International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
***Article Received:** 05 July 2023 **Revised:** 25 August 2023 **Accepted:** 15 September 2023*

Analyze Predict and Classify Water Quality and Usage of Water using Machine Learning Techniques

Roshini L[1], Dr. C.R.K. Reddy[2]
[1]Research Scholar, Department of Computer Science and Engineering, University College of Engineering, Osmania University
Assistant Professor in Department of ET, CVR College of Engineering
Hyderabad, India
roshini1604@gmail.com
Orchid: 000-0001-9709-0289
[2]Professor and Head, Department of Computer Science & Engineering,
MGIT, Hyderabad, India-500044
crkreddy_cse@mgit.ac.in
Orchid : 0000-0003-2246-3930

Abstract- As important as water is for humans, it is also crucial for all livestock and crops. Direct groundwater consumption by crops and livestock can affect both, perhaps causing crop failure or sickness in livestock if the quality is subpar or becomes unusable. Knowing whether the groundwater is usable will allow for the proper usage of the water. Specific crops that can survive that water quality can be grown by farmers. The main objective of this paper is to determine where and how the water can be used while also classifying the water's quality into one of several classifications. data is collected from open source of Telangana ground water quality data 2020. Water quality is identified and assessed with it target parameters as WQI, Classification 1, RSC, TDS, classification. WQI gives the one value for n number of parameters the water and its usage is assessed with its grades as good, moderate, very good and poor whereas Classification 1 is assessed with 2 values as mineral rich MR and poor safe PS, RSC When RSC usage surpasses the permitted limit, irrigation suffers (>2.5). TDS as the target variable assess the salinity of water which assessed with grades. Classification is assessed with 9 types of parameters.

Keywords-Water Purity, water quality indicator, Classification 1, TDS, RSC.

I. INTRODUCTION

In this paper we present an effective predictive classification model that can accurately classify water quality based on physicochemical parameters. The primary aim is to accurately predict and classify the quality of water samples from different locations using machine learning techniques. This model has the potential to contribute to better water management practices and safeguarding the quality of water resources. Groundwater, a vital natural resource, plays an indispensable role in sustaining life on our planet. It constitutes a significant portion of the Earth's freshwater supply, serving as a primary source of drinking water for billions of people worldwide and supporting various ecosystems. As society's demands for clean and reliable water continue to escalate due to population growth, urbanization, and industrialization, the quality of groundwater becomes a subject of paramount importance. Understanding and safeguarding groundwater quality is crucial not only for human health but also for the ecological integrity of aquatic environments.

II. MOTIVATION

Water is essential for all the crops and livestock's as much as it is important for human beings. Crops and livestock consume direct ground water, and if the quality is not up to the mark, or becomes unusable, then crops and livestock are affected, which may lead to crop failure or livestock's developing diseases. By knowing the quality of the ground water, whether usable or not, the water can be put to appropriate use. Farmers can grow specific crops which can tolerate that quality of water. The disease occurred to animals and crop failure due to impure water cause economical damage to the nation and physical damage to human has motivated to predict and classify water and to find the purpose where the water can use.

III. PROBLEM STATEMENT

Water is an important source of factor to stay healthy, nowadays water got polluted everywhere on the earth, Natural factors that influence water quality are hydrological, atmospheric, climatic, topographical, and lithological factors (Magesh et al., 2013; Uddinet al., 2018). There are many sources of water contamination which pollute water, Therefore, predicting and monitoring water quality is vital for ensuring the availability of clean and safe water for all [2]. Water quality prediction models can help in identifying potential threats to water quality, predicting future changes in water quality parameters, and designing effective strategies to manage and improve water quality. These models utilize various statistical and machine learning techniques to analyze historical data and identify patterns and relationships between different water quality parameters [2]. Safe drinking water is essential in human life drinking bad water harms our inner parts of the body.so we should drink neat and clean water. The quality of irrigation water plays a vital role on humans as what we eat is produced from the crops when they are irrigated. Irrigation water criteria and assessment:

Water quality is assessed with its parameters such as salinity and the salinity effects TDS in water for crops if TDS is very high then the water intake in root zone will be very less and so crops won't grow and yield.

The dataset we took consists of various water parameters with its limits and classified as target variables like Classification, Classification 1, RSC and these classifications are defined with its purpose and usage. So, there is a need to classify the purpose of water and define the quality of water; whether usable or not, the water can be put to appropriate use. Farmers can grow specific crops which can tolerate that quality of water.

