

WATER QUALITY INDEX PREDICTION

S V HEMANTH¹, M SRAVANTH², R DHANUSH³

¹ Professor, department of cse, Hyderabad institute of technology, Hyderabad, Telangana

² Student, department of cse, Hyderabad institute of technology, Hyderabad, Telangana

³ Student, department of cse, Hyderabad institute of technology, Hyderabad, Telangana

Abstract. 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. 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. 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. 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.

1. Introduction

1. About Project

As we know only 71% of earth is covered with water bodies so it is very necessary to protect water. Even in the 71% we segregate usage of water for drinking, Cleaning, and for the other uses. This Precious and limited resource, therefore, must be used with care. 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. 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.

1.2 Objectives of the Project

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

1. To check whether the water quality is in compliance with the standards, and hence, suitable or not for the designated use.
2. To monitor the efficiency of a system, working for water quality maintenance
3. To check whether upgradation / change of an existing system is required and to decide what changes should take place.
4. To monitor whether water quality is in compliance with rules and regulations.

1.3 Scope of the Project

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. 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.

2. Literature Survey

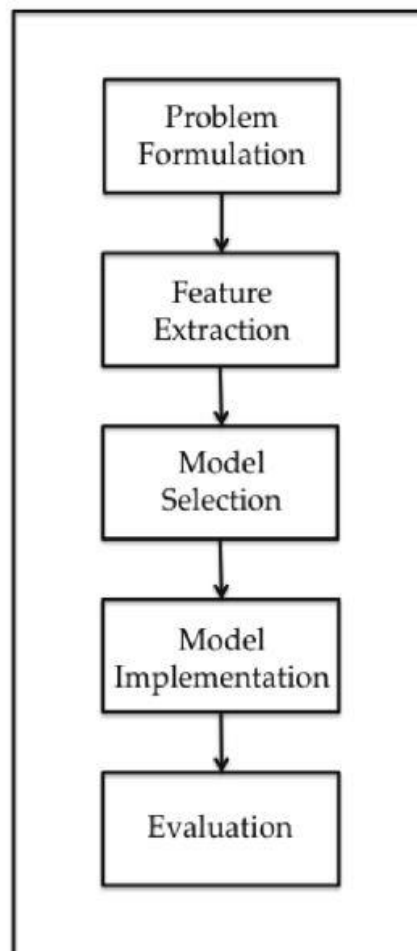
2.1 Existing System

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 System

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.

3. Proposed Architecture



4. Implementation

4.1 Algorithm

Random forest

Random Forest works in two-phase first is to create the random forest by combining N decision tree, and second is to make predictions for each tree created in the first phase.

The Working process can be explained in the below steps and diagram:

Step-1: Select random K data points from the training set.

Step-2: Build the decision trees associated with the selected data points (Subsets).

Step-3: Choose the number N for decision trees that you want to build.

Step-4: Repeat Step 1 & 2. Step-5: For new data points, find the predictions of each decision tree, and assign the new data points to the category that wins the majority votes.

Logistic regression

1. Logistic regression is one of the most popular Machine Learning algorithms, which comes under the Supervised Learning technique. It is used for predicting the categorical dependent variable using a given set of independent variables.

2. Steps in Logistic Regression: To implement the Logistic Regression using Python, we will use the same steps as we have done in previous topics of Regression. Below are the steps: Data Pre-processing step Fitting Logistic Regression to the Training set Predicting the test result Test accuracy of the result (Creation of Confusion matrix) Visualizing the test set result.

4.2 Code Implementation


Anaconda Navigator Anaconda Navigator is a desktop graphical user interface (GUI) included in Anaconda® distribution that allows you to launch applications and easily manage conda packages, environments, and channels without using command-line commands. Navigator can search for packages on Anaconda.org or in a local Anaconda Repository. It is available for Windows, macOS, and Linux. In order to run, many scientific packages depend on specific versions of other packages. Data scientists often use multiple versions of many packages and use multiple environments to separate these different versions. The command-line program conda is both a package manager and an environment manager. This helps data scientists ensure that each version of each package has all the dependencies it requires and works correctly. Navigator is an easy, point-and-click way to work with packages and environments without needing to type conda commands in a terminal window. You can use it to find the packages you want, install them in an environment, run the packages, and update them – all inside Navigator.

Anaconda Prompt Anaconda command prompt is just like command prompt, but it makes sure that you are able to use anaconda and conda commands from the prompt, without having to change directories or your path. When you start Anaconda command prompt, you'll notice that it adds/("prepends") a bunch of locations to your PATH.

Python 3.7. Python is broadly utilized universally and is a high-level programming language. It was primarily introduced for prominence on code, and its language structure enables software engineers to express ideas in fewer lines of code. Python is a programming language that gives you a chance to work rapidly and coordinate frameworks more effectively.

Jupyter Notebook Jupyter Notebook is an open-source, web-based interactive environment, which allows you to create and share documents that contain live code, mathematical equations, graphics, maps, plots, visualizations, and narrative text. It integrates with many programming languages like Python, PHP, R, C#, etc.

