Water Quality Index (WQI)

Joel Swapnil Singh (ITM2017002), Amritesh Bhaskar (IIT2017143),
Sourabh (IIT2017139), and Pratham Singh (IRM2017006)
Indian Institute of Information Technology Allahabad
(Dated: April 23, 2020)

This paper deals with the estimation of water quality by evaluating the Water Quality Index(WQI) for different samples of river Ganga and Sangam. It compares the water quality of river Ganga and sangam seasonally/monthly and characterises the water fit for consumption and other purposes by means of WQI.

Usage: Estimating the Water Quality Index(WQI) for different water samples.

Keywords: Water Quality Index(WQI).

I. PROBLEM IDENTIFICATION

Due to the rapid increase in the population density, various human activities such as bathing, washing clothes, the bathing of animals, and dumping of various harmful industrial waste into the rivers are emerging as a vital cause in the pollution of water bodies such as Ganga and its tributaries, parts of Sangam and many other rivers. This can be a threat to mankind as well as animals as the water quality is degrading day by day. Hence a check on water quality needs to be maintained for consumption and other purposes.

II. INTRODUCTION

A Water Quality Index (WQI) is a means by which water quality data is summarized for reporting the public in a consistent manner as it tells us what the quality of water is for drinking and other purposes. There are different distribution indices for WQI which rate the water accordingly. Some keep lower WQI for excellent water while others keep higher. Here, the WQI measures different parameters of water such as pH, Dissolved Oxygen(DO), Conductivity, Oxygen Reduction Potential(ORP) and Temperature. After which it combines these measures into one score. This calculation produces a score between 0 and 100. The lower the score the better the quality of water.

A. Data-Set

The apparatus used here is Smart water sensor (by Libelium) which is equipped with multiple sensors that measure a dozen of the most relevant water quality parameters such as pH, dissolved oxygen (DO), oxidation-reduction potential (ORP), conductivity (salinity), turbidity and temperature. Following parameters are considered here:

Parameters	DO	pH	ORP	Cond	Temp
Standard Ranges	1-20	6.5-8.5	0.2-0.6	50-400	20-30
Standard Values	10	7	1	225	25
Sensor Range	(1-20) mg/L	pH (0-14)	(0- +-1999) mv)	(0-60000 uS/cm)	(0-100 deg C)

B. Data Preprocessing

Data is preprocessed and the penalties are assigned as they would affect the water quality. The preprocessing table is given below:
(Consider x as the sensed-value and y as the processed value.)

1. *Dissolved Oxygen (DO)*

If x is greater than or equal to 10 then y=1 else y=11-x.

2. *pH*

If x is between 6.5 and 8.5 (both inclusive) then y=0 (only if x is not 7). If x is 7 then y=1. If x is smaller than 6.5 then y=1+6.5-x. If x is greater than 8.5 then y=1+x-8.5.

3. *Oxygen Reduction Potential (ORP)*

If x is between 50 and 400 (both inclusive) then y=1. If x is smaller than 50 then y=1+(50-x)/781. If x is greater than 400 then y=1+(x-400)/781.

4. *Conductivity*

Conductivity is taken as it is.

5. Temperature

If x is between 4 and 30 (both inclusive) then y=1. Otherwise y=1+max(x-30,4-x).

C. Assigning Weights

Weights are assigned on the basis of variations the parameters were showing in the entire dataset. pH having the highest(30), DO(20), Conductivity(3), ORP(2) and Temperature being the least(1). The total sum of the weights assigned is 55. From these assigned weights and the total sum W is evaluated.
Let weight of the parameter be w and the total sum of the weights be T. Then W=w/T for each parameter.

D. Evaluating WQI

After getting weight of parameter(w), Normalised weight(W) and the sensed value(c), q is calculated by formula:

$$q = \frac{c}{s} * 100$$

Using q, the sub-index(si) of that particular parameter is calculated by the formula:

si=w*q

The summation of all sub-indices gives the WQI for the sample.

$$\sum_{n=1}^{5} si = WQI$$

III. SURVEY

Below is a study done on some research papers which give an idea about the Water Quality Index(WQI) along with their techniques, achievements and limitations.

