Urban Water Quality Index Prediction

SmartBridge - Remote Summer Internship Program

Category: Machine Learning

BY:

N.SUSHMA

Index

1. Introduction:-

a. Overview:

i. Water is perhaps the most precious natural resource after air. Though the surface of the earth mostly consists of water, only a small part of it is usable, which makes this resource limited. This Precious and limited resource, therefore, must be used with care.

ii. As water is required for different purposes, the suitability of it must be checked before use. Also, sources of water must be monitored regularly to determine whether they are in sound health or not.

iii. Poor condition of water bodies are not only the indicator of environmental degradation, it is also a threat to the ecosystem. In industries,improper quality of water may cause hazards and severe economic loss. Thus, the quality of water is very important in both environmental and economic aspects. Thus, water quality analysis is essential for using it for any purpose. After years of research, water quality analysis now consists of some standard protocols.

iv. There are guidelines for sampling, preservation and analysis of the samples. Here the standard chain of action is discussed briefly so that it may be useful to the analysts and researchers.

b. Purpose:

Water quality analysis is required mainly for monitoring purposes. Some importance of such assessment includes:

i. To check whether the water quality is in compliance with the standards, and hence, suitable or not for the designated use.

ii. To monitor the efficiency of a system, working for water quality maintenance

iii. To check whether upgradation / change of an existing system is required and to decide what changes should take place.

iv. To monitor whether water quality is in compliance with rules and regulations.

2. Literature Survey

2.1 Existing Problem

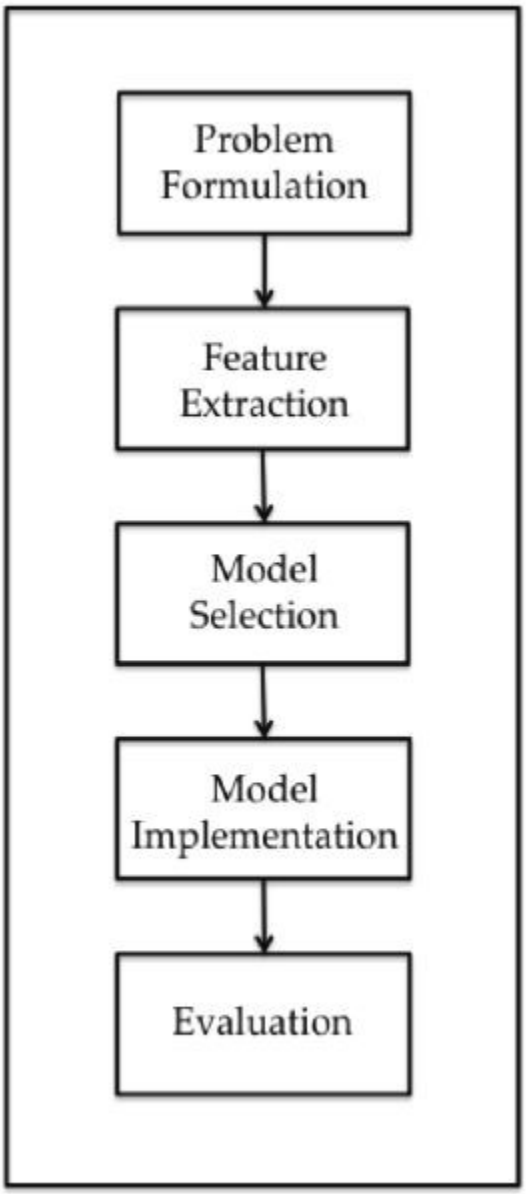
Water is the most abundant and essential resource in India. At present, the quality of water resources is one of the most critical environmental challenges from the national perspective. Increasing population coupled with rapid urbanization, industrialization, and agricultural growth has resulted in the deterioration of water quality in the country. The available water resources in many parts of the country are becoming polluted because of the discharge of untreated sewage, industrial effluents, etc. This pressing situation demands a solution to improve the quality of water resources in the country through water quality management strategies.

2.2 Proposed Solution

With the use of Machine Learning Model, there will be no limitation of the complexity increasing number of variables. This Model and train and test the given factors which Predicts water quality index and with the best performing machine learning model it can effortlessly predict the WQI of water with much higher accuracy than traditional methods.

