Nature Environment and Pollution Technology
An International Quarterly Scientific Journal

p-ISSN: 0972-6268
(Print copies up to 2016)
e-ISSN: 2395-3454

Vol. 21 | No. 5 (Suppl) | pp. 2211-2218 | 2022

Original Research Paper

https://doi.org/10.46488/NEPT.2022.v21i05.015

Open Access Journal

Prediction of Marine Water Quality Index Using a Stacked Classifier Under Machine Learning Architecture

K. Komathy*†
*Department of Information Technology, Academy of Maritime Education and Training, Chennai, India
†Corresponding author: K. Komathy; gomes1960@yahoo.com

Nat. Env. & Poll. Tech.
Website: www.neptjournal.com

Received: 28-04-2022
Revised: 25-06-2022
Accepted: 30-06-2022

Key Words:
Machine learning
Model training
Model optimizing
Ensemble learning
Marine water quality
Marine water quality index

ABSTRACT

The health of humankind is intrinsically associated with the health of the marine and ocean ecosystems. The pollution of the coastal region due to urbanization, for example, principally harms the growth of the ecosystem with poor-quality of water, which aggravates the survival of marine organisms and animals. The toxicity of the contaminated seafood would affect the human-ocean ecosystem thereby bringing down the economic rank of the region as well. Therefore, it is mandatory to assess the quality of the marine and ocean water to initiate any statutory measures to protect the regional marine water against pollution and dumping of toxic matter. This paper, therefore, presented an architecture of machine learning techniques to assist in classifying marine water quality. The proposed framework evaluated various classification models and selected the best fit out of the top-performing algorithms through training and optimizing. The finalized model was a *stacked classifier*, which was then deployed to predict the marine water quality index from the physicochemical and biological properties of the water.

INTRODUCTION

Physical, chemical, and biological processes undertake to transform a marine ecosystem into a complex one. Light, food, and nutrients are vital elements to a healthy marine ecosystem. Also, water temperature, depth, and salinity, as well as the local landscape have an impact on marine ecosystems. Variations of these conditions can result in new compositions of species as part of the marine community. Besides, a microbe present in the marine environment is an important biological indicator of water quality. Say, phytoplankton microbes constitute the basis of the food chain structure in marine environments, which serve as food for aquatic species such as zooplankton, shellfish, and finfish (Sridhar et al. 2006), and keep the biological balance of water quality. Marine microorganisms adapt quickly to changes in nutrients (Robertson & Blaber 1993). According to plant-animal relations showcased by Sverdrup et al. (1942), plants, as autotrophic organisms, transform inorganic material into organic particulate as food for marine animals. Therefore, it is vital phenomenon to monitor the quality of marine water so that the dependents including humankind in the food chain will be able to survive without infections and toxicity.

ORCID details of the authors:

K. Komathy:
https://orcid.org/0000-0002-0721-1762

The water quality index (WQI) is an empirical expression, as per McClelland (1974), which integrates physical, chemical, and microbiological parameters of water quality to derive into a single number. Since 1965, several numerical water quality indices have originated on various approaches depending on the applications and water parameters. Horton started with a simple WQI model (Horton 1965) followed by Brown et al. (1970, 1972) to propose a general WQI. WQIs were derived from associating observed parameter values and the local norms. Different WQI approaches existing currently were derived from either the weighted sum method or amplitude technique. The weighted-sum method was used to generate a sub-index, which helped to normalize the water quality parameters differing in units of measurement (Banda & Kumarasamy 2015). On the other method, the overall WQI was calculated using the deviations of water quality metrics quantitatively from the targets (CCME 2001, Khan et al. 2005, Mostafaei 2014). Akhtar et al. (2021) have elaborated the multi-criteria decision-making (MCDM) approach, which included the processes namely MACBETH and AHP to distinguish the various water usages such as drinking, domestic, irrigation, and industrial requirements. The process involved the steps: collecting the users' demand for each of the criteria to be selected; selecting alternatives to the high-ranked criteria; getting the consumers' ranking of the criteria and alternatives; and a comparison of criteria and their counterparts.

This paper has reviewed the salient features of novel techniques used for arriving at WQI that were proposed by research scientists and brought out a suitable computational model developed over the machine learning platform for classifying the marine water quality and predicting further, the marine water quality index (MWQI) of seawater.

REVIEW OF METHODOLOGIES FOR WATER QUALITY INDEX ESTIMATION

Horton's Model for WQI

In 1965, Horton developed and extended the concept of WQI (Horton 1965). Horton followed the steps for creating an index as (i) selected the water quality indicators such as dissolved oxygen, pH, E.coli, specific conductivity, alkalinity, and chloride, commonly available in Europe, Africa, and Asia countries; (ii) selected rating scales for each indicator so that the sub-index ranges from 0 to 100, where the highest quality was rated 100; and (iii) fixed the weight of the parameters that ranged from 1 to 4.

