

Ideation Phase

Literature Survey On The Selected Project & Information Gathering

Date	02 November 2022
Team ID	PNT2022TMID21084
Project Name	Efficient Water Quality Analysis & Prediction using Machine Learning
Maximum Marks	-

Literature Survey

- The first paper is Predicting and analyzing water quality using Machine Learning. In this paper, data includes the measurements of four parameters which will affect the quality of water. This issue has been addressed in many researches, however more work needs to be done in terms of effectiveness, reliability, accuracy as well as usability of the current water quality management methodologies. To develop a water quality prediction model with the help of water quality factors using ANN and time series analysis was proposed in this paper.
- The Second paper is A Network for Monitoring and Assessing Water Quality for Drinking and Irrigation Purposes. From this paper factors such as industrialization, mining, pollution, and natural occurrences impact the quality of water, as they introduce or alter various parameters that affects the human consumption. The Water Quality Index (WQI) and Irrigation WQI (IWQI) are metrics used to express the level of these parameters to determine the overall water quality. Three ML models - Random Forest (RF), Logistic Regression (LR) and Support Vector Machine (SVM) were considered for the water classification process and results were obtained.
- The Third paper is Quality Monitoring of Abu Dhabi Drinking Water Using Machine Learning Classifiers. Supplying high quality drinking water is important for human health. It is necessary to monitor the drinking water quality to prevent health damage and to avoid water pollution. According to this paper there were five machine learning algorithms that were applied for water quality monitoring using standard water physical and chemical data which was proposed in this paper. These were algorithms Logistic regression, a Support Vector Machines, a K-Nearest Neighbors, a Naive Bayes, and Decision Trees which had been used to obtain the result efficiently.

Information Gathering

1.1 Abstract

Lakes and reservoirs are important water resources. Reservoirs are vital water resources to support all living organisms. They provide clean water and habitat for a complex variety of aquatic life. Apart from this, reservoirs are also helpful to get hydro-electric power, flood control and scenic beauty. Water collected in such resources can be utilized in drought situations also. Unfortunately, these important resources are being polluted and the quality of water is being influenced by numerous factors. The quality of water is deteriorated by anthropogenic activities, indiscriminate disposal of sewage, human activities and also industrial waste. Water quality monitoring of reservoirs is essential in exploitation of aquatic resources conservation. The quality of water helps in regulating the biotic diversity and biomass energy and rate of succession. Moreover, contaminated water can lead to some waterborne diseases and also influences child mortality. In order to reduce the effect of contaminated water, it is essential to assess different aspects of water quality. Predicting water quality parameters a few steps ahead can be beneficial to achieve this. The main objective of this study is to provide fairly accurate predictions for variable data. The study shows that deep learning techniques which use unsupervised learning to provide accurate results as compared to the techniques based on supervised learning. The comparison of results show that robustness can be achieved by denoising autoencoder and deep belief networks and also successfully handling the variability in the data. Merits of the unsupervised learning algorithms are evaluated on the basis of metrics such as mean absolute error and mean square error to examine the error rate of prediction.

1.2 Introduction

Water quality analysis is a complex topic due to the different factors that influence it. This concept is inextricably linked to the various purposes for which water is used. Different needs necessitate different standards. There is a lot of study being done on water quality prediction. Water quality is normally determined by a set of physical and chemical parameters that are closely related to the water's intended usage. The acceptable and unacceptable values for each variable must then be established. Water that meets the predetermined parameters for a specific application is considered appropriate for that

application. If the water does not fulfill these requirements, it must be treated before it may be used. Water quality can be assessed using a variety of physical and chemical properties. As a result, studying the behavior of each individual variable independently is not possible in practice to accurately describe water quality on a spatial or temporal basis. The more challenging method is to combine the values of a group of physical and chemical variables into a single value. A quality value function (usually linear) represented the equivalence between the variable and its quality level was included in the index for each variable. These functions were created using direct measurements of a substance's concentration or the value of a physical variable derived from water sample studies. The major goal of this research is to examine how machine learning algorithms may be used to predict water quality.

1.3 Requirements

Hardware requirements

Operating system- Windows 7,8,10

Processor- dual core 2.4 GHz (i5 or i7 series Intel processor or equivalent AMD)