1. Time Scale Changes in the Water Quality of the Ganga River, India and Estimation of Suitability for Exotic and Hardy Fishes

This paper deals with the time scale changes in the water quality of river Ganga and also estimates the survival and growth of various aquatic creatures. The data set comprises of various physicochemical parameters like Temperature, pH, Total Dissolved Solids (TDS), Electrical Conductivity (EC), Total Hardness(TH), Sulphates, Phosphates, Alkalinity, Nitrate, Chloride, Dissolved Oxygen (DO), Biological Oxygen Demand (BOD) and Chemical Oxygen Demand (COD) which are taken out with the help of Multi probes ITS -701, Automated Oxygen Analyser and Perkin Elmer UV/VIS spectrophotometer model Lambda 25. The results were sorted on the basis of different seasons viz. Summer, Monsoon, and Winter. The mean and standard deviation were evaluated for each component mentioned above by the formulae:

$$\sum_{i=1}^{n} \frac{x_i}{n} = E[X]$$

$$Var(x) = E[X^2] * [E(X)]^2$$

The outcomes are mainly emphasized on COD (Chemical Oxygen Demand) and BOD(Biochemical Oxygen Demand). Chemical Oxygen Demand (COD) is a method of estimating how much oxygen would be depleted from a body of receiving water as a result of bacterial action whereas Biochemical oxygen demand (BOD) is the amount of dissolved oxygen needed (i.e. demanded) by aerobic biological organisms to break down organic material present in a given water sample at certain temperature over a specific time period. Through these two factors, the survival and growth of aquatic creatures are identified as higher values of BOD and COD can favor some species but can also be a threat to others.

2. Effect of physicochemical and biological parameters on the quality of river water of Narmada, Madhya Pradesh, India

The purpose of the study was development of Water Quality Index (WQI) using eight parameters pH, Temperature, Total Dissolved Solids (TDS), Turbidity, Nitrate-Nitrogen, Phosphate, Biological Oxygen Demand (BOD), Dissolved Oxygen (DO) measured at six different sites along the river Narmada. The methods used for calculation of WQI are Weighted Arithmetic Water Quality Index (WAWQI), National Sanitation Foundation Water Quality Index (NSFWQI) and Canadian Council of Ministers of the Environment Water Quality Index (CCMEWQI) in this study. The Water Quality Index is employed as it incorporates the different physical, chemical and biological parameters for the determination of water quality indices using the various mathematical equations. It is observed from the result that the water quality index is in the range of excellent in the winter season and falls from poor to unsuitable for drinking purposes in monsoon season at all the sites studied using the WAWQI method. However NSFWQI method shows that the quality of water for the monsoon season was medium for all sites. And finally the results from CCMEWQI method shows that the quality of water falls in the range of poor to fair for all the sites. The major conclusion to this study was that the water quality of river Narmada was found to be suitable for human consumption in the winter and summer season as the values of parameters found to be (pH 7.7–8.48, TDS 108–234 mg L 1 , Turbidity 0.01–178.25 NTU, Nitrate-Nitrogen 0.03–3.14 mg L 1 , Phosphate 0.01–0.52 mg L 1 , BOD 0.35–2.18 mg L 1 and DO 2.4–7.8 mg L 1) as per the standard values prescribed by different regulatory bodies. On comparison of the methods used for evaluation of water quality of Narmada River at different sites, WAWQI method provides better idea about the water quality.

3. Internet of things enabled real time water quality monitoring system

The purpose of this paper is to present a delineated overview of recent works carried out in the field of smart water quality monitoring in terms of application, communication technology used, type of sensor employed etc. Also, a low cost, less complex water quality monitoring system is proposed. The system embodies a TI CC3200 single chip microcontroller with in-built Wi-FI module and ARM Cortex M4 core which can be connected to the nearest WiFi hotspot for internet connectivity so it can send data from sensors to the cloud, a fabricated sensor that includes a solar cell, Li-ion battery, a power and transmission module which is directly interfaced to the controller, and a LCD for viewing measured parameters. The proposed experimental setup is used to measure five parameters namely Turbidity (TU), Oxidation Reduction Potential (ORP), Electrical Conductivity (EC), pH and compared with drinking water quality standards defined by WHO. To demonstrate the working of the system, and the various options for data analysis, measured quantities of contaminants such as salt and soil are mixed with 350 ml of pipe water and testing is performed exemplifying other features like providing online data to consumers.