The Horton WQI model employed physicochemical indicators of water quality such as DO, pH, coliforms, specific conductance, carbon, chloroforms extract, alkalinity, and chlorides. The parameters were chosen based on their importance, the relative effect of other parameters, and the reliability of the data. Horton applied a linear scaling function and assigned sub-index values based on concentration level or contamination level. The sub-index value ranged from 0 to 100, where 0 indicated worst quality and 100 to excellence. The parameter weights were determined using the Delphi method (Taylor et al. 2003). The panel of experts was asked to assign a weight of 1 to 4 to various water quality parameters. Final WQI was obtained by aggregating an additive function given by Equation (1) as,

$$WQI = \left[\frac{\sum_{i=1}^{n} W_i S_i}{\sum_{i=1}^{n} W_i}\right] m_1 m_2 \qquad \ldots(1)$$

where S_i is the sub-index of the i^{th} variable; W_i is the relative weight of the i^{th} variable; m_1 is a temperature correction factor, which is equal to 0.5 for temperatures below 34°C; and otherwise, it is 1; m_2 is a pollution correction factor equal to 0.5 or 1. Selecting the right parameters for WQI was one of the challenges in the Horton model.

National Sanitation Foundation Water Quality Index (NSFWQI) Method

NSFWQI used the weighted arithmetic mean function to assess the significance of all water quality parameters obtained from 142 water quality assessment experts. To convert the score to weights, the most important parameter was assigned a temporary weight of 1, and the weights of all other parameters were derived by dividing the weight of the highest parameter by the average of the individual importance scores. The final weight for each parameter was obtained by dividing the individual temporary weights by the sum of the temporary weights so that the final sum of the weights, $\sum_{i=1}^{n} W_i = 1$. NSFWQI was estimated using Equation (2) as,

$$Additive\ quality\ index, WQI = \sum_{i=1}^{n} W_i\, Q_i \qquad \ldots(2)$$

where W_i is the unit weight of the i^{th} parameter; Q_i is the sub-index of the quality parameter derived from the parameter-to-value conversion curve in the interval of 0 to 100, and n is the number of parameters. Based on the WQI values obtained in the range of 0 to 100, quality classes would then be assigned as *excellent, good, fair, and bid* accordingly. McClelland (1974) found that though the additive model is a good choice when all the parameters are within a reasonable range it lacks the sensitivity of a low-impacted parameter on the overall water quality. Hence, a multiplicative model was originated (McClelland 1974), which is given by Equation (3),

$$Multiplicative\ quality\ index, WQI = \prod_{i=1}^{n} Q_i^{\,w_i} \qquad \ldots(3)$$

where Q_i is the quality of the i^{th} parameter, a value between 0 to 100; W_i is the unit weight of the i^{th} parameter, and n is the number of parameters.

Oregon Water Quality Index (OWQI) Method

Originally developed in 1970 by Brown et al. (Brown 1970, Brown 1972, Cude 2001), OWQI had been improved in subsequent years after understanding more about the water quality behavior. The OWQI, a statistical tool is used to analyze a defined set of water quality parameters and generate a rating that describes the overall water quality of a particular monitoring site. The OWQI score takes a value from 10 to 100. The OWQI is also helpful for assessing water quality for recreational purposes like fishing and swimming. The OWQI approach utilized the arithmetic mean and is given by Equation (4),

$$OWQI = \sqrt{\frac{n}{\sum_{i=1}^{n} \frac{1}{SI_i^{\,2}}}} \qquad \ldots(4)$$

where n is the number of parameters, and SI_i is the sub-index of the i^{th} parameter (Darvishi et al. 2016). The weighted harmonic helped the impaired parameter to influence the OWQI by its presence and also observed that OWQI was dependent on the environmental changes hence their impacts on water quality were considered. However, all the toxic elements for health such as bacteria, metals, and toxins (Tyagi et al. 2013) were not part of it.

