JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

"WATER QUALITY ANALYSIS OF NIRA RIVER "

[1]Dr. Neeta Kishor Dhane, [2]Dr. Vaishali Vilas Patil
[1]Assistant Professor , [2]Assistant Professor
[1]Department of Statistics,
[1]Tuljaram Chaturchand College, Baramati, India

Abstract : Water is one of the important aspects of human life and rivers are one of the main sources of water but rivers become contaminated due to chemical waste, human activities, etc. Nira is a river that flows through the Indian state of Maharashtra. It is a tributary of the Bhima River and flows through the Pune and Solapur districts of Maharashtra. The study is based on secondary data related to the quality of water from the Nira river. The source of data is https://www.mpcb.gov.in/water-quality/pune/17. Naïve Bayes technique is used to predict water quality. After comparing the water quality of Nira river it is observed that 100% of Nira river water was nonpolluted during the covid 19 lockdown period.

Keywords: **Nira river, Naïve Bayes.**

INTRODUCTION

The Nira river rises in the western ghats of Pune district, flows through Bhor taluka, Shirwal Taluka Satara District, Solapur District, and reaches the Bhima Basin at Nira Narsingpur near Akluj, and then meets the Bhima Basin at Nira Narsingpur near Akluj. It is a tributary of the Bhima River that passes through Maharashtra's Pune and Solapur districts. Between 180 13.528' N and 170 58.237' N, and 730 32.357' E and 750 8.458' E, the Nira river basin is located. The drainage area of the basin is 6,879.60 km2. With average temperatures ranging between 20 and 28 °C (68 and 82 °F), the Nira basin has a hot semi-arid climate bordering on tropical wet and dry. Most of the 722 mm (28.43 in) of annual rainfall in the city fall between June and September, and July is the wettest month of the year.

DATA: - This dataset consists of monthly data of water from Nira river from the year 2008 to 2020.

Water Quality Parameters: - In this project **6** water quality parameters are involved.

1) **pH**:- pH is a measure of how acidic/basic water is. The range goes from 0 to 14, with 7 being neutral.
2) **D.O:-** Dissolved Oxygen(DO) is a measure of how much oxygen is dissolved in the water -the amount of oxygen available to living aquatic organisms. The amount of dissolved oxygen in a river can tell us a lot about its water quality.
3) **B.O.D:-** Biological Oxygen Demand (B.O.D) is a measure of oxygen required to remove waste organic matter from water in the process of decomposition aerobic bacteria. B.O.D is used, often in wastewater-treatment plants, as an index of the degree of organic pollution in water,
4) **C.O.D: - The** Chemical Oxygen Demand (COD) is a measure of water and wastewater quality. The COD test is often used to monitor water treatment plant efficiency.
5) **Nitrate: -** Basically, any excess nitrate in the water is a source of fertilizer for aquatic plants and algae. In many cases, the amount of nitrate in the water is what limits how much plants and algae can grow.
6) **Fecal Coliform: -** Fecal Coliform is a bacteria in aquatic environments that indicated that the water has been contaminated with the fecal material of man or another animal. I indicate the presence of sewage contamination of a waterway and the possible presence of other pathogenic organisms.

Also this data is recorded on 5 water stations of Nira River as, Shindewadi ,Sangavi , U/s of Jubilant Organosis Pune , Sarola Bridge and D/s of Jubilant Organosis Pune .

We first pre-processed the data by substituting missing values with their means and using the Mardia test to ensure that they were normal. The boxplot is used to check for outliers and to see if the data is balanced. We calculate the WQI (Water Quality Index) which summarizes water quality information in a readily-understood format. If WQI lies between 50-100 then water is classified as non-polluted, if WQI lies between 38-50 then water is classified as polluted and if WQI is 38 and less then water is classified as heavily polluted. We then categorize the WQI as heavily polluted, non-polluted, and polluted. WQI represents the best means to communicate and categorize water-quality levels in assessments of water suitability for various applications. Then we perform exploratory analysis for better visualization of our data. In this exploratory analysis, we calculate descriptive analysis, correlation matrix.

We balance the train data via sampling after splitting the data into training and testing datasets. The model is then fitted to the training dataset using Naïve based model and the confusion matrix is created to test it on the test dataset.

MS-Excel and R Studio are used to conduct the statistical analysis.

