

A Water Quality Prediction Using Machine Learning Techniques

KRISHNAVENI T , PRABADEVI P

Department of computer science

K.R College Of Arts And Science,Kovilpatti,India.krishthangamkdm916@gmail.com

Department of computer science

K.R College Of Arts And Science,Kovilpatti,India. thangaprabhadevi@gmail.com

Abstract:

Drinking water quality is one of the most important factors affecting human health. However, the quality of water not good in many countries, especially in developing countries and insufficient drinking water has caused many diseases in water. In recent years, many types of pollution such as human waste, garbage and vehicle waste have made water, which is the most important nutrient source of the human body. It has many health benefits and helps prevent diseases. Increasing pollution affects water quality and people’s quality of life. Water quality is estimated using classification techniques such as decision tree and support vector machines. It used to predict the water quality.

Keywords— classification, decision tree, SVM.

1.INTRODUCTION

Dirty water is associated with the word pollution. It expresses the presence of impurities and problems in the water. This water not suitable for drinking and is harmful to animals ,birds and humans. Water pollution is the source of many deadly diseases that cause injuries, diseases and even kill human civilization.

Water pollution is a big problem threatening the world. In order to solve the problem, we first find its source. About 70% of the world is water, but humans and animals consume only 1% of fresh and healthy water. People drink fresh water in their daily activities, so must be careful to protect it from pollution.

2.DATASET

The Water Quality database for this research work has been taken from the (KAGGLE) has been acquired from UCI machine learning repository. The dataset has a total of 10 fields, one to nine attributes. The last column is for showing true indicates if water is safe for human consumption where 1 means Potable and 0 means Not potable. (0) Water is not safe to drink and (1) Water is safe to drink.

S.NO	Attribute Name	Type
1	PH	Numeric
2	Hardness	Numeric
3	Solids	Numeric
4	Chloramines	Numeric
5	Sulfate	Numeric
6	Conductivity	Numeric
7	Organic_carbon	Numeric
8	Trihalomethanes	Numeric
9	Turbidity	Numeric

3. LITERATURE SURVEY

PAPER 1: Theyazn H.H AldhyanI et al. [1] predict the water quality Using the various classification algorithm to predict the accuracy value. This architecture obtained 59% accuracy, whereas for the SVM method, the two-step algorithm alongside the SVM technique has achieved 81% classification accuracy.

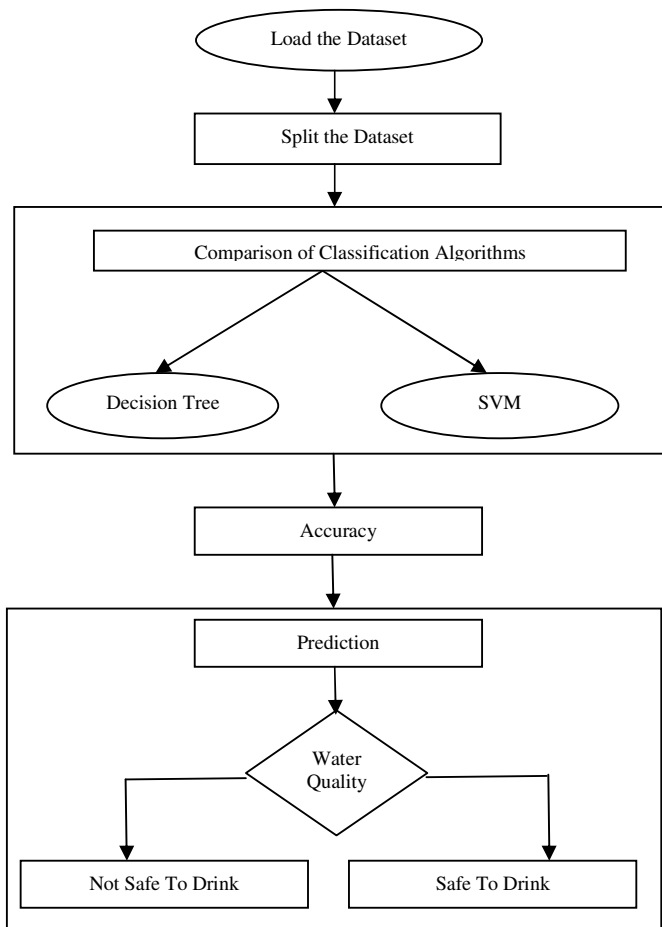
PAPER 2: Rafia Mumtaz et al. [2] Water Quality Prediction Using Supervised Machine Learning: Introduce a method for the prediction of water quality using the variants of Decision

Tree. The modalities used in this technique are the Single Decision Tree (SDT), Boosted Decision Tree (BDT), and Decision Tree Forest (DTF). The decision is taken by training the data set and after that testing. The outcomes presented that the accuracy obtained by the accuracy of 81% in the testing phase.