Weighted Arithmetic Water Quality Index (WAWQI) Method

WAWQI's strategy classified water quality from the most influential water quality indicators. It is a simple and easy technique to utilize. It is observed that most scientists have utilized the WAWQI to decide the water quality of surface waters and groundwater. The algorithm of WAWQI is given below:

- Identify the water quality indicators for WQI calculation
- Calculate the *proportionality constant, k* value from, $k = \frac{1}{\sum_{i=1}^{n} \frac{1}{S_i}}$ where S_i is the standard value for the i^{th} parameter and n is the number of parameters considered.
- Q_i, the quality rating for the i^{th} parameter is estimated by, $Q_i = 100 * \{\frac{V_{oi} - V_i}{V_{Si} - V_i}\}$ where V_{oi} is the observed value from the sample for the i^{th} parameter; Vi is the ideal value of the it^{h} parameter in distilled water; V_{si} is the standard or permissible value for the i^{th} parameter
- W_i, the unit weight of the individual parameter is derived from, $W_i = k/S_i$
- Finally, the water quality index (WQI) is arrived at by Equation (5) as,

$$WQI = ((\sum_{i=1}^{n} W_i Q_i) / \sum_{i=1}^{n} W_i) \quad \ldots(5)$$

The authors of (Chandra et al. 2017) have studied the groundwater quality of Vijayawada, India using the WAWQI model for analyzing the physicochemical properties such as pH, total dissolved solids (TDS), Cl, SO4, Na, K, Ca, Mg, and total hardness (TH) at nineteen different stations of the study area. Oni et al. (2016) studied the impact of municipal solid waste on the quality of the groundwater and surface water bodies at Ado Ekiti, Nigeria by using the algorithm of WAWQI. However, this index did not include all the parameters, which could describe the quality of a water source and this index only quantified the direct impact of physiochemical indicators (Tyagi et al. 2013)

Canadian Council of Ministers of the Environment Water Quality Index (CCME WQI) Method

CCME WQI (CWQI 2022, CCME 2001) is implemented generally to test a multi-boundary water quality against the permissible limit kept by the user. The WQI considers three variables such as scope, frequency, and amplitude to produce the water quality at that particular location with the benchmarks chosen. The output is a number ranging from 0 to 100, where a score of 100 indicates that all variables meet the selected benchmarks. F1 represents the scope variable in terms of failed parameters, which does not satisfy the benchmark at least once, and Equation (6) is arrived as,

$$F_1 = \frac{No.\ of\ failed\ parameters}{Total\ no.\ of\ parameters} * 100 \quad \ldots(6)$$

F_2 indicates the frequency variable, in terms of failed tests, which do not meet the target as given in Equation (7),

$$F_2 = \frac{No\ of\ failed\ tests}{Total\ no.\ of\ tests} * 100 \quad \ldots(7)$$

F_3 refers to the amplitude of the test values, by which they did not meet the benchmark. The non-compliance with the test values is calculated as:

(i) The test value should not exceed the target: $excursion_i = \left\{\frac{Failed\ Test\ Value}{Target}\right\} - 1.0$

(ii) The test value should not fall below the target: $excursion_i = \left\{\frac{Target}{Failed\ Test\ value}\right\} - 1.0$

Then, the sum of excursions under the non-compliance category is calculated as, *Normalized sum of excursions* $(nse) = \frac{\sum_{i=1}^{n} excursion_i}{No.of\ tests}$

F_3 is then calculated as given in Equation (8),

$$F_3 = \left\{\frac{nse}{0.01nse + 0.01}\right\} \quad \ldots(8)$$

CCME WQI method suits to calculate the quality index for different water bodies and it does not get influenced by the presence of missing data. The limitation of the method stated by the authors (Călmuc et al. 2018) is that physiochemical indicators included in the estimation of WQI are selected in such a way that they do have the same level of influence. The water quality estimated is also partial since the other physiochemical indicators and biological indicators are not considered in the algorithm. Lastly, the F_1 variable is not consistent when the quality indicator list has only very few parameters

Bhargava Method

The Bhargava approach considered the parameters corresponding to each of the applications separately and applied the sensitivity function that selects a value between 0 and 1 and the results are accumulated using the geometric mean

(Bhargava 1983a, 1983b, Noori et al. 2017). The geometric mean-based WQI suggested by Bhargava is expressed in Equation (9) as,

$$WQI = \left(\prod_{i=1}^{n} f_i()\right)^{1/n} * 100 \quad \ldots(9)$$

where function f_i () indicates the sensitivity function for the i^{th} parameter, which is related to the weighting of the i^{th} parameter and varies from 0 to 1; and n is the number of water quality parameters considered. Water quality index (WQI), used six water quality parameters namely dissolved oxygen (DO), biochemical oxygen Demand (BOD), most probable number (MPN), turbidity, total dissolved solids (TDS), and pH measured at eight different stations along the river basin. Rating curves were drawn based on the quality of river water and gave importance with weight to every parameter. Lastly, a rating scale in the order of 0 to 4 was used to classify water quality in each of the study areas such as, *excellent* to *poor* range by the Bhargava WQI method.