Therefore, the problem statement is to develop a machine learning model that can predict water quality accurately and suggest the purpose of its usage, thus ensuring the sustainability of water resources and improving the quality of life for millions of people worldwide. This research work will address the root causes of water pollution and provide insights into the combined impact of biophysical and environmental factors, enabling effective monitoring and prediction of water quality.

IV. QUALITY OF WATER

A. Predict and classify the quality of water.

This model has the potential to contribute to better water management practices and safeguarding the quality of water resources.

1. **Data Collection**: This process involves the collection of water quality dataset from various districts, mandals, and villages. The water dataset contains parameters such as pH, E.C, TDS, CO3, bicarbonate (HCO3), chloride (Cl), fluoride (F), nitrate (NO3), sulfate (SO4), sodium (Na), potassium (K), calcium (Ca), magnesium (Mg), total hardness, sodium adsorption ratio (SAR), and more. These parameters provide insights into the chemical composition of the water samples.
2. **Classification Target**: The research focuses on the predicting and classifying water quality samples distinct classes based on established water quality standards. The classification involves two key aspects: "Classification" and "RSC Classification" (Residual Sodium Carbonate). These class labels are indicative of water quality levels, such as "C3S1" and "P.S." Each class label represents a specific quality category, allowing for the differentiation of water samples based on their chemical characteristics.
3. **Feature Analysis:** The study involves a thorough analysis of the dataset's features to understand their individual and combined effects on water quality. This analysis helps identify which parameters are most influential in determining water quality classifications.
4. **Machine Learning Model Development**: The core objective is to build a predictive classification model using machine learning algorithms. The model will be trained on the dataset, utilizing the various physicochemical parameters as input features and the provided water quality classifications as target labels.
5. **Performance Evaluation:** Accuracy, precision, recall, and F1-score are some of the acceptable performance measures that will be used to assess the constructed classification model. These measures evaluate how well the algorithm can assign water samples to the appropriate quality groups.

B. Dataset

Data collected:

Data is collected from Telangana Open Data portal,
Telangana State, India.
This data contains samples tested from various districts.

There are 3 files, each, or year 2018, 2019 and 2020 contains post-monsoon season groundwater.

Quality details are:

ground_water_quality_2020_post.csv can be combinedly used.

Each dataset contains 26 columns such as:

serial num (sno), District, Mandal, Village, Lattitude, Longitude, Chemicals (such as Ca, Mg, CO3 etc), Total Hardness of the water, Total dissolved solids, RSC, SAR, and the target variables 'Classification' and 'Classification1'.

sno	district	mandal	village	temp_id	long_gis	lat_gis	gwl	season	pH	E.C	TDS	CO3	HCO3	Cl	F	NO3	SO4	Na	K	Ca	Mg	T.H	SAR	Classification	RSC meq/L	Classification.1
1	ADILAB/	Adilabad	Adilabad	1001	78.525	19.668	14	Pre-monsoon 2020	7.8	1671	1069	0	470	230	0.5	3.42241	31.5	154	13	72	77.79	499.868	2.99478	C3S1	-0.5974	P.S.
2	ADILAB/	Bazarhatn	Bazarhatr	1002	78.351	19.459	9.4	Pre-monsoon 2020	8	545	348.8	0	180	80	1.1	4.22768	13	94	12	16	14.59	99.9753	4.08746	C2S1	1.6005	MR
3	ADILAB/	Gudihatno	Gudihatn	1007	78.512	19.526	18	Pre-monsoon 2020	8	738	472	0	220	30	0.5	28.3859	9.5	22	1	40	38.9	259.934	0.59328	C2S1	-0.7987	P.S.
4	ADILAB/	Jainath	Jainath	1009	78.64	19.731	5.3	Pre-monsoon 2020	8	1154	739	0	300	100	0.7	140.923	14.8	81	3	56	68.07	419.885	1.71867	C3S1	-2.3977	P.S.
5	ADILAB/	Narnoor	Narnoor	1010	78.853	19.496	4.5	Pre-monsoon 2020	7	1042	667	0	370	30	0.7	163.068	10.3	39	4	96	63.21	499.893	0.7584	C3S1	-2.5979	P.S.
6	ADILAB/	Neradigon	Neradigo	1011	78.412	19.294	9.9	Pre-monsoon 2020	8	1063	680	0	300	100	0.6	62.4086	14.5	65	14	64	53.48	379.91	1.44992	C3S1	-1.5982	P.S.
7	ADILAB/	Talamadu	Talamadu	1013	78.397	19.633	8.1	Pre-monsoon 2020	8	788	504	0	310	40	0.4	42.0755	10.5	59	4	40	43.76	279.926	1.53321	C3S1	0.6015	P.S.
8	ADILAB/	Tamsi	Tamsi	1014	78.427	19.681	6.3	Pre-monsoon 2020	8	811	519	20	200	110	0.7	11.4751	16	144	1	16	14.59	99.9753	6.26164	C3S1	2.4005	MR
9	ADILAB/	Utnoor	Utnoor	1015	78.769	19.379	6.6	Pre-monsoon 2020	8	422	270	0	160	30	0.5	13.4883	8.25	41	2	24	19.45	139.967	1.50676	C2S1	0.4007	P.S.

