ORIGINAL RESEARCH PAPER
Volume - 9 | Issue - 8 | August - 2020 | PRINT ISSN No. 2277 - 8179 | DOI : 10.36106/ijsr
INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH

AN APPROACH FOR CLASSIFICATION OF HEALTH RISKS ON THE ECOSYSTEM DUE TO WASTE DUMPED ON THE BANKS OF THE RIVER.

Computer Science

Ms. Levendra Magna Soares ME(IT) PCCE Verna Goa

Prof. Megha Ainapurkar* Department of Information Technology PCCE Verna, India *Corresponding Author

ABSTRACT

Waste quantities are growing in all countries all around the world. Every year billions of tons of waste are generated. These wastes are result of activities in our homes, businesses and industries and disposal of all this large amount is an enormous environmental problem with many dimensions .Municipal, industrial and agricultural solid waste and biomass deposits cause large scale pollution of land and water. These products have a longer- lasting adverse effect on the environment. These effects include polluting the aquatic life with radioactive waste and substances like plastic that can enter the food chain leading to disturbance in the ecosystem. The Goal is to use data mining techniques to classify the different health risk based on water pollutants to the above mentioned ecosystem. . Using WQI (Water Quality Index) to assist with the above research.

KEYWORDS

Data mining ; K-Means; Classification; Water quality index.

I. INTRODUCTION

Water pollution occurs when harmful substances— often chemicals or microorganisms—contaminate a stream, river, lake, ocean, aquifer, or other body of water, degrading water quality and rendering it toxic to humans or the environment. Safe drinking water is necessary for human health all over the world. Being a universal solvent, water is a major source of infection. According to world health organization (WHO) 80% diseases are water borne. Drinking water in various countries does not meet WHO standards . 3.1% deaths occur due to the unhygienic and poor quality of water . Discharge of domestic and industrial effluent wastes, leakage from water tanks, marine dumping, radioactive waste and atmospheric deposition are major causes of water pollution. Heavy metals that disposed off and industrial waste can accumulate in lakes and river, proving harmful to humans and animals. Toxins in industrial waste are the major cause of immune suppression, reproductive failure and acute poisoning. Infectious diseases, like cholera, typhoid fever and other diseases gastroenteritis, diarrhea, vomiting, skin and kidney problem are spreading through polluted water . Human health is affected by the direct damage of plants and animal nutrition. Water pollutants are killing sea weeds, mollusks, marine birds, fishes, crustaceans and other sea organisms that serve as food for human. Insecticides like DDT concentration is increasing along the food chain. These insecticides are harmful for h u m a n s It is crucial to analyses and control the water pollution. The exceeding levels of pollutants can be recorded and appropriate plans can be devised to control the water pollution and lessen the effects of it.

A water quality index provides a single number (like a grade) that expresses overall water quality at a certain location and time based on several water quality parameters. The objective of an index is to turn complex water quality data into information that is understandable and useable by the public. This type of index is similar to the index developed for air quality that shows if it's a red or blue air quality day. The use of an index to "grade" water quality is a controversial issue among water quality scientists. A single number cannot tell the whole story of water quality; there are many other water quality parameters that are not included in the index. The index presented here is not specifically aimed at human health or aquatic life regulations. However, a water index based on some very important parameters can provide a simple indicator of water quality. It gives the public a general idea the possible problems with the water in the region.

II. METHODOLOGY

A. Weighted Arithmetic Water Quality Index Method

Weighted arithmetic water quality index method classified the water quality according to the degree of purity by using the most commonly measured water quality variables. The method has been widely used by the various scientists and the calculation of WQI was made by using the following equation:

$$WQI = \sum QiWi / \sum Wi$$

The quality rating scale (Qi) for each parameter is calculated by using this expression :

Qi=100[(Vi-Vo/Si- Vo)]
Where

Vi is estimated concentration of ith parameter in the analysed water Vo is the ideal value of this parameter in pure water

Vo = 0 (except pH =7.0 and DO = 14.6 mg/l)
Si is recommended standard value of ith parameter

The unit weight (Wi) for each water quality parameter is calculated by using the following formula:

Where,
Wi=K/Si
K= proportionality constant and can also be calculated by using the following equation:

$$K = \frac{1}{\sum (1/Si)}$$

WQI Value	Rating of Water Quality	Grading
0-25	Excellent Water Quality	A
26-50	Good Water Quality	B
51-75	Poor Water Quality	C
76-100	Very Poor Water Quality	D
Above 100	Unsuitable for drinking purpose	E

B. K-Means Algorithm

K-Means algorithm is an iterative algorithm that tries to partition the dataset into K pre-defined distinct non- overlapping subgroups (clusters) where each data point belongs to only one group. It tries to make the intra-cluster data points as similar as possible while also keeping the clusters as different (far) as possible. It assigns data points to a cluster such that the sum of the squared distance between the data points and the cluster's centroid (arithmetic mean of all the data points that belong to that cluster) is at the minimum. The less variation we have within clusters, the more homogeneous (similar) the data points are within the same cluster.