Recreational Water Quality Index (RWQI) Method

Recreational water bodies are characterized and classified to determine their suitability for recreational use, based on susceptibility to fecal contamination, development of HABs, and proliferation of specific free-living microbial pathogens (WHO 2003). The water quality criterion is to set a recreational water quality index for fresh bathing waters. The World Health Organization (WHO) aggressively promoted the safety of human health while using recreational waters. In (WHO 2003, WHO 2005, WHO 2021) reports, WHO laid out the guidelines that provide an assessment of the health risks associated with the recreational use of water. Risks arising from contacting the contaminated water include infection from microbes and toxicity of physical and chemical properties of the water.

The RWQI was developed by Almeida et al. (2012) using physical, chemical, and microbiological parameters. The opinions of 17 experts were used from the rating curves and all rating curves were fitted with a coefficient regression bigger than 0.98 for each one. WHO has fixed a maximum value (Almeida et al. 2012) for recreational waters as 2 $mg.L^{-1}$ for Q_i, which is equivalent to 100; with 75 for values up to 0.1 $mg.L^{-1}$; and the Q_i reaches10 for values above 2 $mg.L^{-1}$. The RWQI included the water quality indicators namely pH, turbidity, detergents, nitrate, COD, PO_4^{3-}, total coliforms, coliforms, and enterococci. Equation (10) shows the calculation of RWQI as,

$$RWQI = \prod_{i=1}^{n} {Q_i}^{W_i} \quad \ldots(10)$$

where Q_i is the rating value of the i^{th} parameter expressed as concentration or other analytical measurements and W_i is the weightage given to the i^{th} parameter in such a way, $\Sigma W_i = 1$. W_i is calculated as, $W_i = \frac{\frac{1}{a_i}}{\Sigma\frac{1}{a_i}}$, where a_i is the coefficient that varies from 1 to 4 as per the significance of the parameter i. RWQI is a result ranging from 0 to 100.

THE FRAMEWORK OF MACHINE LEARNING TECHNIQUES FOR THE CLASSIFICATION AND PREDICTION OF MARINE WQI

This Section discusses the design of a framework of machine learning techniques for classifying and further predicting marine water quality index (MWQI) using eight water quality parameters like salinity, dissolved oxygen (DO), dissolved inorganic carbon (DIC), alkalinity, phosphate, nitrate+nitrite, dissolved organic carbon (DOC), and heterotrophic bacteria from the dataset collected by the Center for Microbial Oceanography Research & Education Data System (C-MORE DS 2021). The framework has embedded the machine learning platform to predict marine water quality. In this framework, different supervised machine-learning models were examined to predict water quality more accurately. Fig.1 illustrates the process flow for arriving at the MWQI. The observations of physiochemical and biological parameters collected from 11 cruises were

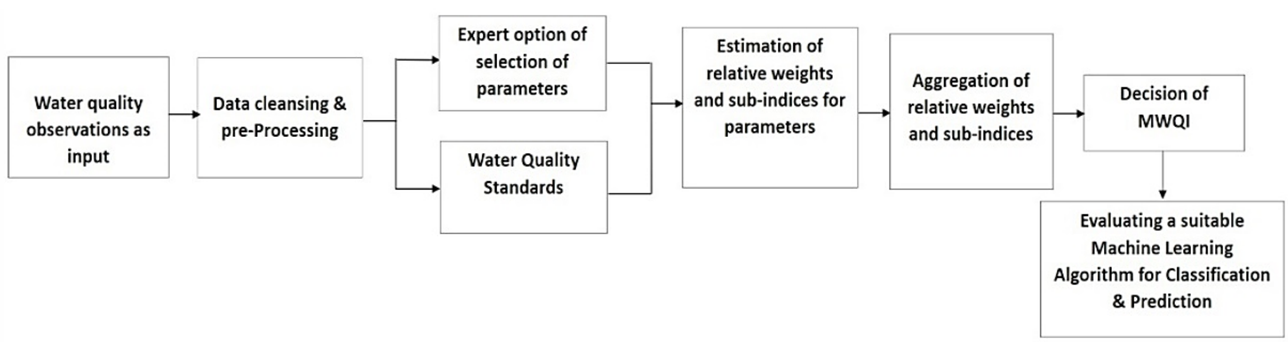

Fig. 1: Process flow of arriving MWQI from the input observations.

used in this study. The data pre-processing phase involved data cleansing using outlier detection, missing data, and normalization. After pre-processing, the dataset taken for the study contained 1361 observations. The weighted arithmetic water quality index (WAWQI) algorithm discussed in Section 2 (iv) was embedded in this framework. The parameters used for estimating MWQI have been thoroughly explored with their significance to marine water quality.