Fig 1: Dataset with relevant columns for water parameters and target variables

C. Water Quality:

To assess water quality, there are three types of parameters: physical, chemical, and biological there are baseline concentration and quality depends on desired use of water parameters accessed are different depending on usage of water like for drinking, irrigation or for livestock.

To ensure safe and healthy water for consumption and other uses, various measures are taken to monitor and improve water quality. These include regular testing and analysis of water samples, implementation of water treatment processes to remove contaminants, and development and enforcement of regulations to prevent pollution and protect water resources.

Research is performed and datasets is collected on ground water quality dataset 2020 from Telangana pollution board which is an open source with the parameters like "pH, E.C, TDS, CO3, HCO3, Cl, F, NO3, SO4, Na, K, Ca, Mg, T.H, SAR", target parameters describe the usage and purpose of water.

The target parameters are:

1. WQI.
2. Classification 1
3. RSC
4. TDS
5. classification

whereas Classification parameter gives the 9 types of classifications C3S1, C2S1, C4S1, C4S2, C3S2, C4S4, C3S3, C4S3, C1S1.which describes sodium and salinity levels, RSC Residual sodium carbonate: RSC is the excess amount of carbonate and bicarbonate. When RSC usage surpasses the permitted limit, irrigation suffers (>2.5). and Classification 1 as target variable in the dataset classifies 2 values M.R Mineral Rich and P.S. Poor safe, WQI gives the single value with its grades.

D. Water Parameter standards:

The standard percentage of these chemicals to maintain pure quality of water can vary depending on the specific water quality standards and regulations in your region. However, I can provide some general information on the acceptable levels of these chemicals in drinking water according to the United States Environmental Protection Agency (EPA) Safe Drinking Water Act (SDWA) Maximum Contaminant Level (MCL) regulations.[14] These MCLs are set to protect public health and are based on the best available science and risk assessments.

Here are the MCLs for the chemicals of the parameters which we are using to determine the quality of water.

	Parameter	Safe Range	Units
0	pH	(6.5, 8.5)	
1	E.C	(0, 750)	µS/cm
2	TDS	(0, 500)	ppm
3	Cl	(0, 250)	mg/L
4	F	(0.7, 1.5)	mg/L
5	NO3	(0, 10)	mg/L (as NO3-N)
6	SO4	(0, 250)	mg/L
7	Na	(0, 200)	mg/L
8	K	(0, 10)	mg/L
9	Ca	(0, 100)	mg/L
10	Mg	(0, 50)	mg/L

Fig 2: water parameter standards

MCL's on the taken parameters.

Maximum Contaminant Level (MCL) is the highest level of a contaminant that is allowed in drinking water, as established by regulatory agencies such as the Environmental Protection Agency (EPA) in the United States. MCL values are set based on health considerations to ensure that the consumption of water remains safe for human health.

E. Target parameters

1. Water quality indicators

The water quality index is a valuable tool for evaluating the suitability of water for various purposes, such as drinking, irrigation, and aquatic life. In this project, we present a code snippet that calculates the WQI for a dataset of water quality parameters. Water quality is accessed with its water quality index, wqi gives one value by taking various water parameters as input and output as WQI and chemicals compositions their index levels ranges, quantity available in the water. The combinations change as the composition's changes, this can show the impact on the next available composition in the

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
***Article Received:** 05 July 2023 **Revised:** 25 August 2023 **Accepted:** 15 September 2023*

water, which might make the water more pollutant. To solve these issues our model Purity regulator-based composition model (PRCM) will describe the solvents to be added within the range and at same time maintain the purity of water sodium and salinity values for irrigation or livestock's. Hence the study took different parameters.

