyze the urban water quality fast and accurate. In the future, more evolution algorithms, like differential evolution and ant colony.

Optimization, will be explored to forecast the water quality. On the other hand, ideas from other data processing methods, like boost learning and weighted timing analysis will be investigated.

11. BIBILOGRAPHY

- https://ieeexplore.ieee.org/document/7494106
- https://www.datascience2000.in/2021/10/water-quality-prediction-using-machine.html
- https://iwaponline.com/wqrj/article/53/1/3/38171/Water-quality-prediction-using-machine-learning
- https://www.researchgate.net/publication/351077205_Efficient_Water_Quality_Prediction_for_Indian_Rivers_Using_Machine_Learning
- https://www.sciencedirect.com/science/article/abs/pii/S004896972200403X