Correlation analysis is to determine the association among water quality parameters involved in the study. Correlation analysis as given in Fig. 2 helps to conclude the positive, negative, or no-correlation effects of those chosen parameters. Table 1 gives the statistic of input parameter values and the outcome of the correlation. The standard deviation shows the average distance of each observation from the mean. A larger value indicates that the data are more spread out. Salinity, for example, showed a low variation, which made 25%, 50%, and 75% of the observations stick around the mean value. Looking at the minimum and maximum range of the DO, DIC, Alkalinity, DOC, and h-bacteria, the standard deviations also proved that the data are widely spread out. The parameters such as DIC, alkalinity, phosphate, nitrate +nitrite, and DOC showed the minimum value and 25%, 50%, and 75% observations as zero, which was reflected in their correlation too. The coefficient of variation (CV) and the index of dispersion (ID) displayed the relative dispersion of the observations around the mean.

The correlation matrix given in Table 2 shows the degree of linear correlation by means of coefficient R between any two parameters. R-value helps to identify the water quality parameters, which may influence the seawater quality. Some of them exhibit negative correlations too. Salinity, DO, DIC, nitrate+nitrite, phosphate, alkalinity, DOC, and h-bacteria from the given dataset were found to be more significant in the water quality. Marine water quality standards of those chosen parameters were also considered for estimating the relative weights and sub-indices and aggregating these two computed outcomes. The estimated value of water quality of each observation was found to be lying in the range of 0 to 100. The water quality results were then classified as *excellent, good, and poor* represented numerically as classes 2, 1, and 0 respectively. The classes that arrived are therefore called *marine water quality index (MWQI)*.

The framework has further employed the machine learning model to classify and predict MWQI, which is shown in Fig. 2. Lastly, the correlation of each parameter with MWQI was studied to understand the linear association between each parameter with MWQI and is represented in Table 3. Determination of the best-fit classification model for the marine water quality adapted training, evaluation, and optimization as discussed in the following sections.

Training and Optimizing the Classification Models

Process flow to determine the best-fit machine learning algorithm for predicting the MWQI has followed: (i) create models, (ii) tune models, (iii) compare models, and (iv) ensemble model. An integrated software development tool, namely, the Anaconda platform unified with python programming (Python 2020) and PyCaret (PyCaret 2022) was used in the model evaluation and optimization. The dataset (C-MORE DS 2021) after pre-processing was split

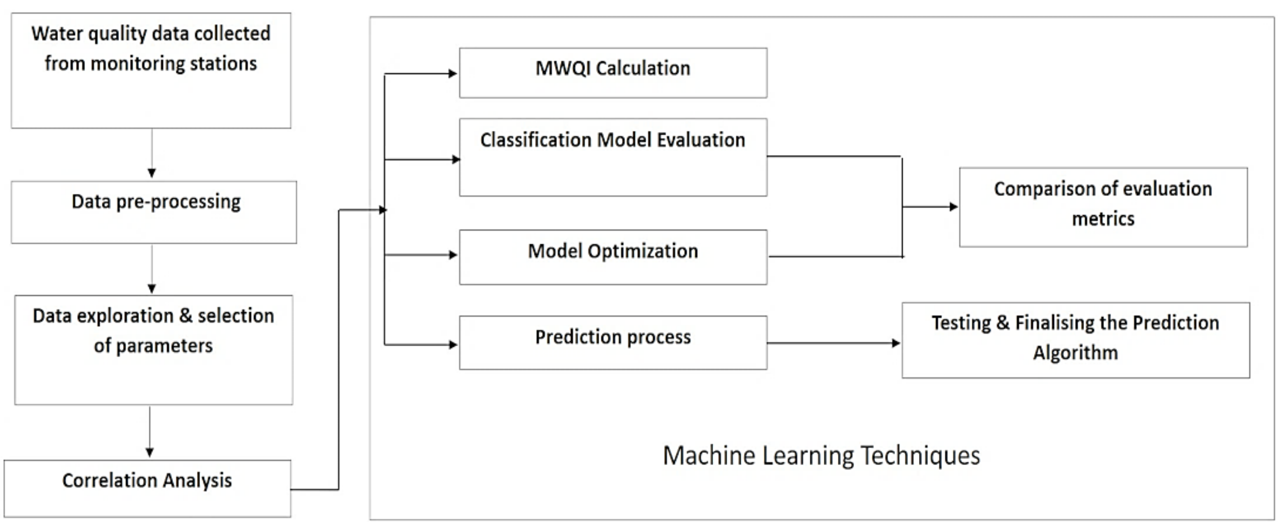

Fig. 2: Framework of machine learning techniques for classifying and predicting MWQI.

Table 1: Descriptive statistics to analyse the input parameters for prediction.