RAM-4GB

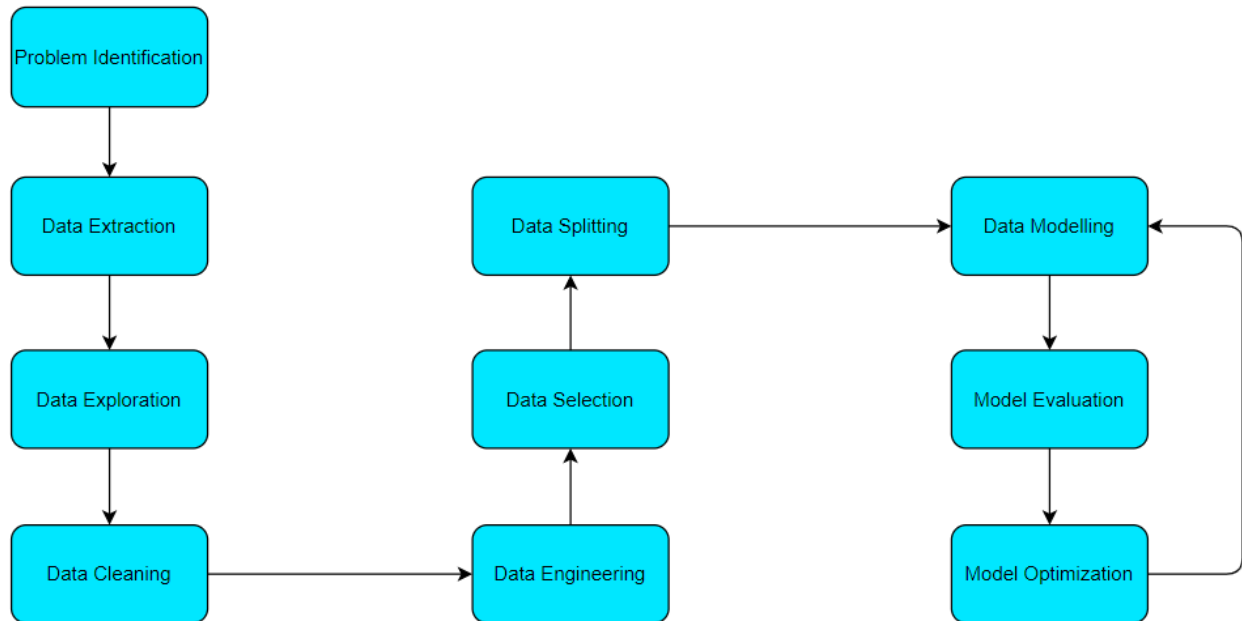
Software Requirements

Python

Jupyter Notebook

Chrome

1.4 Flowchart



Flowchart for Water Quality Analysis & Prediction

1.5 Existing Solution

The most essential things for predicting the water quality are indicator parameters and class labels which are in the dataset. All water quality parameters are unrealistic because it is not only expensive and technically difficult but also fails to deal with the variability in water quality. Water quality testing is very important to check the quality of drinking water to avoid waterborne diseases and improve health. In this paper we have tested the quality of water by measuring several parameters such as pH, total hardness, Solides, Chloramines, Sulfate, Conductivity, Organic_carbon, Trihalomethanes, Turbidity, and Potability.

2. Methodology

There are two primary phases in the system: 1. Training phase: The system is trained by using the data in the data set and choosing the water quality data set, which is a prerequisite to model construction based on the essential parameters that affect water quality. 2. Testing phase: the system is provided with the inputs and is tested for its water quality. The quality of the water is checked by pH value and hardness. And therefore, the data that is used to train the model or test it, has to be appropriate. The system is designed to analyze and predict the quality of the water and hence appropriate algorithms must be used to do the two different tasks. Before the algorithms were selected for further use, different algorithms were compared for its underlying relational structure between the indicator parameters and the class label. The well-suited one for the task was chosen.

3.1 Technical Papers

- **Quality Monitoring of Abu Dhabi Drinking Water Using Machine Learning**

Classifiers , 07-10 December 2021 14th International Conference on Developments in eSystems Engineering (DeSE)

DOI: [10.1109/DeSE54285.2021.9719373](https://doi.org/10.1109/DeSE54285.2021.9719373)

- **Predictive Analytics for Smart Water Management in Developing Regions,**

18-20 June 2018 IEEE International Conference on Smart Computing (SMARTCOMP)

DOI: [10.1109/SMARTCOMP.2018.00047](https://doi.org/10.1109/SMARTCOMP.2018.00047)

3.2 Research Publications

- [1] Brown R. M., McClelland N. I., Deininger R. A., Tozer R. G. 1970 A water quality index- do we dare? Water and Sewage Works, October 1970, 339-343
- [2] Roy R Majumder M. 2018 A Quick Prediction of Hardness from Water Quality Parameters by Artificial Neural Network. International Journal of Environment and Sustainable Development, 17(2/3), 247-257
- [3] Kori R., Parashar S., Basu, D.D. undated Guide Manual: Water and Wastewater Analysis. Central Pollution Control Board, Ministry of Environment and Forest, India
- [4] Colorado Department of Public Health and Environment- Water Quality Control Division (CDPHE-WQCD). "Primary Drinking Water
- [5] Postel, S. L. Daily, G.C. and Ehrlich P.R. (1996) Human appropriation of renewable fresh water. Science 271:785-788.
- [6] Sim, T. S. and Duraka B. J. (1987) Coliphage counts: Are they necessary to maintain drinking water safety. BiotMicro J. Appl. Microtech:223-226.
- [7] Opinya, G.N., Pameijer L.H, and Gron, P. (1987) Analysis of Kenyan drinking water. East Afr. Med. J. 21:194-201.
- [8] Smith, B. A. (1974) Feeding overstrength cow milk to babies, Brit. Med. J. 4:741.
- [9] Hunter, Paul R. Colford, Jark M. Lechevalier; Mark W. Binder S. And Berger, Paul s. (2001) Waterborne Diseases In: Emerging infectious Diseases (conference Panel Summaries) 7 (3): p. 544 supplement, June 2001.
- [10] Robert ``Bob'' Masters (2001) Drugs on tap In: Emerging Contaminants. Water & Wastewater International August, 2001 16(4): 11-12.
- [11] Basavaraj Simpi, S.M. Hiremath, KNS Murthy, K.N.Chandrashekarappa, Analysis of Water Quality Using Physico-Chemical Parameters Hosahalli Tank in Shimoga District, Karnataka, India, Global Journal of Science Frontier Research,2011 pg 31 -34.
- [12] Harender Raj Gautam and Rohitashav Kumar (2010). Better Groundwater Management Can Usher in India into Second Green Revolution, Journal of Rural Development, Vol.58, No.7, Pp. 3-5.