

Report On Prediction Urban Water Quality Using Machine Learning

- By Sahil Faizal

1.INTRODUCTION

OVERVIEW

With the rapid development of economy and accelerated urbanization, water pollution has become more and more serious. Urban water quality is of great importance to our daily lives. Prediction of urban water quality help control water pollution and protect human health. Water shortage problems in semi-arid areas have become more and more serious in recent years. Recent studies show that a lack of water resources could affect nearly 5.5 billion people in 10 years. Severe water shortages and large volumes of sewage render river and lake water pollution issues serious in arid areas. The water quality of rivers and lakes is becoming central to human and economic development. Therefore, the evaluation and estimation of water quality levels is essential for societal and economic development.

PURPOSE

Understanding the problems and trends of water pollution is of great significance for the prevention and control of water pollution. In order to overcome this kind of problem, I have developed a machine learning model to predict the water quality based on the inputs given by the user and to forecast the predictions. With advances in space information science and with an increasing use of computer applications in recent years, remote sensing has become a useful tool of surface parameter monitoring. It allows one to monitor large scale water bodies that suffer from qualitative problems more effectively.

2.LITERATURE SURVEY

EXISTING PROBLEM

Nowadays majority of people in cities face the issues of water pollution so there is a great necessity of an water quality monitoring system which can give them accurate predictions of water quality , which keeps on varying from depending upon the change in physical and chemical factors of the water source.

PROPOSED SOLUTION

People can use water quality measuring devices that are available in the market to measure the required environmental variables that are required for prediction.

Then users can use the interface website which is hosted on the internet using a browser and then enter the measured values in the specified boxes. Based on the values entered by users the Machine Learning model will predict the water quality index based on the input.

3.THEORITICAL ANALYSIS

BLOCK DIAGRAM



SOFTWARE DESIGN

A User Interface of Predicting Urban Water Quality using Machine Learning was created in software designing using following software technologies:

- 1.Flask
- 2.HTML
- 3.CSS
- 4.Github
- 5.Zoho Writer
- 6.Heroku

4.EXPERIMENTAL INVESTIGATIONS

The water quality index (WQI) has been used to identify threats to water quality and to support better water resource management. So WQI is the best measure which we can use to predict the water quality based on the input features we have on the dataset.

WQI index is calculated using the features like Dissolved Oxygen, pH, Conductivity, Biological Oxygen Demand and Temperature etc. And for training the model effectively we need to scale down the default values in our dataset and multiply them as the their respective standard weights as mentioned by the WHO(World Health Organization). Thus the linear relation between the environment variables is evident.