Water data calculate WQI and finds the behavior that is typical to normal range of water whether the water is miner rich, poor, or classified with high salinity and high sodium centers on the identification of anomalous behavior, or behavior that is not typical of normal operation.

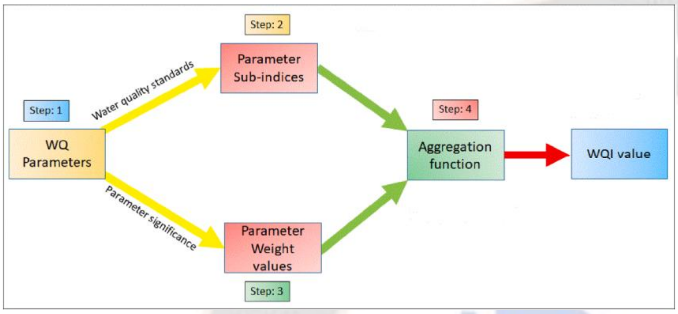

Fig 3. General structure of WQI model.

WQI model structure

The general structure of WQI models is illustrated in Fig. 3 and shows that most WQIs contain four main steps (Abbasi and Abbasi, 2012; Abrahao ˜ et al., 2007; Lumb et al., 2011; Sutadian et al., 2018), namely:

Steps to calculate WQI:

Load Dataset: Dataset considered is from open-source dataset. It contains 355 rows and 27 columns.

Water Parameter selection: Parameter selection is the first step in calculating WQI dataset contains has parameters like "'pH', 'E.C', 'TDS', 'CO3', 'HCO3', 'Cl','F', 'NO3 ', 'SO4', 'Na' , 'K', 'Ca', 'Mg', 'T.H'" are the parameters used to calculate th e WQI.

Assigning Parameter Weights: Each parameter is assigned a weight in the parameter weights dictionary. These weights represent the relative importance of each parameter in determining water quality.

Table 1: parameter weights

Parameters	Units
pH	0.1
E.C	0.15
TDS	0.2
CO3	0.05
HCO3	0.08
CI	0.07
F	0.05
NO3	0.03
SO4	0.05
Na	0.06
K	0.04
Ca	0.06
Mg	0.07
T.H	0.08
SAR	0.07

Above table 1 describes the standard water parameter weights to calculate WQI.

Calculating Sub-Indices: For each water sample, the code calculates sub-indices for each parameter by multiplying the parameter value with its assigned weight. This quantifies the contribution of each parameter to the overall water quality assessment.

Calculating the Overall WQI: The sub-indices are summed to calculate the overall WQI for each water sample. This provides a single value that represents the comprehensive water quality assessment.

```
Calculated WQI values:
0       64.492641
1       21.061445
2       28.258677
3       44.931620
4       42.245614
          ...
350     84.389686
351     52.660483
352     28.398994
353    183.905383
354     62.566600
Name: WQI, Length: 355, dtype: float64
```

Fig 4: calculated WQI values.

Above fig 4 describes WQI values calculated for water quality dataset of 355 rows and 26 columns**.**

Assigning Classifications: Based on a defined threshold value (e.g., 50), the code assigns a classification ("Good" or "Poor") to each water sample. This classification helps interpret the calculated WQI values.

```
Assigned grades:
0           Good
1           Poor
2       Moderate
3       Moderate
4       Moderate
          ...
350    Very Good
351         Good
352     Moderate
353    Excellent
354         Good
Name: Grade, Length: 355, dtype: object
```

fig 5. WQI classified grades.

Grades given to water samples are:

Table 2 water samples grade.

WQI value	Grade
0-24	poor
25-49	moderate
50-70	good
71-99	Very good
Above 100	Excellent

Above table 4 describes the grades given by WQI.

2. Classification 1:

Classification1 in water dataset target variable is used to predict water quality classified into 2 classes MR, P.S.

1. MR: MR describes the water quality as mineral rich and safe water predicts label as 1.

2. PS: P. S. describes the water quality as poorly safe--unsafe water--label as 0.

3. U.S. : U.S. describes the water quality as unsafe – unsafe water – label as 0

Table 5 classification 1 types with labels

Classification 1 types	Classification values	label
P.S	313	0
U.S.	23	0
M.R.	19	1

The above table 5 describes classification.1 with its types 3 l abelled as 1 and 0 is P.S is poorly safe for irrigation are 313 out of 355 and U.S unsafe are 23 which are not used for irrig ation and only M.R. mineral rich water is labelled as 1 are o nly 19 out of 355 are good in minerals can be used for drinki ng, irrigation.