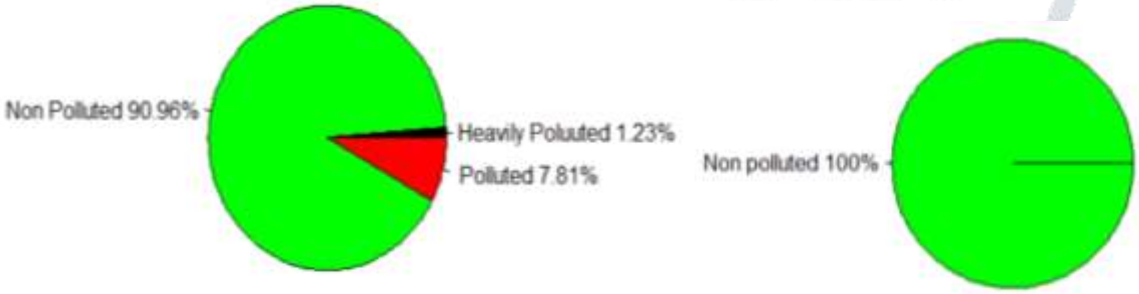
SOURCE OF DATA: - https://www.mpcb.gov.in/water-quality/pune/17

Descriptive Statistics of Dataset: -

pH	DissolvedOxygen	B.O.D	C.O.D	Nitrate	FecalColiform
Min. :4.200	Min. :0.860	Min. : 1.300	Min. : 4.00	Min. :0.0010	Min. : 2.00
1st Qu.:7.820	1st Qu.:4.800	1st Qu.: 4.600	1st Qu.: 16.00	1st Qu.:0.2895	1st Qu.: 65.48
Median :8.100	Median :5.585	Median : 6.400	Median : 20.00	Median :0.7575	Median :140.00
Mean :8.069	Mean :5.372	Mean : 6.906	Mean : 21.64	Mean :1.1379	Mean :154.12
3rd Qu.:8.300	3rd Qu.:6.100	3rd Qu.: 8.325	3rd Qu.: 24.00	3rd Qu.:1.5000	3rd Qu.:225.00
Max. :9.500	Max. :7.800	Max. :46.400	Max. :104.00	Max. :9.8000	Max. :900.00

Comparison of water quality before covid-19 and during covid-19:-

Before Covid-19 **During Covid-19**



Conclusion: - It can be clearly observed that before covid-19, 90.96% of water was nonpolluted, 7.81% of water was polluted and 1.23% of water was heavily polluted whereas see that 100% Nira river water is nonpolluted during covid 19 period.

Correlation Matrix: -

	pH	DissolvedOxygen	B.O.D	C.O.D	Nitrate	FecalColiform
pH	1.00000000	-0.13855425	0.09002088	0.113514106	0.104873781	0.02189539
DissolvedOxygen	-0.13855425	1.00000000	-0.43991566	-0.427300435	-0.057935114	0.03944647
B.O.D	0.09002088	-0.43991566	1.00000000	0.745960564	-0.046050588	0.01990596
C.O.D	0.11351411	-0.42730043	0.74596056	1.000000000	0.001978704	0.03226599
Nitrate	0.10487378	-0.05793511	-0.04605059	0.001978704	1.000000000	0.13679321
FecalColiform	0.02189539	0.03944647	0.01990596	0.032265986	0.136793208	1.00000000

Conclusion: - From the above graph, it is seen that B.O.D and C.O.D. are highly correlated.

Testing Normality: -

Mardia test for multivariate normality: -

	Test	Statistic	p-value	Result
1	Mardia Skewness	7264.14628633036	0	NO
2	Mardia Kurtosis	160.233019717851	0	NO
3	MVN	<NA>	<NA>	NO

H0: Data comes from a multivariate normal distribution.

H1: Data doesn’t come from a multivariate normal distribution.

Here P-value for Mardia skewness and kurtosis is 0

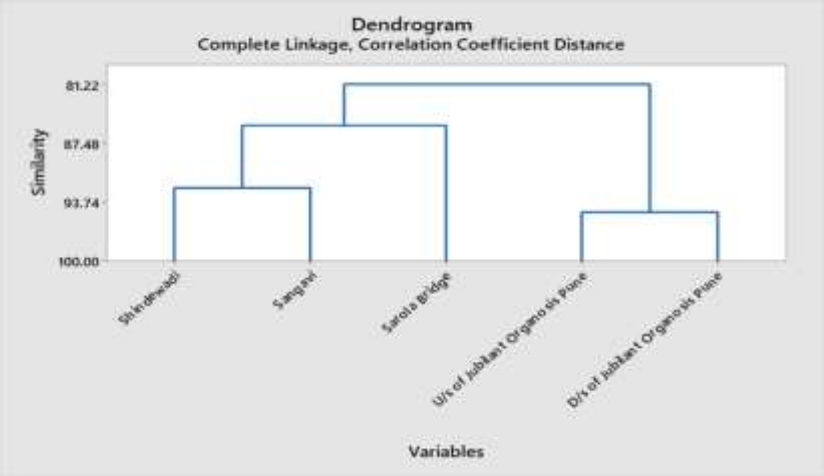
Conclusion: Data doesn’t come from a multivariate normal distribution.