I. RESULTS
1) WQI.

 PRINT ISSN No. 2277 - 8179 | DOI : 10.36106/ijsr

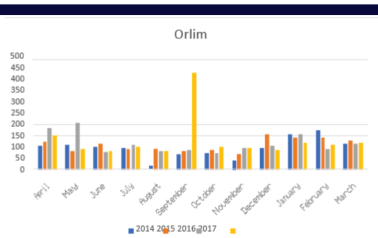

Fig 1.Comparison of WQI at Orlim 2014-2017

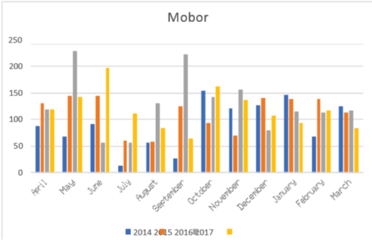

Fig 2.Comparison of WQI at Mobor 2014-2017

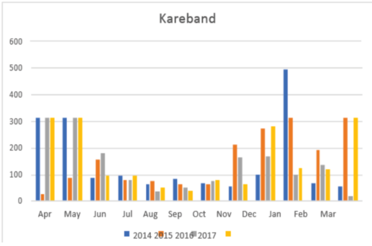

Figure 3.Comparison of WQI at Kareband 2014-2017.

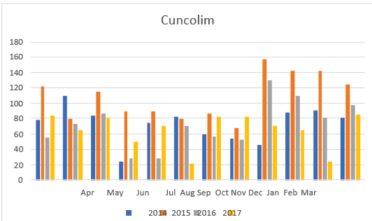

Figure 4.Comparison of WQI at Cuncolim 2014-2017.

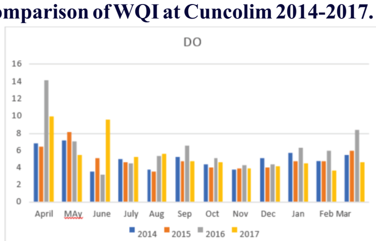

Figure 5.Comparison of Dissolved Oxygen at Orlim 2014-2017

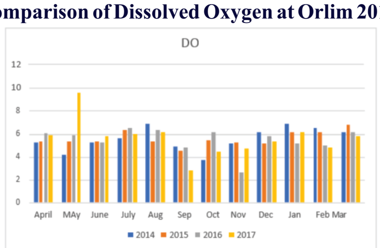

Figure 6.Comparison of Dissolved Oxygen at Mobor 2014-2017

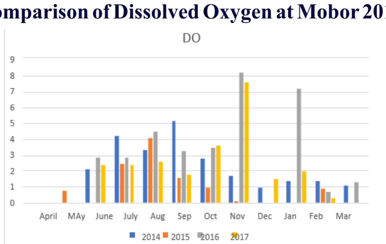

Figure 7.Comparison of Dissolved Oxygen at Kareband 2014-2017

Water Quality index of the present waterbody i.e river Sal is obtained from various physiochemical parameters. The data contains samples of water collected at 4 different locations those are Kareband, Orlim ,Mobor and Cuncolim.

The water quality rating study clearly shows that the status of the waterbody is eutrophic and it is unsuitable for human use.

We can see that the water quality decreases during the months march to may and November to December.

It can be concluded that due the water flow during the rainy season i.e during the months June to October the Water quality improves as compared to the other months.

Bio-chemical oxygen demand is a parameter to assess the organic load in a waterbody. Polluted waters have recorded a higher Bod value. We can see that during the months that have high levels of pollution the BOD is also high.

Dissolved oxygen regulates the distribution of flora and fauna. The concentration of DO is higher in locations where the levels go pollutions are low.

Also BOD and DO inversely proportional to each other if the concentration of DO is higher than the Bio-chemical oxygen demand decreases and when the BOD is higher the Concentration of DO decreases.

2) K-Means.