3. Classification of ground water based on Residual sodium carbonate:

When compared to the alkaline earths (Ca^{2+} and Mg^{2+}),

RSC is the excess amount of carbonate and bicarbonate. When RSC usage surpasses the permitted limit, irrigation suffers (>2.5).

As soil water becomes more concentrated due to evaporation and plant transpiration, the propensity of Ca^{2+} and Mg^{2+} to precipitate increases. These ions then become fixed in the soil through the base exchange process, reducing soil permissibility.

SC = ((CO3 2-) + (HCO3-) - ((Ca2+) + (Mg2+))

Whereas meq /L (milliequivalent per litre) is used to represent concentrations

RSC of no more than 1.25 is secure.

RSC is marginal for values between 1.25 and 2.50.

RSC is not suited if it is larger than 2.50.

4.Classification of ground water for animals and poultry based on TDS as target variable.

Groundwater used for livestock and poultry:

Table 6: Usage of ground water with water parameter level

Water parameter level	Grade	Data classification with Grades Of TDS out of 355 rows
TDS < 1000 mg / L	Super good	264
TDS between 1000 to 3000	very satisfactory	89
TDS between 3000 to 5000	Satisfactory for animals but not poultry	2
TDS between 5000 to 7000	Limited use for animals not fit for poultry	0
TDS > 10,000	bad	0
TDS between 7000 to 10,000	worst	0

The table 6 presents data on water quality parameters in relation to their respective grades and recommended usage for animals and poultry.

The data is categorized into different TDS ranges and their corresponding grades and usage recommendations are provided.

1. Water with a TDS level below 1000 mg / L is determined as “Super good”. This quality of water is

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
***Article Received:** 05 July 2023 **Revised:** 25 August 2023 **Accepted:** 15 September 2023*

deemed suitable for all classes of animals and poultry, indicating no significant adverse effects.

2. TDS range of 1000 to 3000 mg/L, the water is labelled as "very satisfactory." While it remains satisfactory for animals, it might cause temporary mild diarrhea in animals unaccustomed to such levels. Poultry might also exhibit watery droppings when exposed to water nearing the upper limit of this range.
3. TDS levels falling between 3000 and 5000 mg/L, the water is considered "Satisfactory for animals but unfit for poultry." Animals can tolerate this range, although those unfamiliar with it might experience temporary diarrhea, particularly if the water contains predominant sulfate salts. Poultry, however, find this water quality unsuitable, leading to increased mortality, decreased growth, and watery feces, especially in turkeys.
4. Water that contains between 5000 and 7000 mg/L has "Limited use for animals and is unfit for poultry." Animals can use it, except for those that are breastfeeding or pregnant. Animals may reject the water at first until they get used to it since it may have a moderate laxative effect. However, it is inadequate for poultry.
5. Water with a Total Dissolved Solids exceeding 10,000 mg/L is labelled as "Not recommended." Such water quality is deemed unsuitable for all classes of animals and poultry due to its adverse effects.
6. TDS Within the range of 7000 to 10,000 mg/L, the water has "Very limited use." Pregnant and lactating cows, horses, and sheep, along with their young, face considerable risk when exposed to this water. Older ruminants and horses might tolerate it better, but it remains unfit for poultry and likely unsuitable for swine.

This comprehensive table data underscores the critical relationship between TDS levels, water grades, and their impact on different animal species, providing valuable insights into the recommended usage based on water quality.

5.Classification as ground water with Classification as the Target classes:

The above dataset described in fig 4 Classification as the target variable consists of 9 classes mentioned in table 4 with the level of sodium and salinity and the purpose and suitability is evaluated.