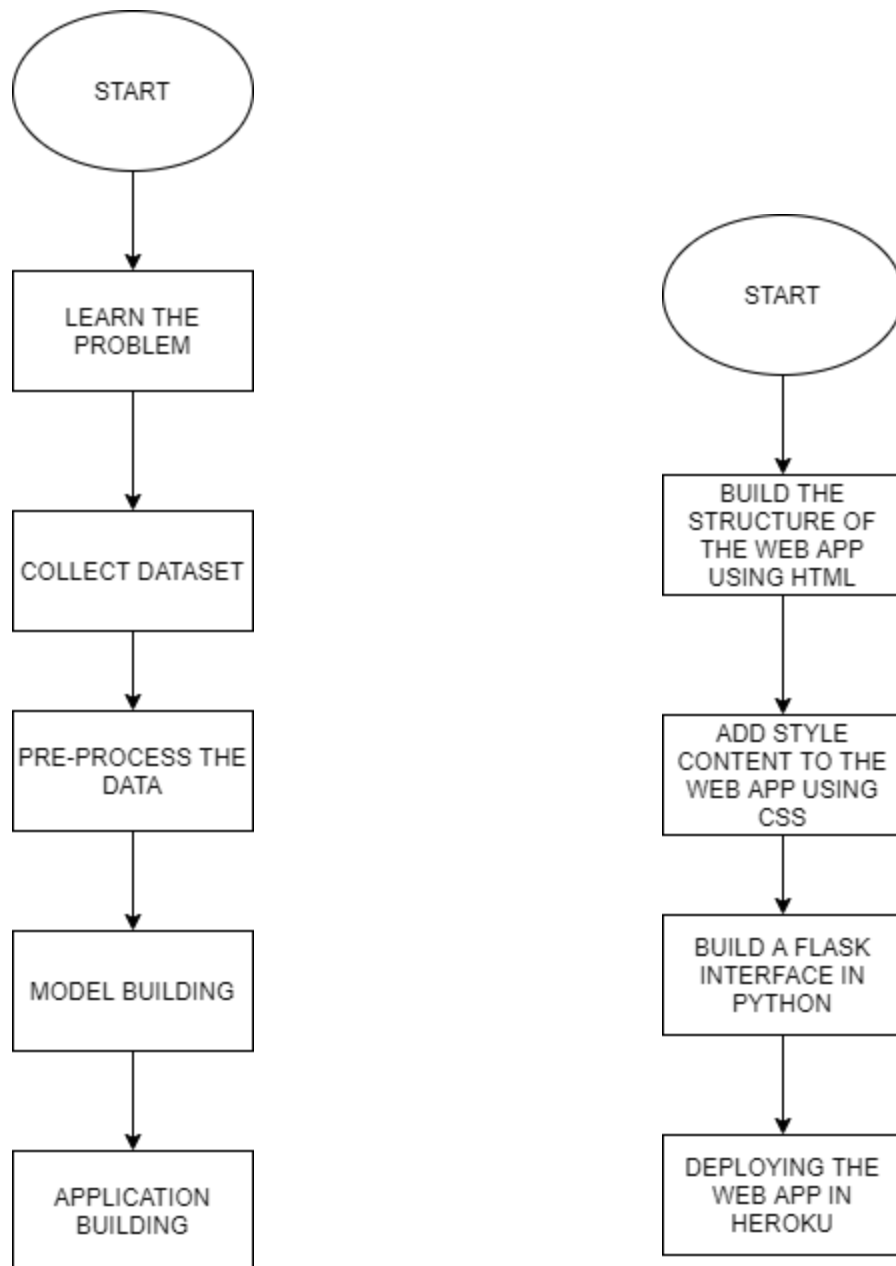
It was found that the dataset had a large number of outliers but their removal could result in nearly 55% of the entire dataset which is considered to be a loss for the training of the model. Also many "NULL" values were found in the dataset, which were replaced

URBAN WATER QUALITY PREDICTION

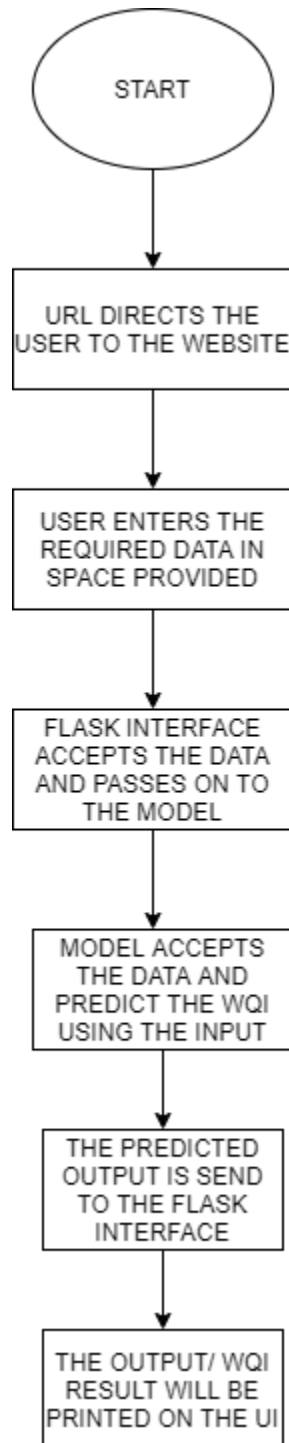
by the respective column means after type casting all the required values into required data types. After that WQI was computed and the column was added to existing dataframe. Then data visualization was done using Seaborn and Matplotlib Libraries.

5.FLOWCHART

FLOW OF STEPS FOR ML MODEL & USER INTERFACE DEVELOPMENT



WORKFLOW OF THE SOFTWARE



6.RESULT

The Urban Water Quality Prediction Model was trained & tested using Jupyter Notebook. As I have implemented the project using Multiple Linear Regression which was used because of the fact that multiple factors or features present in the dataset showed a linear relationship with the WQI(predicted value) which was calculated as per the WHO standards. The most important feature on which the water quality index depended on was 'Income Composition of Resources'. The User Interface was successfully implemented by using HTML, CSS and Flask. The integrated web application was then deployed in Heroku & after filling input fields in the UI, we obtained our predicted value.