PAPER 3: Md. Mehedi Hassan et al. prediction of Water Quality Index (WQI) Using Machine Learning Algorithms. The quality of water has a direct influence on both human health and the environment. Water is utilized for a variety of purposes, including drinking, agriculture, and industrial use. The water quality index (WQI) is a critical indication for proper water management.

PAPER 4: Nida Nasir et al. water quality classification using machine learning algorithms Monitoring water quality to implement a water quality prediction using machine learning techniques. In this technique, our model predicts that the water is safe to drink or not using some parameter like PH value, Hardness, etc.

4. ARCHITECTURE



5.METHODS

CLASSIFICATION USING DECISION TREE

Decision trees are constructed via an algorithmic approach that identifies ways to split a data set based on different conditions. It is one of the most widely used and practical methods for supervised learning. They can be used for both classification and regression tasks. It is a tree-structured classifier, where internal nodes represent the features of a dataset, branches represent the decision rules and each leaf node represents the outcome. In a Decision tree, there are two nodes, which are the Decision Node and Leaf Node. Decision nodes are used to make any decision and have multiple branches. Leaf nodes are the output of those decisions and do not contain any further branches. The logic behind the decision tree can be easily understood because it shows a tree-like structure.

Decision Tree Terminologies:

Root Node: Root node is from where the decision tree starts. It represents the entire dataset, which further gets divided into two or more homogeneous sets.

Leaf Node: Leaf nodes are the final output node, and the tree cannot be segregated further after getting a leaf node.

Splitting: Splitting is the process of dividing the decision node/root node into sub-nodes according to the given conditions.

Branch/Sub Tree: A tree formed by splitting the tree.

Pruning: Pruning is the process of removing the unwanted branches from the tree.

Parent/Child node: The root node of the tree is called the parent node, and other nodes are called the child nodes.

SUPPORT VECTOR MACHINE (SVM) :

Support Vector Machine (SVM) is a supervised machine learning algorithm which can be used for both classification and regression challenges. It is a discriminative classifier formally defined by a separating hyperplane. The objective of SVM algorithm is to find a hyperplane in an n-dimensional space that distinctly classifies the data points. The dimension of the hyperplane depends upon the number of features. SVM selects extreme points/vectors which help to create hyperplane. Support vectors represent the obvious cases, hence the technique is called a support vector machine.

6.CONCLUSION

This work reports the impact of different Machine Learning techniques for water quality prediction. The performance analysis of two algorithms namely Decision Tree and Support Vector Machine (SVM) is performed and the accuracy is obtained. When it comes to accuracy, it is observed that Decision Tree algorithms outperforms best more than the Support Vector Machine (SVM). Hence by using the Decision Tree algorithm, the

prediction of water quality is performed and the result is obtained.

REFERENCES

1. Shoba Nath., Raju M.B., Rajagopalan K. and Nandakumar P., Central Groundwater Board, World Bank aided Hydrology Project, Water Resources Organisation, PWD, Chennai, Workshop on Water Quality issues in Tamil Nadu, India, 25 (2019).

2. Kulasekaran A., ChettiaGounder K. and Chellapandian K., Water Quality Problems in Tamil Nadu, Workshop on Water Quality issues in Tamilnadu, World Bank aided Hydrology Project, organised by Water Resources Organisation, PWD, 14 (2015).

3. APHA., Standard methods for the examination of water and waste water, Edition. American Public Health Association, Washington D.C. 21st. Edition (2005).

4. Emamgholizadeh S, Kashi H, Marofpoor I, Zalaghi E (2013) Prediction of water quality parameters of Karoon River (Iran) by artificial intelligence-based models. *Int J Environ Sci Technol* 11:645–656.

5. Rahim NA, Ahmad Z (2013) Features selection in water quality prediction in neural network using canonical correspondence analysis (CCA). In: 6th international conference on process systems engineering (PSE ASIA), pp 25–27, June 2013.

6. Solanki A, Agrawal H, Khare K (2015) Predictive analysis of water quality parameters using deep learning. *Int J Comput Appl* 125(9):29–34.

7. Fatemi, A. (2020). Strategies and policies for water quality management of Gharasou River, Kermanshah, Iran: A review. *Environmental Earth Sciences* (2020) , 79, 254.

8. Zhang, Q., Panpan, X., & Qian, H. (2020). Groundwater quality assessment using improved water quality index (WQI) and human health risk (hhr) evaluation in a semi-arid region of Northwest China. *Exposure and Health*, 2020(12), 487–500.

9. Karthick, T., Gayatri, D., Kohli, T. S., & Pandey, S. (2018). Prediction of Water quality and smart water quality monitoring system in IoT environment. *International Journal of Pure and Applied Mathematics*, 118(20). ISSN: 1314–3395 (On-line Version).

10. Kumar, V., Sharma, A., Chawla, A., Bhardwaj, R., Kumar, A., & Thukral, A. K. (2016). Water Quality Assessment of River Beas, India, using Multivariate and Remote Sensing Techniques. *Environment Monit Assess.*