Theoretical Analysis

a. **Block Diagram**



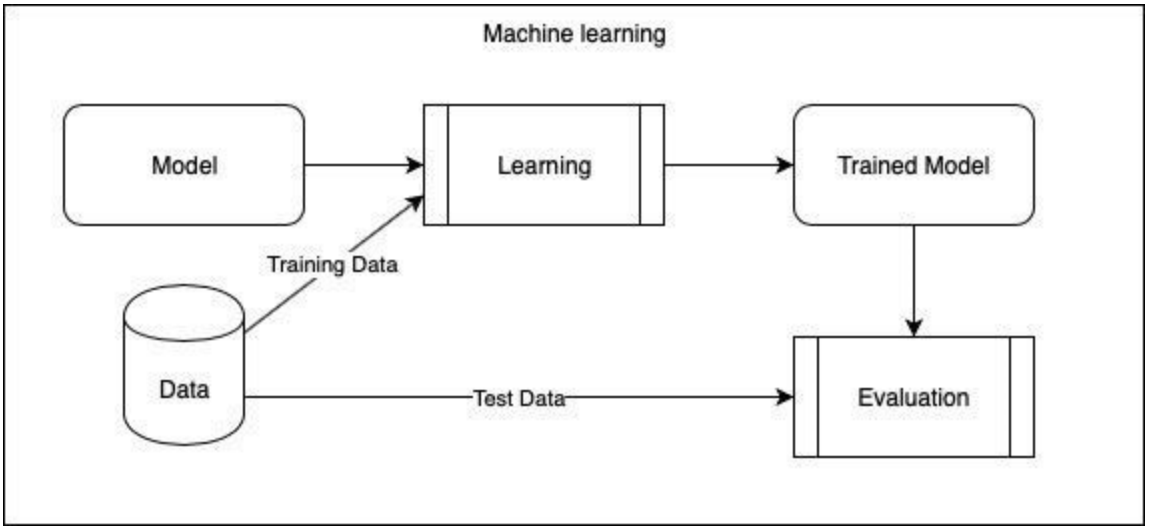
b. Hardware/Software Designing:

Python, Python Web Frameworks, Python for Data Analysis, Python for Data Visualization, Data Pre-processing Techniques, Machine Learning, Regression Algorithms.

Experimental Investigations:

In our project, we have used the Indian Water quality dataset. The data that is used in this project originally comes from the kaggle machine learning dataset. We got to know all the required parameters to predict WQI and we also analysed different models and concluded the best model for predicting the output.

Flowchart :

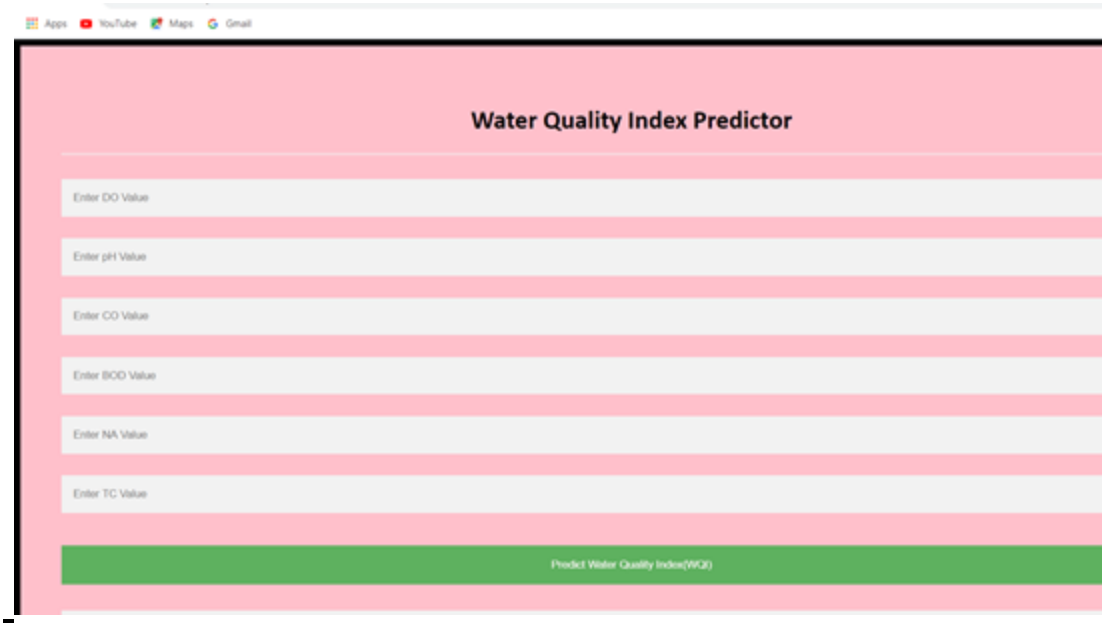


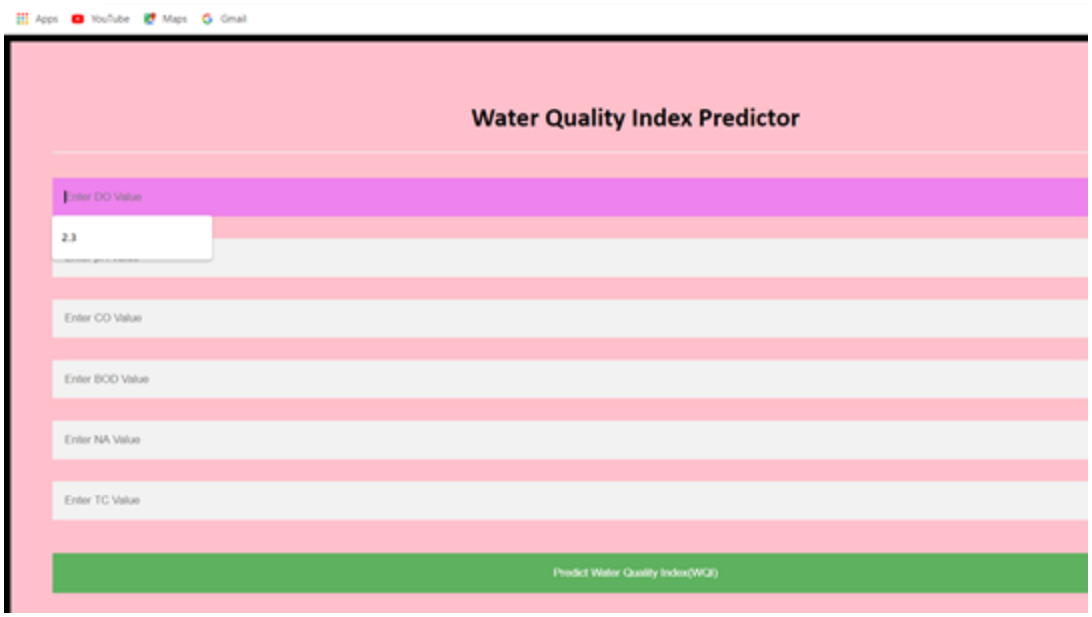
Result:

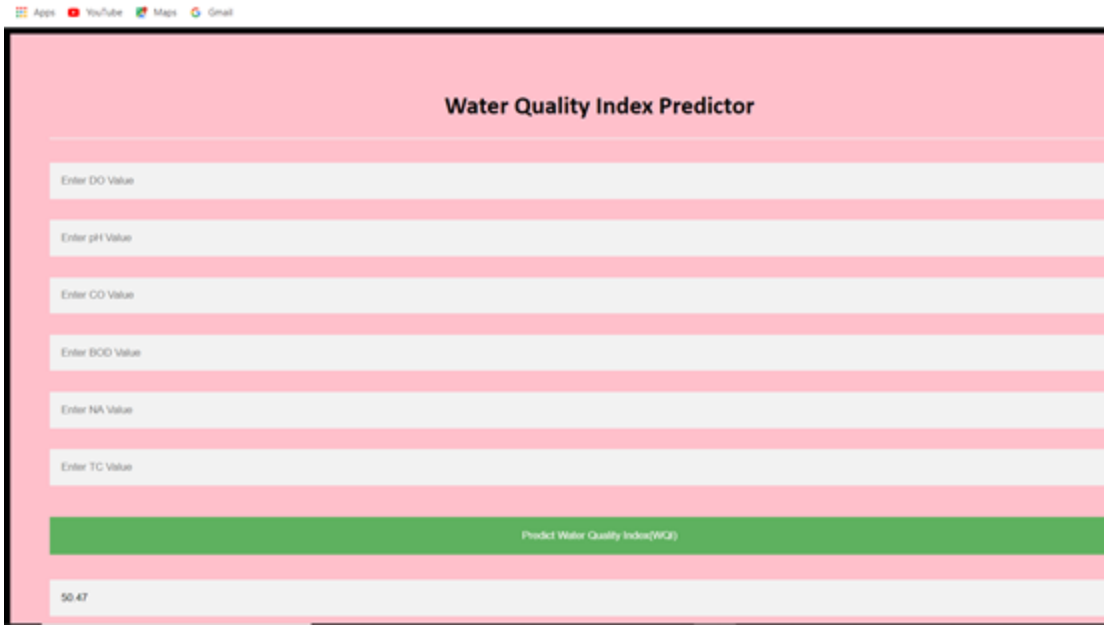
We have used Random Forest Classifier and linear regression to make predictions and compared their performance. Random Forest has highest accuracy and is a good choice for this problem. Random Forest trains the model with subsets of data sampled from the training data, this will make our model more accurate.

In this project, the Random Forest algorithm is used to predict its performance. The results show 97% accuracy.

Snapshots:


Apps
YouTube
Maps
Gmail
Water Quality Index Predictor
Enter DO Value
Enter pH Value
Enter CO Value
Enter BOD Value
Enter NA Value
Enter TC Value
Predict Water Quality Index(WQI)


Apps
YouTube
Maps
Gmail
Water Quality Index Predictor
Enter DO Value
2.3
Enter CO Value
Enter BOD Value
Enter NA Value
Enter TC Value
Predict Water Quality Index(WQI)



Advantages & disadvantages:

Advantages:

The benefits of this model are:

1. No human interference is required
2. Easy interface
3. Accurate calculations ● Faster Results

Disadvantages:

1. Random forests have been shown to fit over certain noisy classification or regression problems.
2. For data with different values, attributes with more values will have a greater

impact on random forests, so the attribute weights generated by random forests on such data are not credible

APPLICATIONS:

Water quality Index analysis using Random Forest algorithm tells the wqi index of water by giving basic parameters.

Predicted results are very accurate

Minimizing the physical efforts

Conclusion:

This project helps the researchers and common people to predict the WQI Very easily

Future Scope:

The future Scope of this project is, we can also remove GSM hardware and we can receive the emergency alerts to mobile phones from cloud servers through the internet.

Bibliography:-

https://www.kaggle.com/anbarivan/indian-water-quality-data

Data from **water_dataX.csv** dataset.

APPENDIX