Descriptive Statistics	Salinity $\mu mol.kg^{-1}$	DO $\mu mol.kg^{-1}$	DIC $\mu mol.kg^{-1}$	Alkalinity $\mu mol.kg^{-1}$	Phos $\mu mol.kg^{-1}$	Nitrate +Nitrite $\mu mol.kg^{-1}$	DOC $\mu mol.kg^{-1}$	H-Bact $CFU.100mL^{-1}$	Water quality %
Observations	1361	1361	1361	1361	1361	1361	1361	1361	1361
Mean	35.12	200.86	437.43	498.65	0.086	0.353	3.563	455.80	59.25
Std dev	0.253	40.11	837.76	954.50	0.274	1.946	15.46	188.68	6.25
Min	34.039	0.0	0.0	0.0	0.0	0.0	0.0	123.70	50.47
25%	34.95	204.10	0.0	0.0	0.0	0.0	0.0	337.40	55.62
50%	35.203	208.20	0.0	0.0	0.0	0.0	0.0	441.10	57.68
75%	35.293	213.60	0.0	0.00	0.10	0.07	0.0	528.90	60.65
Max	35.988	265.90	2328.1	2409.0	2.71	29.64	83.50	2469.7	100.0
CV	0.007	0.199	1.915	1.914	3.175	5.519	4.339	0.4139	0.105
ID	0.002	8.011	1604.4	1827.07	0.870	10.74	67.11	78.11	0.659

Table 2: Correlation matrix of marine water quality indicators.

	Salinity	DO	DIC	Alkalinity	Phos	Nit+Nit	DOC	H-bact
Salinity	1.0000							
DO	0.3689	1.0000						
DIC	0.0280	0.0204	1.0000					
Alkalinity	0.0320	0.0308	0.9994	1.0000				
Phos	-0.2338	-0.3443	0.3305	0.3050	1.0000			
Nit+Nit	-0.2276	-0.3943	0.2107	0.1855	0.9410	1.0000		
DOC	-0.1097	0.0554	-0.1230	-0.1231	-0.0688	-0.0383	1.0000	
H-Bact	-0.3420	-0.0166	0.1131	0.1109	0.2651	0.1355	-0.0496	1.0000

into training, validation, and test data with eight water quality parameters. Training and validation data were used to evaluate the classification models under the machine learning framework. Training the classification models is to select an optimum one in classifying marine water quality. PyCaret is a python-based machine learning library, that assists the framework for fitting the best model. To optimize the selection further, tuning the hyperparameters of the individual model was also added. The performance metrics namely, accuracy, AUC, recall, precision, F1, kappa, MCC, and TT helped to appraise the top-performing models as demonstrated in Table 4. From the comparison, it is found that the decision tree algorithm was the best-fit one. Additionally, ensemble and stacking processes would also help to identify a more suitable algorithm.

Table 3: Correlation of MWQI with each Water Quality Indicator.

MWQI	1.000
Hbact	0.646
Phos	0.562
Nit	0.525
Doc	0.264
Dic	0.135
alk	0.122
Csal	-0.478
Coxy	-0.661

Ensemble Learning Techniques in Prediction

Machine learning models sometimes favor biasing and therefore, ensemble learning is preferred to recover from biasing. An ensemble is a machine-learning approach, in which several machine-learning algorithms are bundled together and trained. The top four classification methods from Table 4 namely, decision tree, ada boost, gradient boosting, and random forest were tuned first and then ensembled using bagging or boosting algorithms. Stacking was finally applied to combine the four ensembled algorithms. Evaluation of the stacked classifier by their performance was appraised as portrayed in Table 5. The finalized prediction algorithm was saved as *MWQI Classifier* in the server where the algorithm could be deployed and worked stand-alone thereon.

Table 4: Comparison of the performance of classification models.

Classifier Model	Accuracy	AUC	Recall	Precision	F1	Kappa	MCC	TT (Sec)
Decision tree	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.0110
Ada boost	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.0120
Gradient boosting	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.0830
Random forest	0.9987	1.0000	0.9986	1.0000	0.9993	0.9850	0.9860	0.2040
Extra trees	0.9974	0.9995	0.9973	1.0000	0.9986	0.9701	0.9720	0.1700
Logistic regression	0.9961	1.0000	0.9986	0.9973	0.9979	0.9494	0.9531	0.0490
Light gradient boosting machine	0.9948	1.0000	0.9973	0.9973	0.9973	0.9345	0.9391	0.0310
SVM - linear kernel	0.9935	0.0000	0.9959	0.9973	0.9966	0.9252	0.9301	0.0140
Ridge classifier	0.9935	0.0000	0.9986	0.9946	0.9966	0.9041	0.9150	0.0150
Linear discriminant analysis	0.9935	1.0000	0.9973	0.9959	0.9966	0.9195	0.9251	0.0110
K Neighbors	0.9895	0.9691	0.9973	0.9919	0.9945	0.8453	0.8611	0.0280
Naive Bayes	0.9777	0.9986	0.9767	1.0000	0.9881	0.8031	0.8246	0.0100
Dummy classifier	0.9580	0.5000	1.0000	0.9580	0.9786	0.0000	0.0000	0.0100
Quadratic discriminant analysis	0.8777	0.8787	0.8849	0.9881	0.9181	0.4815	0.5397	0.0120

Table 5: Performance of the Stacked Classifier Model Finalized for Marine Water Quality Prediction.

Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC
Stacking Classifier	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Testing the Prediction Algorithm

MWQI classifier was loaded in the integrated machine learning platform of Anaconda, Python, and PyCaret and tested the sample data accessed from the given water quality dataset. Table 6 displays a sample output from the prediction. The outcome of the prediction was 100% accurate.

CONCLUSION

This paper proposed a machine learning framework for evaluating the classification models, optimizing the models, and finally selecting the best-fit model with higher performance for the given dataset and application. The chosen model was then applied for prediction. The classification method called *stacked classifier* was proposed as a top-performing machine learning model under the supervised learning category to classify the marine water quality. The application dataset from C-MORE DS with eight selected physio, chemical, and biological properties are involved to train, evaluate and test the model for arriving at the marine water quality index (MWQI). The estimated and predicted values of MWQI matched with 100% accuracy.

ACKNOWLEDGMENT

This study was granted by the Department of Science and Technology (DST), India under optimal water usage for the industrial sector (OWUIS).

Table 6: A sample of predicted output.

Salinity $\mu mol.kg^{-1}$	DO $\mu mol.kg^{-1}$	DIC $\mu mol.kg^{-1}$	Alkalinity $\mu mol.kg^{-1}$	Phos $\mu mol.kg^{-1}$	Nitrate +Nitrite $\mu mol.kg^{-1}$	DOC $\mu mol.kg^{-1}$	H-Bacteria $CFU.100mL^{-1}$	MWQI calculated	MWQI predicted	Score
34.578	211.2	0	1.33	6.75	0	1543.9	85.307	0	0	1
34.94	214.1	0	0	0	0	538.9	58.658	1	1	1
35.289	211.2	0	0	0	0	665.5	62.060	1	1	1
34.854	265.8	0	0.77	0	0	1540.1	75.588	0	0	1
35.175	210.7	0	0.12	0.01	0	575.1	60.025	1	1	1
35.195	212.2	0	0	0	0	140.8	50.468	0	1	1

REFERENCES

Akhtar, N., Ishak, M. I. S., Ahmad, M. I., Umar, K., Md Yusuff, M. S., Anees, M. T., Qadir, A. and Ali Almanasir, Y.K. 2021. Modification of the water quality index (WQI) process for simple calculation using the multi-criteria decision-making (MCDM) method: A review. Water, 13(7): 905. https://doi.org/ 10.3390/w13070905

Almeida, C., González, S.O., Mallea, M. and González, P. 2012. A recreational water quality index using chemical, physical and microbiological parameters. Environ, Sci. Pollut. Res., 19(8): 3400-3411. https://doi.org/10.1007/s11356-012-0865-5

Banda, T.D. and Kumarasamy, M.A. 2020. Development of water quality indices (WQIs): A review. Pol. J. Environ. Stud., 29(3): 2011-2021. https://doi.org/10.15244/pjoes/110526

Bhargava, D.S. 1983a. Most rapid BOD assimilation in Ganga and Yamuna rivers. J. Environ. Eng., 109(1): 174-188. https://doi.org/10.1061/(ASCE)0733-9372(1983)109:1(174).

Bhargava, D.S. 1983b. Use of water quality index for river classification and zoning of Ganga River. Environ. Pollut. Ser. B. Chem. Phys., 6(1): 51-67. https://doi.org/10.1016/0143-148X(83)90029-0

Brown, R.M., McClelland, N.I., Deininger, R.A. and O'Connor, M.F. 1972. A water quality index—crashing the psychological barrier. Environ. Qual., 15: 173-182. https://doi.org/10.1007/978-1-4684-2856-8_15

Brown, R.M., McClelland, N.I., Deininger, R.A. and Tozer, R.G. 1970. A water quality index-do we dare. Water Sew. Works, 117(10): 339-343.

Călmuc, V.A., Călmuc, M., Topa, M.C., Timofti, M., Iticescu, C. and Georgescu, L.P. 2018. Various methods for calculating the water quality index. Ann. Dunarea De Jos Univ. Galati. Fascicle II Math. Phys. Theor. Mech, 41(1): 171-178. https://doi.org/10.35219/ann-ugal-math-phys-mec.2018.2.09.