Checking similarities between Nira river stations: -

Correlation Coefficient Distance, Complete Linkage

Amalgamation Steps

Step	Number of	Similarity	Distance	Clusters	New	Number of obs	
	Clusters	level	level	joined	cluster	in new cluster	
1	4	94.7857	0.104287	3	5	3	2
2	3	92.1954	0.156093	1	2	1	2
3	2	85.5739	0.288522	1	4	1	3
4	1	81.2185	0.375630	1	3	1	5

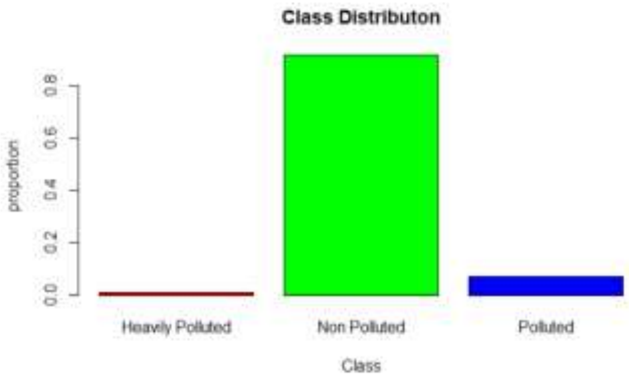


Conclusion: - Above dendrogram has used correlations between stations to identify the similarities between them. Here we can see that no two stations are 100% similar. The stations Shindewadi , Sangavi and Sarola Bridge shows greater than 80% similarity. Also, stations U/s of jubilant Organosis Pune & D/s of jubilant Organosis Pune shows greater than 90% similarity. So within our dataset we have identified two clusters {Shindewadi, Sangavi , Sarola Bdridge} & {U/s of jubilant organosis Pune & D/s of jubilant organosis pune}.

Data Partition

Our train data set have 624 observation and test data set have 156 observations.

Bar plot of class distribution:-



Conclusion: - Above graph highlights class imbalance. So, it highlights that proportion of non-polluted water is highest and proportion of heavily polluted water is least.

Balancing the data: -

Up-sampling: -

Heavily Polluted	Non Polluted	Polluted
571	571	571

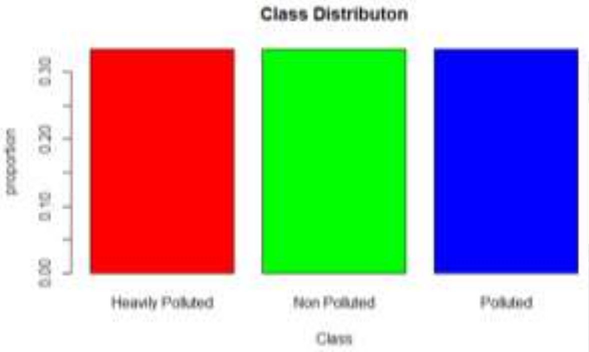
Here we can see that after balancing the train data our train dataset has an equal no of water samples for each class i.e 571 water samples.

Proportion table of balanced data: -

Heavily Polluted	Non Polluted	Polluted
0.3333333	0.3333333	0.3333333

Here we can see that our water is equally polluted, heavily polluted and nonpolluted i.e 33.33%

Bar plot of balanced data:-



Conclusion: - This graph highlights class balance. That is proportion of all classes are equal.

Data mining algorithms for classification: -

Naive Bayes Model

Conclusions: If the approximate value of pH is 8.64, Dissolved Oxygen is 2.04, B.O.D is 19, C.O.D is 36.18, Nitrate is 0.83 and Fecal Coliform is 146.14 then our water is heavily polluted. If the approximate value of pH is 8.04, Dissolved Oxygen is 5.59, B.O.D is 6.39, C.O.D is 20.59, Nitrate is 1.3 and Fecal Coliform is 212.81 then our water is nonpolluted. If the approximate value of pH is 8.31, Dissolved Oxygen is 3.21, B.O.D is 11.26, C.O.D is 30.22, Nitrate is 1.69 and Fecal Coliform is 199.87 then our water is polluted. For pH , Dissolved Oxygen and Nitrate sd is close to zero i.e low for all classes it indicates that data points tend to be very close to the mean. For B.O.D , C.O.D and Fecal Coliform sd is not close to zero i.e high for all classes it indicates that data points are spread out over a large range of values.

Prediction of naïve bayes model :-

	Heavily Polluted	Non Polluted	Polluted	pH	DO	B.O.D	C.O.D	Nitrate
1	0.9922590	9.349890e-08	7.740905e-03	8.95	1.84	12.8	36	0.19
2	0.3739438	1.621687e-04	6.258941e-01	8.68	2.45	11.2	32	2.80
3	0.9078054	2.068107e-06	9.219249e-02	8.90	2.50	13.5	40	0.70
4	1.0000000	3.629835e-48	1.785770e-10	8.41	0.86	46.4	36	0.11
5	0.9928741	8.711657e-16	7.125886e-03	7.96	2.40	28.0	36	0.14
6	0.9940053	1.002287e-08	5.994710e-03	8.80	1.40	13.0	40	1.10

	FecalColiform	Class
1	225.00	Heavily Polluted
2	195.00	Heavily Polluted
3	84.22	Heavily Polluted
4	70.00	Heavily Polluted
5	170.00	Heavily Polluted
6	275.00	Heavily Polluted

Conclusions: -These are first few rows of our balanced train data. There is 99.2% chance that water is heavily polluted for 1st water sample it maybe because of all parameter values are not within the permissible range. Similarly for other 5 samples, they also belong to heavily polluted class.

Confusion Matrix:-

Prediction	Heavily Polluted	Non Polluted	Polluted
Heavily Polluted	0	0	1
Non Polluted	0	134	0
Polluted	2	9	10

Overall Statistics

Accuracy: 0.9231

95% CI : (0.8695, 0.9596)

No Information Rate : 0.9167

P-Value [Acc > NIR] : 0.4583

Kappa : 0.6211

Mcnemar's Test P-Value : NA

Statistics by Class:

	Class: Heavily Polluted	Class: Non Polluted	Class: Polluted
Sensitivity	0.00000	0.9371	0.90909
Specificity	0.99351	1.0000	0.92414
Pos Pred Value	0.00000	1.0000	0.47619
Neg Pred Value	0.98710	0.5909	0.99259
Prevalence	0.01282	0.9167	0.07051
Detection Rate	0.00000	0.8590	0.06410
Detection Prevalence	0.00641	0.8590	0.13462
Balanced Accuracy	0.49675	0.9685	0.91661

Conclusions: Here accuracy is 92.31%. Accuracy shows how accurate is the naïve bayes classifier in predicting whether water is polluted, heavily polluted or non-polluted with actual data. So, when training data is about 80% then the accuracy achieved is about 92.31%.

CONCLUSIONS:

1. The stations Shindewadi , Sangavi and Sarola Bridge shows greater than 80% similarity. Also, stations U/s of jubilant Organosis Pune and D/s of jubilant Organosis Pune shows greater than 90% similarity
2. 91.54% of Nira river is non polluted , 7.31% of Nira river is polluted and 1.15% of Nira river is heavily polluted.
3. Before covid 19 (in our data from Jan 2008 to Feb 2020) 90.96% of Nira river was non-polluted and during covid 19 (in our data from March 2020 to dec 2020) 100% Nira river water was non-polluted. So as the coronavirus induced lockdown and reduced industrial activities, the water quality of Nira river has improved.

REFERENCES

[1] Mosleh Hmoud Al-Adhaileh and Fawaz Waselallah Alsaade "Modelling and Prediction of Water Quality by Using Artificial Intelligence" Sustainability 2021, 13, 4259. https://doi.org/10.3390/su13084259

[2] Theyazn H. H Aldhyani , Mohammed Al-Yaari, Hasan Alkahtani, and Mashael Maashi

[3] "Water Quality Prediction Using Artificial Intelligence Algorithms" Hindawi Applied Bionics and Biomechanics Volume 2020, Article ID 6659314

[4] Kumar, Pradip | Kaushal, Rajendra Kumar | Nigam, Anjani K. "Assessment and Management of Ganga River Water Quality Using Multivariate Statistical Techniques in India: Asian Journal of Water, Environment and Pollution, vol. 12, no. 4, pp. 61-69, 2015

[5] Liya Fu , You-Gan Wang "Statistical Tools for Analyzing Water Quality Data" DOI: 10.5772/35228 · Source: InTech

[6] Sanjeev Gour1, Mamta Gour2 "Study on Water quality of Narmada River by analyzing physicochemical and biological parameters using random forest model "International Journal of Computer Sciences and Engineering Open Access Research Paper 7(1), Jan 2019 E-ISSN: 2347-2693

[7] Shailesh Jaloree , Anil Rajput, Sanjeev Gour " Decision Tree approach to build a model for water Quality" Binary Journal of Data Mining & Networking 4 (2014) 25-28

[8] https://www.mpcb.gov.in/water-quality/pune/17 Water Quality Status Of Maharashtra 2018-19