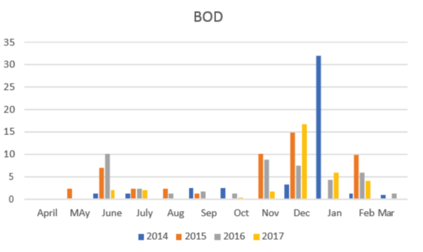

Figure 8.Comparison of BOD at Kareband 2014-2017

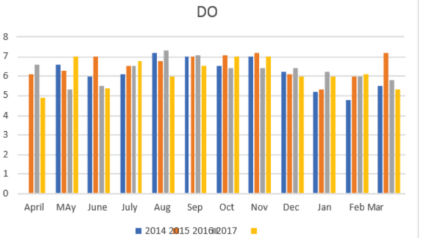

Figure 9.Comparison of Dissolved Oxygen at Cuncolim 2014-2017

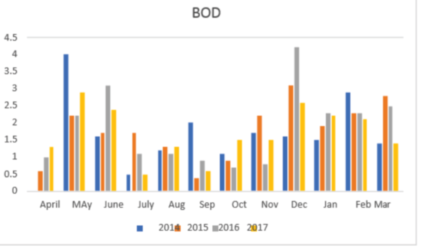

Figure 10.Box plot comparing FC of 3 clusters at Orlim.

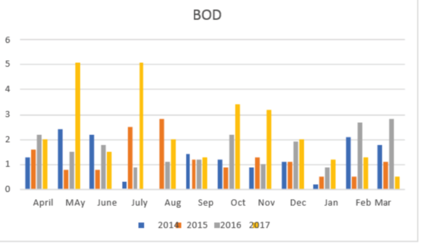

Figure 11.Box plot comparing TC of 3 clusters at Orlim.

 PRINT ISSN No. 2277 - 8179 | DOI : 10.36106/ijsr

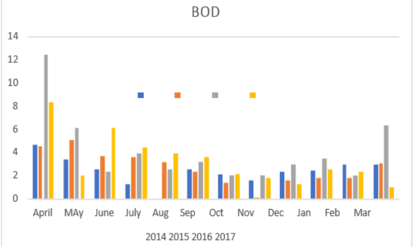

Figure 12.Figure 8.Comparison of BOD at Orlim 2014-2017

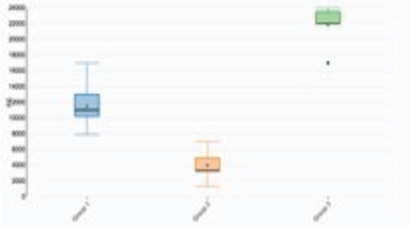

Figure13.Box plot comparing BOD of 3 clusters at Orlim.

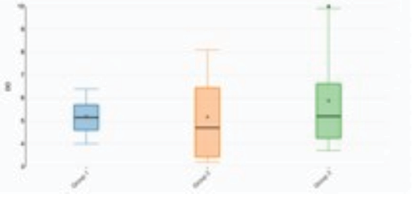

Figure 14.Box plot comparing DO of 3 clusters at Orlim.

Figure 15.Box plot comparing TC of 3 clusters at Orlim.

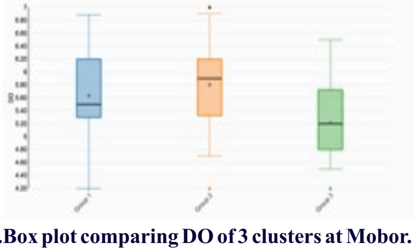

Figure 16.Box plot comparing DO of 3 clusters at Mobor.

Figure 17.Box plot comparing FC of 3 clusters at Orlim.

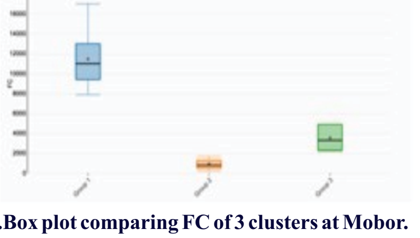

Figure 18.Box plot comparing FC of 3 clusters at Mobor.

Figure 19.Box plot comparing BOD of 3 clusters at Mobor.

Figure 20.Box plot comparing TC of 3 clusters at Mobor.

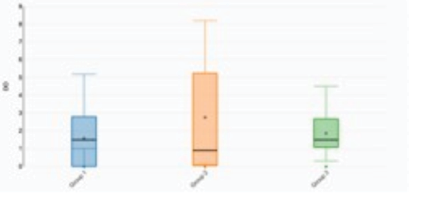

Figure 21.Box plot comparing DO of 3 clusters at Kareband.

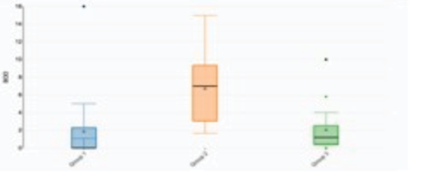

Figure 22.Box plot comparing BOD of 3 clusters at Kareband.

Figure 23.Box plot comparing FC of 3 clusters at Kareband.

Figure 24.Box plot comparing FC of 3 clusters at Kareband.