The target classes are:

Table 7: Target class Classification water analysis

S.NO	Classification	Water quality analysis	Data classified out of 347 pre-monsoon seasons
1	**C1S1**	low sodium and salt levels is suitable for crops.	2
2	C2S1	Low sodium and medium salinity water is mostly used for irrigation.	77
3	**C3S1**	The low sodium and high salinity waters necessitate adequate drainage. Choose crops that can withstand salt well.	227
4	**C3S2**	High salinity and medium sodium fluids which call for good drainage.	14
5	**C3S3**	These waters' high sodium and salinity	2
6	**C4S1**	Waters with very low sodium content and excessive salinity	17
7	**C4S2**	Medium salt and very high salinity fluids	11
8	**C4S3**	very high salinity and high sodium waters produce harmful level of exchangeable sodium.	2
9	**C4S4**	very high sodium and very high salinity, not suitable for irrigation	2

An experimental analysis is performed on ground water dataset of pre-monsoon water quality data 2020. The data set has Classification as the target variable with total number of 374 water quality data for pre-monsoon season as C1S1, C2S1, C3S1, C4S1. Water is classified based on water parameters and its concentration of chemical values.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
***Article Received:** 05 July 2023 **Revised:** 25 August 2023 **Accepted:** 15 September 2023*

A. Counting the Occurrences of Each Value in the 'Classification' Column:

```
import pandas as pd

# Load the dataset and select relevant columns
data = pd.read_csv('ground_water_quality_2020_pre .csv')

# Count the occurrences of each value in the 'Classification.1' column
classification_counts = data['Classification'].value_counts()

# Display the counts
print("Counts of 'Classification' values:")
print(classification_counts)
```

Fig 5- Code for Counting the occurences

This Python code snippet shows how to use the panda's module to load a dataset from a CSV file and count how many times it occurs of each unique value in a specified column, in this case, the 'Classification' column. The counts are then displayed in a tabular style.

1. Importing Libraries: The first step is to import those necessary libraries. To work with data frames, we import pandas as 'pd' in the present example.
2. Loading the Dataset: We load the dataset into a pandas Data Frame called 'data' from a CSV file named 'ground_water_quality_2020_pre.csv'. This Data Frame will enable us to work with the data more effectively.
3. Counting Occurrences: We use the value_counts() method to count the occurrences of each unique value in the 'Classification' column and save the result in a variable called 'classification_counts'. This method returns a Series containing the counts, where the index represents the unique values, and the values represent their individual counts.
4. Displaying the Counts: Finally, we use the print () method to print out the counts of each unique value in the 'Classification' column. The following list highlights the number of times each category appears in the dataset.

Output:

The code will produce an output like this:

- C3S1: This category appears 227 times in the 'Classification' column.
- C2S1: This category appears 77 times.
- C4S1: This category appears 17 times.
- C3S2: This category appears 14 times.
- C4S2: This category appears 11 times.
- C4S4: This category appears 2 times.
- C4S3: This category appears 2 times.
- C1S1: This category appears 2 times.
- C3S3: This category appears 2 times.
- O.G: This category appears 1 time.

Each line represents a unique value (classification category) found in the 'Classification' column of the dataset, followed by the number of times that category occurs.This information is valuable for understanding the distribution of different categories within the dataset. This output shows the counts of each unique value in the 'Classification' column, which can be valuable for understanding the distribution of categories in your dataset.

V. *Experimental Results*

Table 8: Dataset after classification with target variable

sno	Classification	Classification(grades)	RSC meq / L	RSC Grade(grades)	TDS	TDS_Grade(grades)	Classification.1	classification.1(grades)	WQI	WQI(Grade)
1	C3S1	1	-0.59737	Secure	1069.44	Super Good	P.S.	0	64.493	Good
2	C3S1	1	1.600493	Secure	348.8	Super Good	MR	0	21.061	Poor
3	C2S1	0	-0.79868	Secure	472.32	Super Good	P.S.	1	28.259	Moderate
4	C2S1	1	-2.3977	Secure	738.56	Super Good	P.S.	1	44.932	Moderate
5	C2S1	1	-2.59786	Unsuitable	666.88	Super Good	P.S.	1	42.246	Moderate
6	C3S1	1	-1.59819	Secure	680.32	Super Good	P.S.	0	41.384	Moderate
7	C3S1	0	0.60148	Secure	504.32	Super Good	P.S.	0	31.216	Moderate
8	C3S1	0	2.400493	Secure	519.04	Super Good	MR	0	30.146	Moderate
9	C3S1	1	0.400658	Secure	270.08	Super Good	P.S.	0	16.723	Poor

Water data (table 8) after classifying and predicting the quality of water by giving grades to the target variables like in (table 1) dataset is the classification of water quality in (table 8) we have classified and predicted the quality of water and its usage and purpose for classification we have 9 types of classifications are C3S1,C2S1,C1S1.., are classified according to the salinity of ground water and given grades as 1 and 0 1 describes its good for irrigation purpose and 0 which is not suitable for usage, and RSC is residual sodium carbonate is classified with secure, unsuitable ,very good…, TDS is given with grades as super Good , Good…,Classification are given with 0 as poorly safe (PS) and 1 as mineral rich(MR),WQI is the water quality index given with grades as poor, Good, Moderate…. All the classifications with grades we can classify how much of water is useful with its purpose.