7.ADVANTAGES AND DISADVANTAGES

ADVANTAGES

- The UI made is easy to use and can be scaled further by adding more functionality.
- Model is capable of providing accurate results for the user inputs.
- Critical factors needed for the precise prediction of the water quality is put into use for the model creation.
- Can be accessed easily by anyone from around the world with ease.
- As the Machine Learning model is integrated within the UI people who are unaware of the ML technicality can also utilise this website.

DISADVANTAGES

- Error in data can result in wrong predictions.
- The model require some more manual tuning.
- As the model is made using the available data the insights will also be biased towards this dataset there can be other crucial factors for determining the water quality.
- Nearly 45% of the data comprises of outliers ie; data is not clean and can be biased to some values.

8.APPLICATIONS

- The model can be implemented in Pollution Control and Monitoring Organizations for prediction of water quality.
- The model can be utilized by the residents in urban areas to monitor the quality of water they are consuming.
- The model can also be used at Water Distribution Institutions, which plays key role of water supply in urban areas so as to provide clean water to the people.

9.CONCLUSION

The Prediction of Urban Water Quality using Machine Learning was successfully implemented. The UI was created using HTML and CSS as Front-End and Flask for the Back-End. Even though a working Machine Learning model is built the software has to undergo more upgrading in order to be made production ready. As nearly 45% of the data comprises of the outliers the model can be biased also. Main requirement for the further development of the software is precise and accurate data. The UI has been tested & deployed successfully.

10.FUTURE SCOPE

Prediction of Water Quality using Machine Learning can be a foundation for predicting various factors that affect human life like air quality, rainfall, agricultural productivity etc. Some additional features can be included by using an enhanced dataset with less outliers and more sensible data. Also, the UI can be improved by adding some visualization and interactive features to it. The analysis of Big Data using Machine Learning can provide some good insights to various natural and human made factors that affect people around the world.

A sensor network can be implemented for the real time monitoring of water quality. As the sensor data can be sent to the cloud in which further processing can be done and conclusions can be drawn from visualizations and the predictions made. Through this the concerned authorities become aware about the recent trends on water quality and necessary steps can be undertaken to prevent deterioration of the human health and natural wealth.

11.BIBLIOGRAPHY

1. <https://www.analyticsvidhya.com/blog/2020/04/how-to-deploy-machine-learning-model-flask/>
2. <https://towardsdatascience.com/data-preprocessing-concepts-fa946d11c825>
3. <https://www.itrcweb.org/RISK-3/Content/4.%20Data%20Evaluation.htm>
4. https://www.statisticssolutions.com/regression-analysis-multiple-regression/?__cf_chl_jschl_tk__=e0a600a6cd7a128fb545245a4cf8ed7a6215ad18-1610649623-0-AZAzBHmABjEUMFmeL6HjhrTJhN-
5. <http://www.ijstr.org/final-print/oct2019/Water-Quality-Monitoring-Using-Machine-Learning-And-lot.pdf>

12.APPENDIX

SOURCE CODE

- **Github Application Link:** <https://github.com/sahilf2001/Urban-Water-Quality>
- **Github Model Link:**
https://github.com/sahilf2001/Urban_Water_Quality_Prediction_Model_File
- **UI Link:** <https://urbanwaterqualitypredictor.herokuapp.com/>
- **Dataset Link:** <https://www.kaggle.com/anbarivan/indian-water-quality-data>
- **Demonstration Video Link:**
https://drive.google.com/file/d/1_iU1YdytvSzjBrJ0rGGEIVQswp3vbMh6/view?usp=sharing

UI OUTPUT SCREENSHOT

WQI Prediction

Enter the integer values as input

Temperature

Dissolved Oxygen(mg/l)

pH value(0-14)

Conductivity(umhos/cm)

Biological Oxygen Demand(mg/l)

Nitrate Content(mg/l)

Total Coliform(MPN/100ml)Mean

Year

Predict

URBAN WATER QUALITY PREDICTION

WQI Prediction

Enter the integer values as input

35

5

8

98

4

1

156

2014

Predict

Water Quality Index would 272.0