4. Drinking water quality assessment and its effects on residents' health in Wondo genet campus, Ethiopia

Healthy and clean drinking water is basic need and right of every common people. Many developing countries like Ethiopia are still on there way of providing healthy drinking needs to every person and as a result many unprotected ways of water consumption are making there way to common people homes like river,etc. Keeping this in mind and considering WHO's facts the 80% of worlds diseases are causes of improper sanitation wherein Ethiopia's 60% included this study was carried out in Wondo Genet Campus for water quality assessment and further providing appropriate water treatment mechanism. Many essential physio-chemical parameters were taken into consideration like- turbidity, temperature, TDS, EC and chemical constituents like pH, calcium, sodium, chloride,etc. Samples were taken from 10 different parts of campus that were major water body entries. Main focus given onto taking water samples carefully with help of every soul in campus thankfully. Methodology used was first included selecting best 10 locations then letting water run through them 5 minuted prior to ensure any chemical involvement. Further analysis for every components was carried out and checked there lie in standard ranges. Coliforms are major factors in deciding drinking water quality as more of there concentration is more prone to bacteria. As a result almost all locations were found to be perfectly fine in terms of there water sample components except one had no coliform which was good but 3 had more than permissible, hence were looked into more carefully.

5. Assessment of Ganga river ecosystem at Haridwar, Uttarakhand,India with reference to water quality indices

The sacred and holy river Ganges is one of the oldest known river in the country. Many organizations with help of many physicochemical parameters has tried guess the same for many years and have put many possibilities and facts in front about the current situation of the water quality of the holy river. Its a long river covering many states and crowded cities of around 70 towns and 1000s of villages, hence domestic and industrial waste has been a major pollution factor in past few decades as results show. Another few concerns include Kumbh the covers millions' gathering in a short span. Earlier methods have been used for calculating WQI with parameters like turbidity, OD, BOD, COD,CO2 etc in many locations in UP written in C++ programs.We will be focusing on three popular methods for the same and have carried out the evaluation WQI for past 10 years.

1st method is River Ganges Index of Ved Prakash (2012), it is very widely used of weighted manipulation of assigning weights to parameters according to their importance and summing up the products of weight. Second method uses WQI by NSF as its basis, its an organization under WHO had carried out index development very effectively with 142 experts working on 35 parameters and finally decided to go with 11 parameters. Along with weights ratings are assigned to parameters. 3rd method is of Arithmetic weighted index in which WQI was rather calculated simple arithmetic mean, using the same for calculating quality rating scale for parameters and rest was kept same. After being evaluated for past 10 years the results weren't that satisfying as none of them resulted the same. As a conclusion for all methods we can surely say the water quality for holy river has been ranged poor to good which was expected from discussion above too.

TABLE I. A brief description of the above SURVEY.