Table 9: Machine learning model to find Accuracy

SNO	Machine learning model WQI	WQI	Classification	Classification.1	RSC Grades	TDS Grades
I 1	Logistic Regression Accuracy	0.66	0.96	0.69	0.97	0.96
2	Support Vector Machine Accuracy	0.94	0.91	0.72	0.96	0.97

3	Decision Tree Accuracy	0.96	0.99	0.96	0.96	0.99
4	Naive Bayes Accuracy	0.86	0.90	0.85	0.59	0.96
5	Random Forest Accuracy	0.97	0.99	0.84	0.94	0.99
6	K-Nearest Neighbors Accuracy	0.93	0.82	0.68	0.87	0.99
7	Gradient Boosting Accuracy	0.93	0.99	0.96	0.99	0.99
8	XGBoost Accuracy	0.94	0.99	0.96	0.97	0.99

Machine Learning models(table 9) are compared with its accuracy for Classification, Classification.1,RSC,TDS XGBoost is giving highest accuracy among all as 0.99,0.96,0.97,0.99 and only Random forest gave accuracy for WQI

VI. Conclusion

By using these datasets one can train an ML Classification model to classify the water quality into one of the multiple classes and know where and how the water can be used with the above fig water quality analysis ground water dataset 2020 is classified with more C2S1 which is low sodium and medium salinity this water is mostly used for irrigation next classification is C4S1 water with very low sodium and high salinity which is not suitable for irrigation. Out of 347 data only 227 are suitable for irrigation purpose, and 2 sets are used for crops, remaining data is not suitable for irrigation or crops.

The above table 3 describes the water parameter TDS grades given to the dataset according to table 7.

Water data with TDS (Total Dissolved Solids) graded as super good with 264 are suitable for animals and poultry, and TDS with grade “very satisfactory” 89 which are satisfactory for animals but may cause temporary mild diarrhea and can be used for poultry and this analysis says that ground water quality dataset 2020 is suitable for animal and poultry but can cause temporary mild diarrhea for 89 out of 355 data.

Overall in the dataset only at 19 places can be used for drinking ,264 suitable for livestocks out of 355 data.

REFERENCES

[1] Krishnan, K.S.D. Multiple Linear Regression-Based Water Quality Parameter Modeling to Detect Hexavalent Chromium in Drinking Water. In Proceedings of the 2017 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET), Chennai, India, 22–24 March 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 2434–2439.

[2] Ighalo JO, Adeniyi AG (2020) A comprehensive review of water quality monitoring and assessment in Nigeria. Chemosphere 260:127569

[3] Bhoi SK, Mallick C, Mohanty CR (2022) Estimating the Water Quality Class of a Major Irrigation Canal in Odisha, India: ASupervised Machine Learning Approach. Nat Environ Pollute. Technol 21(2):433–446

[4] Prasad B.C. and Narayana T.S., Nat. Environ. Poll. Tech., 3, 10 (2004)

[5] Garg A., Kripalani C. and Brighu U., Nat. Environ., Poll. Tech., 3, 1(2004) 73.

[6] Ahmed AN, Othman FB, Afan HA, Ibrahim RK, Fai CM, Hossain MS, Elshafie A (2019) Machine learning methods for better water quality prediction. J Hydrol 578:1–18

[7] Pant, R.R.Spatio temporal variations of hydrogeo chemistry and its controlling factors in the Gandakiriver basin, Central Himalaya Nepal. Sci. Total. Environ. 2018, 622–623, 770–782. [CrossRef]

[8]. Khatunm, M. Phytoplankton assemblage with relation to water quality in Turag River of Bangladesh. Casp. J. Environ. Sci. 2020, 18, 31–45.

[9]. Rose-Rodrígue, R. Water and fertilizers are used efficiently in two hydroponic systems for tomato production. Hortic. Bras. 2020, 38, 47–52. [CrossRef]

[10]. Trombadore, O. Effective Data Convergence, Mapping, and Pollution Categorization of Ghats at Ganga River Front in Varanasi; Springer: Berlin/Heidelberg, Germany, 2020; Volume 27, pp. 15912–15924.