Canadian Water Quality Index (CWQI) 2022. Accessed on 11th February 2022. https://www.gov.nl.ca/ecc/waterres/quality/background/cwqi/

CCME 2001. Canadian water quality guidelines for the protection of aquatic life: CCME Water Quality Index1.0 Technical Report. Canadian Council of Ministers of the Environment (CCME). Accessed on 10th Feb 2022. https://prrd.bc.ca/wp-content/uploads/post/prrd-water-quality-database-and-analysis/WQI-Technical-Report-en.pdf

Center for Microbial Oceanography: Research & Education Data System (C-MORE DS) 2021. Data Accessed on 7thSeptember 2021. https://hahana.soest.hawaii.edu/cmoreserver/datasearch/data.php

Chandra, D.S., Asadi, S.S. and Raju, M.V.S. 2017. Estimation of water quality index by weighted arithmetic water quality index method: a model study. Int. J. Civil Eng. Tech., 8(4): 1215-1222.

Cude, C.G. 2001. Oregon water quality index is a tool for evaluating water quality management effectiveness 1. JAWRA J. Am. Water Resour. Assoc., 37(1): 125-137. https://doi.org/10.1111/j.1752-1688.2001.tb05480.x

Darvishi, G., Kootenaei, F. G., Ramezani, M., Lotfi, E. and Asgharnia, H. 2016. Comparative investigation of river water quality by OWQI, NSFWQI, and Wilcox indexes: Case study: The Talar River–Iran). Arch. Environ. Protect., 42(1): 41-48. https://doi.org/10.1515/aep-2016-0005

Horton, R.K. 1965. An index number system for rating water quality. J. Water Pollut. Contr. Fed., 37(3): 300-306.

Khan, A. A., Tobin, A., Paterson, R., Khan, H. and Warren, R. 2005. Application of CCME procedures for deriving site-specific water quality guidelines for the CCME Water Quality Index. Water Quality Research Journal, 40(4): 448-456.

McClelland, N.I. 1974. Water quality index application in the Kansas River basin. U.S. Environmental Protection Agency Region 7, Kansas City, Missouri. Accessed on 4th February 2022. National Service Center for Environmental Publications (NSCEP). https://www.epa.gov/nscep

Mostafaei, A. 2014. Application of multivariate statistical methods and water-quality index to the evaluation of water quality in the Kashkan River. Environ. Manag., 53(4): 865-881. https://doi.org/10.1007/s00267-014-0238-6

Noori, M.M., Abdulrazzaq, K.A. and Mohammed, A.H. 2017. Evaluation of water quality using Bhargava water quality index method and GIS, case study: Euphrates river in Al-Najaf City. Int. J. Sci. Res., 6(7): 1286-1295. https://doi.org/10.21275/art20175545

PyCaret Official. 2022. Accessed on 2nd January 2022. https://pycaret.gitbook.io/docs/get-started/functions.

Python for Windows. 2020. Accessed on 12th June 2020. https://www.python.org/downloads/windows/

Robertson, A.I., and Blaber, S.J.M. 1992. Plankton, epibenthos, and fish communities. Coast. Estuar. Stud., 29: 173-224. https://doi.org/10.1029/CE041p0173

Sridhar, R., Thangaradjou, T., Kumar, S.S. and Kannan, L. 2006. Water quality and phytoplankton characteristics in the Palk Bay, southeast coast of India. J. Environ. Biol., 27(3): 561-566.

Sverdrup, H.U., Johnson, M.W. and Fleming, R.H. 1942. The Oceans: Their Physics, Chemistry, and General Biology. Prentice-Hall Inc., New York.

Taylor, J.G. and Ryder, S.D. 2003. Use of the Delphi method in resolving complex water resources issues. JAWRA J. Am. Water Resour. Assoc., 39(1): 183-189. https://doi.org/10.1111/j.1752-1688.2003.tb01570.x

Tyagi, S., Sharma, B., Singh, P. and Dobhal, R. 2013. Water quality assessment in terms of water quality index. Am. J. Water Resour., 1(3): 34-38. https://doi.org/10.12691/ajwr-1-3-3

World Health Organization (WHO). 2003. Guidelines for Safe Recreational Water Environments. Volume 1, Coastal and Fresh Waters. https://apps.who.int/iris/handle/10665/42591

World Health Organization (WHO). 2005. Guidelines for Safe Recreational Water Environments. Volume 2: Swimming Pools and Similar environments. https://www.who.int/publications/i/item/9241546808

World Health Organization (WHO). 2021. Guidelines on Recreational Water Quality. Volume 1: Coastal and Freshwaters. https://www.who.int/publications/i/item/9789240031302