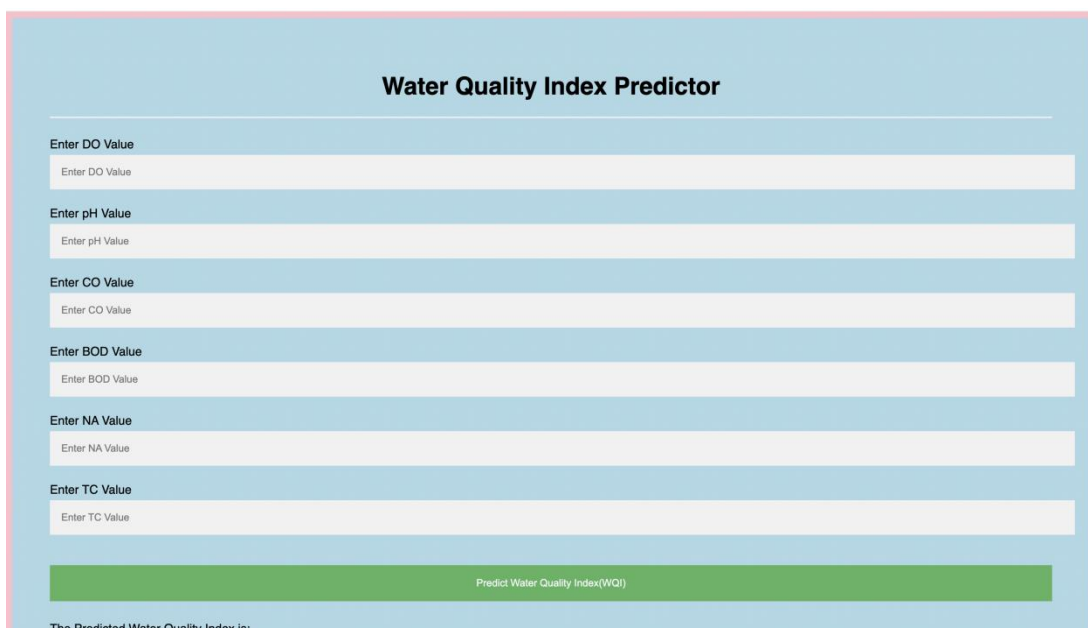
1. All in one place: As you know, Jupyter Notebook is an open-source web-based interactive environment that combines code, text, images, videos, mathematical equations, plots, maps, graphical user interface and widgets to a single document.



The screenshot shows a web form titled "Water Quality Index Predictor". It features seven input fields for different water quality parameters: DO Value, pH Value, CO Value, BOD Value, NA Value, and TC Value. Each field is labeled "Enter [Parameter] Value" and contains a placeholder text "Enter [Parameter] Value". Below these fields is a green button labeled "Predict Water Quality Index(WQI)". Underneath the button, there is a label "The Predicted Water Quality Index is:" followed by a text area containing the placeholder "{(wqi)}".

2. Easy to convert: Jupyter Notebook allows users to convert the notebooks into other formats such as HTML and PDF. It also uses online tools and nbviewer which allows you to render a publicly available notebook in the browser directly.
3. Easy to share: Jupyter Notebooks are saved in the structured text files (JSON format), which makes them easily shareable.
4. Language independent: Jupyter Notebook is platform-independent because it is represented as JSON (JavaScript Object Notation) format, which is a language-independent, text-based file format. Another reason is that the notebook can be processed by any programming language, and can be converted to any file formats such as Markdown, HTML, PDF, and others.
5. Interactive code: Jupyter notebook uses ipywidgets packages, which provide many common user interfaces for exploring code and data interactivity.

5.Result



The screenshot shows a web form titled "Water Quality Index Predictor". It features seven input fields for different water quality parameters: DO Value, pH Value, CO Value, BOD Value, NA Value, and TC Value. Each field is labeled "Enter [Parameter] Value" and contains a placeholder text "Enter [Parameter] Value". Below these fields is a green button labeled "Predict Water Quality Index(WQI)". Underneath the button, there is a label "The Predicted Water Quality Index is:" followed by a text area containing the placeholder "{(wqi)}".

6. Conclusion

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.

7. Future Scope

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

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

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

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

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

8. References

1. <http://sersec.org/journals/index.php/IJAST/article/download/33228/18386/>
2. <http://sersec.org/journals/index.php/IJAST/article/view/33228>
3. <https://realpython.com/python-web-applications/>
4. Umair Ahmed, Rafia Mumtaz, Hirra Anwar, Asad A Shah and Rabia Irfan, 2019, Efficient Water Quality Prediction Using Supervised Machine Learning, MDPI, vol.11, no. 2210, pp. 1-14
5. <https://www.digitalocean.com/community/tutorials/how-to-make-a-web-application-using-flask-in-python-3>
6. M. Qun, G. Ying, L. Zhiqiang and T. Xiaohui, "Application of Comprehensive Water Quality Identification Index in Water Quality Assessment of River," 2009 WRI Global Congress on Intelligent Systems, 2009, pp. 333-337, doi: 10.1109/GCIS.2009.166.
7. Hongfang Lu, Xin Ma, Hybrid decision tree-based machine learning models for short-term water quality prediction, Chemosphere, Volume 249, 2020, 126169, ISSN 0045-6535