LITERATURE SURVEY

INTRODUCTION

Water is an inorganic, transparent, and colourless chemical substance required for the survival of most existing organisms and humans. Water quality has a direct impact on public health and the environment. Water is used for various practices, such as drinking, agriculture, and industry. Recently, water sports and entertainment development has greatly helped attract tourists. Water is utilised for multiple purposes, including drinking, agriculture, and industrial use. The water quality index (WQI) is a critical indication of proper water management.

USE CASE

The purpose of this study was to analyse and find the water quality from the water dataset of 1991 from India using ML algorithms. Therefore, designing a model that predicts water quality is essential to control water pollution and alert users in case of poor quality detection. Motivated by these reasons, we use machine learning algorithms in this study to develop a model capable of predicting the water quality index and then the water quality class. Data visualisations were implemented to reveal significant findings such as missing values.

Many studies have been conducted to address water quality problems. Most works employ manual laboratory analysis and statistical analysis to

assist in regulating water quality, while other studies use Machine Learning methods to help to obtain optimised solutions to water quality problems.

EXISTING SOLUTIONS

The proposed system is intended to determine portability. It is divided into two phases, one for training and the other for testing. The following procedures are carried out in both sections. The data set was chosen as follows: The collection of essential parameters that affect water quality, identification of the number of data samples, and definition of the class labels for each data sample present in the data are all factors that go into selecting the water quality data set, which is a prerequisite to model construction. Ten indicator parameters make up the data sets used in this study. pH value and hardness are examples of these factors. The proposed approach, however, is not constrained by the number of parameters or the selection of parameters. A k-fold cross-validation technique is employed to set the learning and testing framework in this study, corresponding to each data sample in the data set.

Using this technique, the dataset is separated into k-disjoint sets of equal size, each with roughly the same class distribution. In turn, this division's subsets are utilised as the test set, with the remaining subsets serving as the training set. These are the Decision Tree (DT) and K-Nearest Neighbour (KNN) methods. Each strategy takes a different approach in terms of the underlying relational structure between the indicator parameters and the class label. As a result, each technique's performance for the same data set is likely to differ. Validating the

performance of different classifiers on an unknown data set: Data mining provides several metrics for validating the performance of different classifiers on an unknown data set. A repeated cross-validation procedure in the Matlab caret package created the learning and testing environment. The following procedure was used to apply the classification algorithm:

1. The data set was split into training (80%) and testing (20%). (20 percent).

2. The training set was subjected to repeated cross-validation, with the number of iterations fixed to Classifiers being trained in this manner.

3. The model's optimal parameter configuration was selected, resulting in maximum accuracy.

4. The model was scrutinised.

Conclusions

Probability determines the quality of water, which is one of the essential resources for existence. Traditionally, testing water quality requires an expensive and time-consuming lab analysis. This study looked into an alternative machine learning method for predicting water quality using only a few simple water quality criteria. To estimate, a set of representative supervised machine learning algorithms was used. It would detect water of lousy quality before it is released for consumption and notify the appropriate authorities.

Machine learning has been widely used as a powerful tool to solve problems in the water environment because it can be applied to predict water quality, optimise water resource allocation, manage water resource shortages, etc. Modelling and prediction of water quality are very important for the protection of the environment. Developing a model by using advanced artificial intelligence algorithms can be used to measure the future water quality.

References

1. PCRWR. National Water Quality Monitoring Programme, Fifth Monitoring Report (2005–2006); Pakistan Council of Research in Water Resources Islamabad: Islamabad, Pakistan, 2007.

2. Ling, J.K.B. Water Quality Study and Its Relationship with High Tide and Low Tide at Kuantan River. Bachelor's Thesis, Universiti Malaysia Pahang, Gambang, Malaysia, 2010.

3. Mehmood, S.; Ahmad, A.; Ahmed, A.; Khalid, N.; Javed, T. Drinking Water Quality in Capital City of Pakistan, 2013.

4. Abbasi, T.; Abbasi, S.A. Water Quality Indices; Elsevier: Amsterdam, The Netherlands, 2012.

5. Abyaneh, H.Z. Evaluation of multivariate linear regression and artificial neural networks in prediction of water quality parameters. J. Environ. Health Sci. Eng. 2014, 12, 40.