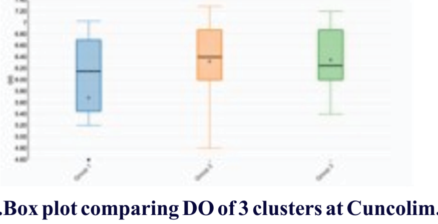

Figure 25.Box plot comparing DO of 3 clusters at Cuncolim.

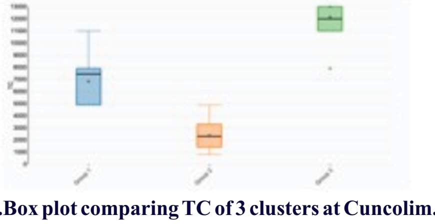

Figure 26.Box plot comparing TC of 3 clusters at Cuncolim.

 PRINT ISSN No. 2277 - 8179 | DOI : 10.36106/ijsr

Figure 27.Box plot comparing BOD of 3 clusters at Cuncolim.

Figure 28.Box plot comparing FC of 3 clusters at Cuncolim.

Applying K-Means algorithm we decide the data into 3 clusters. We cluster the data at each location for the period of 4 years i.e 2014-2017.The parameters that we have considered for this clustering includes BOD,DO,TC and FC. The tolerable limit of DO is greater than 4 for BOD iT IS less than 3 for FC it is less than 2500 and TC it is less than 5000.

It is observed that at location Orlim cluster 3 the min value of DO is 4and max value id 14.1 ,the min value for BOD is 0 and max value is 12.5,the min value of FC is 0 and max value is 2300 ,and for TC the min value is 0 and max value is 1300.

When we post the box plot we can see that cluster 3 comes in the satisfactory condition i.e all the paramerts are in under the tolerance levels. That indicates low levels of pollution which implies a healthy ecosystem surrounding the river location .

At the second location Mobor we observe that cluster 3 has the values of all paramerts under tolerance level .Hence we can conclude that the water in satisfactory. Which implies very low levels of pollution.

At the third location Kareband we observer that the the values of FC and TC are above the tolerance limit that above 2500 and 5000 respectively. Here we can imply that the water is highly polluted.

At the Fourth location i.e Cuncolim we observe that cluster 2 the values of cluster 2 comes under the satisfactory condition which implies water pollution is under control at this location. As compared to Orlim and Mobor, the values of the paramerts are not completely satisfactory .

Therefore we can conclude that the pollution levels of at river Sal are satisfactory at Orlim and Mobor .At Kareband the levels of pollution very high. At Cuncolim it is at a moderate level .

III. CONCLUSION.

In this project firstly we have found the WQI of river Sal at 4 different locations. We have created a summarization in which months we have high and low levels of pollutions. We can conclude that pollution depends on the climate factors also .We have seen that during the monsoons we see decease in the levels of pollution as compared to during other time of the year.

Using K-Means we have seen that by using the most important paramerts i.e Dissolved oxygen ,Bio-chemical oxygen demand, Total coliform and Fecal Coliform we can cluster the data .These paramerts are can used to determine if the pollution levels are satisfactory or not satisfactory. The health of the environment mostly depends on these parameters .If there is sufficient dissolved oxygen the aquatic life will be healthy. Also the micro-organisms will have sufficient oxygen to decompose .If the water is pollution free or less pollution that implies less breakout of disease ,a healthy river will lead to healthy environment.

IV. REFERENCES

1. An Approach for classification of Health Risks Based on Air Quality Levels Ranjana Waman Gore, Computer Science and Engineering MArathwada Institute of technology for Aurangabad, Maharashtra
2. River Water Pollution Analysis using High Resolution Satellite Images - A Survey, Hinal J.Patel Dept. of Information Technology Dharmsinh Desai University, Nadiad, India
3. Water Quality Index Using IOT Rajat Verma1, Laxmi Ahuja2, Sunil Kumar Khatri3Amity Institute of Information Technology, Amity University Uttar Pradesh, Noida, India
4. Data Analysis, Quality Indexing and Prediction of Water Quality for the Management of Rawal Watershed in Pakistan . Maqbool Ali, Ali Mustafa Qamar Department of Computing School of Electrical Engineering and Computer Science (SEECS) National University of Sciences and Technology (NUST) Islamabad, Pakistan
5. Determination of Water Quality Index and Suitability of an Urban Waterbody in Shimoga Town, Karnataka K. Yogendra* and E.T. Puttaiah Department of P.G. Studies and Research in Environmental science
6. Physico-chemical parameters for testing of water – A review Patil. P.N, Sawant. D.V, Deshmukh. R.N Department of Engineering Chemistry, Bharati Vidyapeeth's College of Engineering