Paper Number	Technique(s)	Objectives	Achievement(s)/Data Source	Method Limitation(s)
1	Mean, Standard Deviation , COD and BOD.	Estimation of survival of Aquatic Creatures by checking the changing water quality of river Ganga.	Successful identification of the survival and growth of various species of aquatic creatures on the basis of BOD and COD.	The water quality evaluation was completely based on COD and BOD which sometimes may not be enough since there are various other parameters to go through before concluding anything.
2	Weighted Arithmetic Water Quality Index (WAWQI), National Sanitation Foundation Water Quality Index (NSFWQI), Canadian Council of Ministers of the Environment Water Quality Index (CCMEWQI)	Development of Water Quality Index (WQI) using eight parameters pH, Temperature, Total Dissolved Solids (TDS), Turbidity, Nitrate-Nitrogen, Phosphate, Biological Oxygen Demand (BOD), Dissolved Oxygen (DO)	On collecting the water quality parameters of Narmada river with national and international standards, it is concluded that water quality of Narmada river is found to be suitable for all the sites studied except in monsoon season and WAWQI method is the one that provides better idea about the water quality.	The number of parameters which exceed the criteria tends to be more heavily weighted than parameters which routinely exceed the criteria. As a result, a site with one sampling event that exceeds criteria for multiple parameters may score worse than a site which consistently exceeds the criteria for one parameter.
3	Internet of things, Real time monitoring, Cloud Computing	To provide a detailed survey of recent work in the area of smart water quality monitoring and present a low cost, less complex smart water quality monitoring system	Survey on the tools and techniques employed in existing smart water quality monitoring systems. A low cost, less complex water quality monitoring system has been developed that can be further improved.	Log life expectancy of system, High power consumption Communication overhead.
4	Chemical and Microorganism analysis.	Check water quality at Wondo Genet Campus and provide appropriate water treated mechanism.	Successful sample collection, analysis and treatment action.	Analysis need to be carried separately for each parameter constituent making its prone to mistakes.
5	Weighted Manipulation form (Ved Prakash), Quality Transformation form (NSF-WQI), Arithmetic Weighted Index.	Evaluate river Ganges water quality.	WQI evaluated successfully for past 10 years using 3 methods.	The 3 methods were able to provide range as results didn't agreed on the same.

IV. RESULTS AND DISCUSSION

A. Day-Wise Distribution of WQI

This Chart(FIG. 1.) illustrates the variation of WQI over the span of a year on two holy water bodies Ganga and Sangam. Dataset was reduced on daily basis, then median was applied, to form a nice depicting graph of WQI. The red curve shows the variation in WQI of Sangam and the blue curve shows the variation of river Ganga. In March there was an enormous growth. In following months WQI went down to 30-40 in December.

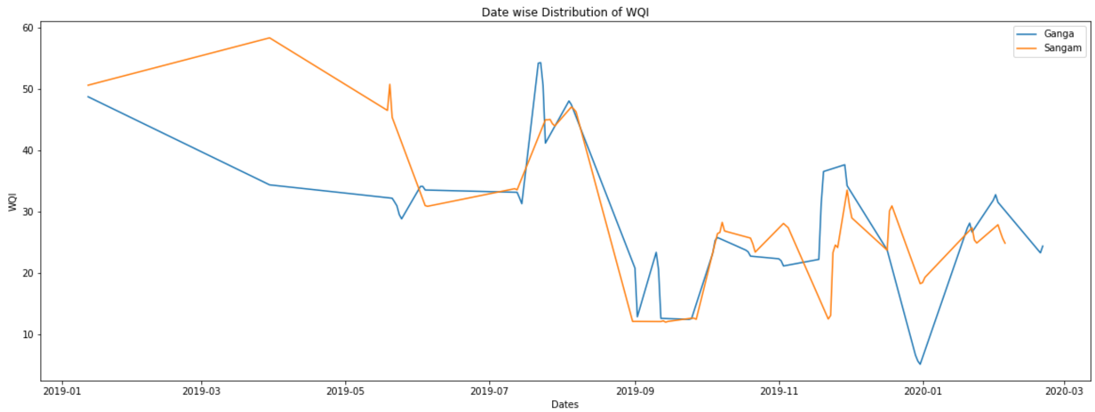

FIG. 1. WQI distribution on a daily basis.

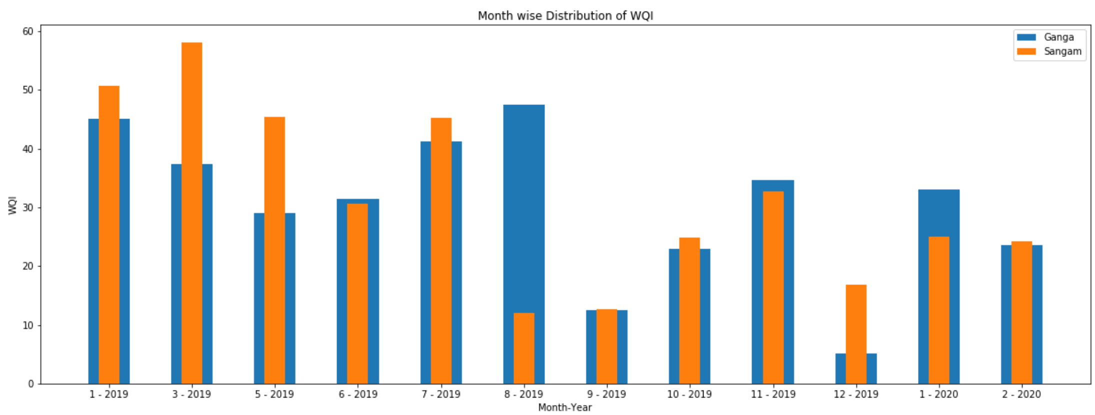

FIG. 2. WQI distribution on a monthly basis.

B. Month-Wise Distribution of WQI

To get the results of the behavior of water quality on a monthly basis the dataset is grouped accordingly and the median is evaluated for every month. After which a histogram(FIG. 2.) is plotted depicting the comparison between the water quality of river Ganga and Sangam by means of WQI. The orange bar shows the WQI of Sangam in respective month and the Blue bar depicts for river Ganga. From this Bar graph, It can be concluded that there is variation of WQI in both the water bodies over the months.

C. ML Model analysis on different datasizes

Now, dataset was trained on many well-known machine learning models for comparing the accuracy on the different dataset attributes. While training, test & training data was divided in 3:1 to increase the productivity. For Training 3 ML models were used : Decision Tree, Support-Vector Machine, Random Forest. All the models were trained on different data sizes for better analysis. Fortunately, average accuracy of dataset was above 99% which is good sign for performance of the models. Results of both the Rivers were almost same. Following figure shows the accuracy of all models in different dataset of both rivers.

Dataset Quantity	Decision Tree		SVM		Random Forest	
	Sangam	Ganga	Sangam	Ganga	Sangam	Ganga
10000	0.9997	0.9856	0.9991	0.9943	0.9997	0.9883
20000	0.9981	0.9927	0.9973	0.9850	0.9981	0.9925
30000	0.9997	0.9947	0.9990	0.9906	0.9997	0.9947
40000	0.9972	0.9959	0.9957	0.9903	0.9975	0.9974
50000	0.9966	0.9959	0.9958	0.9914	0.9971	0.9934
Total	0.9967	0.9959	0.9967	0.9912	0.9978	0.9961

V. CONCLUSION

Maintaining the water quality is of utmost importance as it plays a vital role in catering most of our essential and day to day needs. The results obtained from the above figures show average water quality throughout the year. Although it depicts the best water quality in the month of December for river Ganga and in the month of August for Sangam.
All the three models were successful to a great extent in terms of accuracy and there was almost no difference among the results obtained from the three. Although Random forest does a better job in terms of processing power. These results provide a great estimation about the water quality of river Ganga and Sangam.

VI. REFERENCES

1. Tiwari A, Dwivedi AC and Mayank P "Time Scale Changes in the Water Quality of the Ganga River, India and Estimation of Suitability for Exotic and Hardy Fishes", HYCR.
2. Nidhi Gupta, Pankaj Pandey, Jakir Hussain "Effect of physicochemical and biological parameters on the quality of river water of Narmada, Madhya Pradesh, India", Elsevier, in 2017.
3. S. Geetha and S. Gouthami "Internet of things enabled real time water quality monitoring system", Springer, in 2017.
4. Yirdaw Meride and Bamlaku Ayenew "Drinking water quality assessment and its effects on residents' health in Wondo genet campus, Ethiopia",Springer, in 2016.
5. R. Bhutiani, D. R. Khanna, Dipali Bhaskar Kulkarni, Mukesh Ruhela "Assessment of Ganga river ecosystem at Haridwar, Uttarakhand,India with reference to water quality indices", Springer, in 2014.