[11]. King, K.W.; Williams, M.R.; Macrae, M.L.; Fausey, N.R.; Frankenberger, J.; Smith, D.R.; Kleinman, P.J.A.; Brown, L.C. Phosphorus Transport in Agricultural Subsurface Drainage: A Review. J. Environ. Qual. 2015, 44, 467–485. [CrossRef]

[12]. Hupfer, M.; Lewandowski, J. Oxygen Controls the Phosphorus Release from Lake Sediments—A Long-Lasting Paradigm in Limnology. Int. Rev. Hydrobiol. 2008, 93, 415–432. [CrossRef]

[13]. Tammeorg, O.; Möls, T.; Niemistö, J.; Holmroos, H.; Horppila, J. The actual role of oxygen deficit in the linkage of the water quality and benthic phosphorus release: Potential implications for lake restoration. Sci. Total Environ. 2017, 599–600, 732–738. [CrossRef]

[14]. Søndergaard, M.; Jensen, J.P.; Jeppesen, E. Role of sediment and internal loading of phosphorus in shallow lakes. Hydrobiologia 2003, 506–509, 135–145. [CrossRef]

[15]. Cooke, G.D.; Welch, E.B.; Peterson, S.A.; Nichols, S.A. Restoration and Management of Lakes and Reservoirs, 3rd ed.; Taylor & Francis/CRC Press: Boca Raton, FL, USA, 2005; p. 591.

[16]. Klapper, H. Control of Eutrophication in Inland Waters; Ellis Hornwood: New York, NY, USA, 1991; pp. 116–266.

[17]. Lossow, K.; Gawro ´nska, H. Lakes—Review of restoration methods. Munic. Rev. 2000, 9, 91–106.

[18]. Kuha, J.K.; Palomäki, A.H.; Keskinen, J.T.; Karjalainen, J.S. Negligible effect of hypolimnetic oxygenation on the

trophic state of lake Jyväsjärvi, Finland. Limnologica 2016, 58, 1–6. [CrossRef]

G. Eason, B. Noble, and I. N. Sneddon, “On certain integrals of Lipschitz-Hankel type involving products of Bessel functions,” Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955. *(references)*

[19] J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[20] I. S. Jacobs and C. P. Bean, “Fine particles, thin films and exchange anisotropy,” in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.

[22] K. Elissa, “Title of paper if known,” unpublished.

[23] R. Nicole, “Title of paper with only first word capitalized,” J. Name Stand. Abbrev., in press.

[24] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, “Electron spectroscopy studies on magneto-optical media and plastic substrate interface,” IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].

[25] M. Young, The Technical Writer’s Handbook. Mill Valley, CA: University Science, 1989.

[26] Electronic Publication: Digital Object Identifiers (DOIs):

Article in a journal:

[27] D. Kornack and P. Rakic, “Cell Proliferation without Neurogenesis in Adult Primate Neocortex,” Science, vol. 294, Dec. 2001, pp. 2127-2130, doi:10.1126/science.1065467.

Article in a conference proceedings:

[28] H. Goto, Y. Hasegawa, and M. Tanaka, “Efficient Scheduling Focusing on the Duality of MPL Representatives,” Proc. IEEE Symp. Computational Intelligence in Scheduling (SCIS 07), IEEE Press, Dec. 2007, pp. 57-64, doi:10.1109/SCIS.2007.357670.

[29] Neha Radhakrishnan and Anju S Pillai Comparison of Water Quality Classification Models using Machine Learning

[30] Abbasi, T., Abbasi, S.A., 2012. Water-Quality Indices. Water Quality Indices. Elsevier, pp. 353–356. https://doi.org/10.1016/B978-0-444-54304-2.00016-6.

[31] Lumb, A., Sharma, T.C., Bibeault, J.-F., 2011. A Review of Genesis and Evolution of Water Quality Index (WQI) and Some Future Directions. Water Qual. Expo. Heal. https://doi.org/10.1007/s12403-011-0040-0.

[32] OLASUPO O. AJAYI 1 , ANTOINE B. BAGULA 1 , HLONIPHANI C. MALULEKE 1 , ZAHEED GAFFOOR 2 , NEBO JOVANOVIC2 , AND KEVIN C. PIETERSEN 3 “WaterNet: A Network for Monitoring and Assessing Water Quality for Drinking and Irrigation Purposes”. IEEE Xplore Full-Text PDF: