IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-7, Issue-1, Jan 2019 E-ISSN: 2347-2693

Study on Water quality of Narmada River by analyzing physicochemical and biological parameters using random forest model

Sanjeev Gour[1*], Mamta Gour[2]

[1]Department of Computer Science, Career College, Bhopal, India
[2]Department of Chemistry, Govt. Nutan Girls College, Bhopal, India

Corresponding author: sunj129@gmail.com

Available online at: www.ijcseonline.org

Accepted: 06/Nov/2018, Published: 31/Jan/2019

Abstract- Rivers are most important recourse for consumption as freshwater for human being. The quality of water resource can be measured and indentified by its physicochemical and biological parameters. This paper presents a study of water quality of River Narmada through analysis of some of its selected physicochemical and biological parameters. For this study, secondary dataset with 7 water quality parameters collected during years from 1990 to 2012 of River Narmada of Hoshangabad district of Madhya Pradesh has been used. The objective of this study to assess the quality of water of Narmada River in comparison with Surface Water Quality Standards for Indian Rivers. For this assessment, one of algorithm of data mining called random forest is used which is incorporate as built in model of an analytical tool called Rapid Miner. This study show that quality of river water is not favorable and goes down day by day due to domestic and industrial effluents mixing in the river at town. The finding of this study gives definitely an alert message to water resource managers for the improvement of water quality management by proper treatment and monitoring the River.

Keywords- Water Quality, Narmada River, Classification, Random Forest model.

I. INTRODUCTION

Today there are many management and treatment policies of freshwater for drinking have been developed recently [1]. As the quality of water is highly variable with respect to time and space and affected due to both human and natural factors, it is big challenge to monitoring and managing the water quality plan for respective authority. Many advanced techniques are using for analyzing large number of database of water quality of river collected on daily basis [2]. Narmada River is considered a lifeline of Madhya Pradesh and Gujarat states. Many studies indicates that quality of water of Narmada is getting poor and found unsuitable for drinking at some places due to significant alteration of physicochemical and biological parameters like Temperature, pH, BOD/COD ratio, ammonium nitrogen and nitrate etc. [3,4,5].

With this fact, authors have tried to analyzing these water quality parameters of Narmada River using one of Data mining method and extracted much useful information about the quality of water. For this study, a large secondary water quality dataset of River Narmada have collected in Excel form from Central Water Commission laboratory Bhopal and analyzed using one of analytical software tools named Rapid Miner. This tool generates a Random forest model in terms of random or decision trees.

Authors have interpreted all generated random trees with considering all main physicochemical and biological parameters of water as primary predictor variable. The results extracted from the model may be very useful to understand the behavior of quality of river water in perspective to some of physicochemical and biological parameters of water.

II. DATA MINING TECHNIQUE

Data mining is nowadays a one of emerging techniques if data science for analyzing big dataset in many data centric applications because its methods have great capability to discover the hidden pattern from the large dataset [6]. This is one of the main reasons that the analysts (authors) use this method for this study.

II.I CLASSIFICATION

The Classification process of Data Mining is about predicting the target class (i.e. classify which category) on the basis of existing data for which category membership is already known. In this way, this process builds a model that predicts the category of new data by using learned and trained rules. One of such algorithm for classification is random forest, which is used for this study.

II.II RANDOM FOREST ALGORITHM

Random forest algorithm is a supervised classification algorithm. This algorithm builds the forest with a number of decision trees. The set of these decision trees that uses subspace sampling is known as a random forest. It is widely used in many application areas such as Banking, Medicine, Stock market and E-commerce. Random forest model break into number of decision trees randomly and each node of a tree represents a splitting rule for one specific Attribute. Each random tree will predict different target or category class (outcome) for the same rule and by considering each predicted class; analyst can extract patterns and relationships between attributes.

III. ABOUT RAPID MINER STUDIO

Rapid Miner Studio is a popular unified platform for data mining and predictive analytics in any data centric domain. It provides an easy visual design environment for making various analytics models with their built in machine learning algorithms. It has a huge collection of various machine learning algorithms, data preparation, exploration functions and visualization tools for analysis of various kind data. Rapid miner supports all kind of data mining procedures so this is why analyst used the latest version of Rapid Miner Studio for his analytical study.

IV. ABOUT DATA SET

For this study, secondary data of water quality of river Narmada of one of district of Madhyapradesh named Hoshangabad is used. Analyst (author) has collected this dataset from the Laboratory of Central Water Commission (CWC) State Baseline Bhopal. The dataset was in MS-Excel format and in form of summary report of water quality parameters for the period of year 1990 to 2012. Sample data table and format is given in Table 1.

Table 1– Secondary water quality dataset

Year	pH_GEN	Temp	BOD	COD	DO	NO3-N	O_PO4	NH3-N	Class Value
1990	8.2	26.6	0.6	21.6	7.1	0.54	0.043	0.03	A
1991	8	27.1	0.9	0.86	7.4	0.16	0.021	0.01	B
1992	8.2	27.1	0.6	0.87	8.1	0.49	0.01	0.02	A
1993	8	27.3	1.6	39.1	6.3	0.61	0.027	0.03	A
1994	8.3	28.2	1.5	24.3	6.35	0.28	0.012	0.02	A
1995	8.3	26.3	1.6	27.6	7.5	0.58	0.021	0.01	B
1996	8.3	29.4	0.9	28.3	7.4	0.47	0.048	0.015	B
1997	8.1	27.3	1	15.7	7.4	0.16	0.01	0.16	C
1998	8.3	27.3	0.7	22.6	6.9	0.71	0.039	0.03	B
1999	8.1	28.9	1.6	18	7.9	0.34	0.025	0.05	C
2000	8.1	26.7	1.2	27.9	7.1	0.58	0.019	0.02	A
2001	8.2	27.3	1.1	30	6.9	0.31	0.023	0.02	A
2002	8.3	27.6	1.4	25.3	6.8	0.51	0.017	0.05	B
2003	8.3	28	1.3	25.9	6.2	1.04	0.148	0.052	B
2004	8.2	27.7	1.4	35.4	6.1	0.28	0.086	0.06	B
2005	8.1	27.8	1.4	44.4	6.1	2.41	0.138	0.078	B
2006	8.2	27.1	1.2	29.2	6.1	1.89	0.085	0.065	C
2007	8.1	28.6	1.3	33.8	6.1	2.29	0.095	0.36	C
2008	8.1	27.9	1.8	39.6	7.1	1.81	0.405	0.05	C
2009	8.2	28.2	1.2	39.2	6.1	0.52	0.151	0.8	C
2010	8.2	27.3	1.2	26.4	6.1	1.04	0.03	0.854	D
2011	8.3	29.5	0.6	26.2	6.1	1.49	0.152	0.458	D

IV.I ABOUT NARMADA RIVER

Narmada River (IHD book, CWC: 2009) is one of the 13 prominent rivers of India, about 98,797 sq km of total water-shed area is covered by it [7]. Narmada is considered as the lifeline and west flowing river, in the state of Madhya Pradesh. The water quality of river Narmada which is fifth largest river in the Indian subcontinent get worse from past decades due to sewage waste water of Paper Mill industry and therefore concentration of industrial waste chemicals increases pollutants in the Narmada River [8,9]. So these facts encourage to analysts (authors) to study more about the quality of surface water of Narmada River [10, 11, and 12].

IV.II ABOUT WATER QUALITY PARAMETERS

Water's nature is of course, dynamic and its quality varies with respect to time and space. In order to analysis the quality of water, it is necessary to consider some the major physic-chemical and biological components of water. These physical, chemical, or biological components affect water quality. Table-2 gives a summary of some of the more common water quality parameters which are used to assess the quality of water for this study.

PARAMETER	UNIT
PHYSICAL	
pH_GEN	pH units
Temp	deg C
CHEMICAL	
NH3-N	mg N/L
NO3-N	mgN/L
O-PO4-P	mg P/L
BIOLOGICAL/BACTERIOLOGICAL	
BOD3-	mg/L
COD	mg/L
DO	mg/L

Table 2- Water quality parameters

IV.III DESCRIPTION OF CLASS ATTRIBUTE

The entries for class label (i.e. predicted target class) parameter "Class Value" in dataset are calculated or measured on the basis of Surface Water Quality Standards (as per IS: 2296) [13] . According to it, surface water quality can be divided into four classes; Class A - Extra clean fresh surface water resources use for conservation which are not necessary to pass through the water treatment processes, Class B- Very clean fresh surface water resources use for consumption that require ordinary water treatment processes, Class C - Medium clean fresh surface water resources use for consumption, but should be passed through an ordinary treatment process before use; and last Class D- Fairly clean fresh surface water resources use for consumption, but requires special water treatment processes before use.

V. EXPERIMENTAL PROCESS

From the 'Repository' panel of Rapid Miner, import the dataset from local disk and set the format of data compatible for the prediction process. Also assigns the role and type of all water quality parameters of loaded dataset. In this experiment the role of water quality parameter 'Pollution class' is set to be class label (target class) and polynomial as type. The types of all column parameters are set to be real except year.

Now store the dataset in local repository as CSV file. It is time to parameter setting for random forest modeling like number of tree, splitting criteria, confidence, minimal depth, minimal gain, leaf size etc. by pre-pruning. Now retrieve the dataset from repository panel by dragging in the middle of screen and now create process by adding and connecting operators. Operators are connected together via their Ports. In this experiment Rapid Miner analyst connect the output port of dataset with input port of process called Random forest and connect the output port of process to result port on the right side of selected process panel by dragging the lines between ports. (Figure-1). After connected the elements of the process, analyst execute the process by pressing the button 'Run' and the operators in process will perform their actions i.e. the set of random tree are generated at its result window.

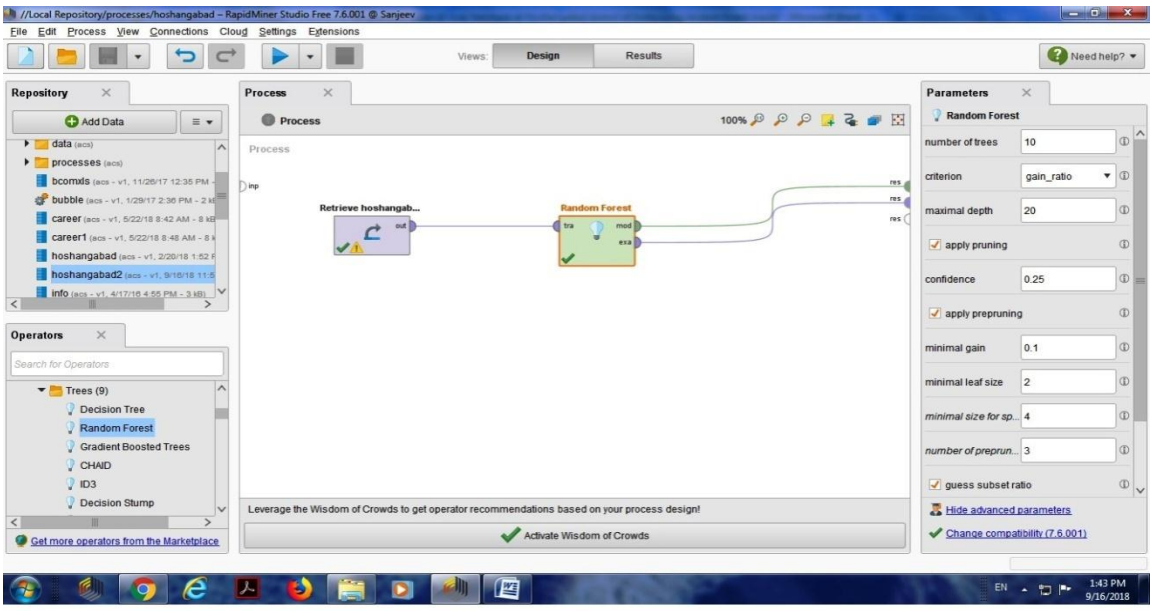

Figure 1–screen shot for design view in Rapid Miner.

VI. EXPERIMENT RESULTS

Random forest model generate exact 10 Decision/Random trees for exploratory analysis of large data sets of surface water of river Narmada .One of tree is shown in figure-2.This can help us to predict the target class and extract the patterns in water quality parameters. Some of extracted patterns or predictive relationship from model described here as results:

The interpretation of random tree 1 and 2 where "Temperature" is the predictor variable for primary split in tree1 and "Year "is for tree 2 respectively, shows that after year 1994-95 the quality of surface water gradually goes down two levels from A to C due to increasing temperature in surface water and the heat water getting out from the factories may be one of main causes for this.

It has clear observed from random tree 3 that ammonia nitrogen (nh3-n) in figure 2 is the most influential parameter among all rest parameters as the quantity of ammonia nitrogen in surface water exceed the particular point (in this case 0.038 mg/l), the quality of water is goes down from class A to D. (i.e. Best to worst). This tree also shows one of interesting relationship of ammonia nitrogen with pH value. With little variation of ammonia nitrogen (up or down), the increasing value in pH will take the quality of water towards poor from one level.

Nitrate nitrogen (no3-n) is another important influential parameter as the quantity of Nitrate nitrogen in surface water exceed the particular point (in this case 1.265 mg/l), the quality of water is goes down from class A to D. (i.e. Best to worst). Some interesting analysis can also be made using charts generated by Rapid Miner. (Figure-3)

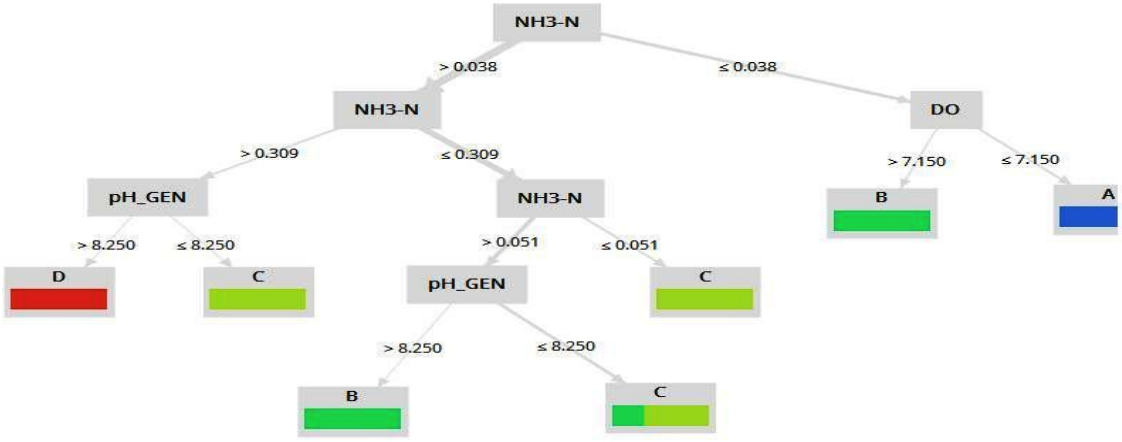

Figure 2- One of random trees generated by Rapid Miner

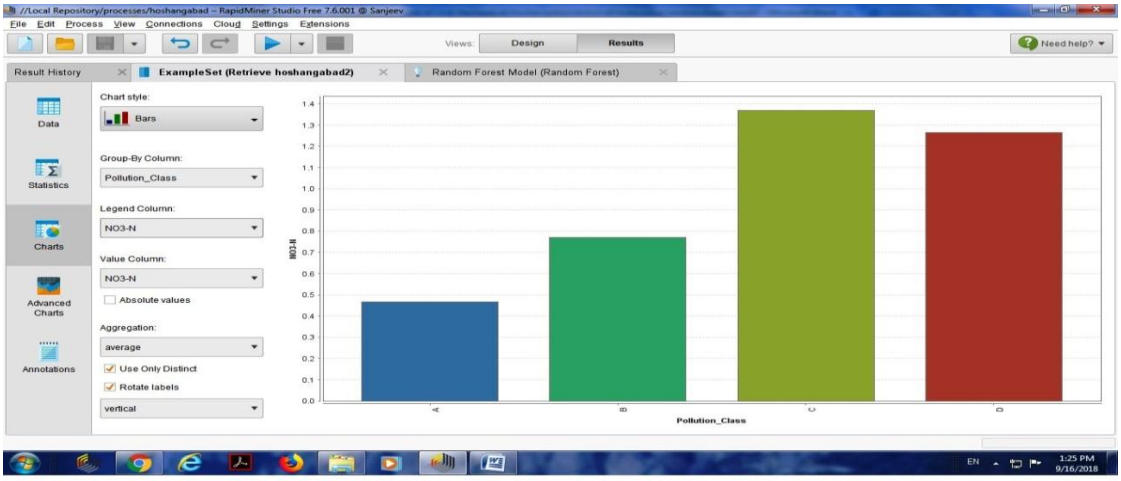

Figure 3- Bar-Chart between two parameters generated by Rapid Miner

BOD and COD both have been widely adopted as a measure of pollution effect. In this way random tree 5 gives a better understanding the quality of water as it clearly showed the relationship between these two parameters which indicates and measure the oxygen levels in the water. The quality of water went down from A to C as COD level exceed form 39.15 mg/l. Similarly water got much polluted after BOD level increased from 1.5 to 2.0 mg/l.

One of most parameter to measure the quality of water is dissolve oxygen (DO). It is the actual amount of oxygen available in dissolved form in the water. When the DO drops below a certain level, the quality of water unable to continue at a good level. From the model, It is has been clearly seen that the DO level of water of Narmada river continuously decreased from 7.1 to 6.1 mg/l in period of 1990 To 2012. Many such other useful observations and relationships for water quality have been discovered for Narmada River.

VII. CONCLUSION

The result of model generated by Rapid miner tool using random forest algorithm for water quality of Narmada River is presented in the previous section. The observations mined by analyzing different physicochemical and biological parameters of water show that quality of surface water of Narmada is gradually down due to many natural and human activities.

Water temperature is increasing as warm water is discharged into river from the industries. Ammonium nitrogen, nitrate and phosphates components coming in river water from various sources including human, animal and factory sewage. COD/BOD ratio in water gradually increased by the means of domestic sewage as well for septic sewage.
Oxygen levels also goes down through over fertilization of water plants by run-off from farm fields containing phosphates and nitrates. It is urgent need to establish a water treatment plant at Hoshangabad district so that we can save the river Narmada which is a lifeline of Madhya Pradesh. So the finding of this study gives definitely an alert message to water resource managers for the improvement of water quality management by proper treatment and monitoring the River.
Further research study is recommended to analyzing other water quality parameters which directly affects the quality of River water.

ACKNOWLEDGEMENT

We sincerely acknowledge the Central Water Commission India, Bhopal for providing valuable data regarding the water quality of River Narmada. We also have to express my appreciation to authors mentioned in following reference section for referring the part of their reports, surveys, observations, book and research manuscripts during the study of this research although any errors are my own and should not tarnish the reputations of these esteemed professionals.

REFERENCES

[1] Sharma Deepshikha, Kansal Arun, "Water Quality analysis of River Yamuna using water quality index in the national capital territory India", Appl. Water Science, Vol. 1, pp. 147-157, 2011.
[2] Mohamad Ali Fulazzaky, " Water Quality Evaluation System to Assess the Brantas River Water", Water Resour Manage, Vol.23, pp.3019–3033, 2009.
[3] Sharma Shraddha, Vishwakarma Rakesh, Dixit Savita and Jain Praveen," Evaluation of Water Quality of Narmada River with reference to Physicochemical Parameters at Hoshangabad city, MP, India ", Vol. 1,Issue 3, pp.40-48,2011.
[4] Gour Sanjeev ,Gour Mamta," Neural Network Approach In Water Quality Data Analysis For The River Narmada", Binary Journal of Data Mining & Networking, Vol. 4, pp. 49-53, 2014.
[5] Soni Virendra, Khwaja Salahuddin and Visavadia Manish, "Preimpoundmental studies on Water Quality of Narmada River of India", International Research Journal of 12 Environment Sciences, Vol. 2,Issue 6, pp.31-38, 2013.
[6] Shailesh Jaloree, Anil Rajput, Sanjeev Gour," Decision tree approach to build a model for water quality", Binary Journal of Data Mining & Networking, Vol. 4,pp.25-28,2014.
[7] Report, Central Water Commission (CWC)," Integrated Water Hydrological book", Narmada Basin Organization Bhopal. September 2009.
[8] Singh Pawan and Shrivastava Pradeep, "Analysis of Water Quality of River Narmada", International Journal of Current Research, Vol 7,Issue 12, pp. 24073-24076, 2015.
[9] Jaloree Shailesh, Gour Mamta, Gour Sanjeev," Water Quality Assessment using Association Rule Mining for River Narmada", Indian Journal of Science and Technology**,** Vol 9,Issue 10 ,pp.1-5, 2016.
[10] Kushram Parvati," Physico-chemical studies on Narmada River water at Dindori (M.P.) India", International Journal of Applied Research, Vol.2,Issue 2, pp. 226-228, 2016.
[11] Jain Dilip, Rahi D.C, Verma Shailza, " Assessment of Water Quality Status by Using Water Quality Index (WQI) Method in Narmada River in Jabalpur Madhya Pradesh", International Journal for Scientific Research & Development, Vol.4,Issue 12, pp.122-124,2017
[12] Katakwar Mukesh," Water quality and pollution status of Narmada river s Anjan Tributary in Madhya Pradesh, India" , International journal of current Research an Academic Review, Vol. 2,Issue 11, pp. 93-98, 2014.
[13] Report, **"**Surface Water Quality Standards" (as per IS: 2296), Chapter-6, Table 9.1 pp.245-252 and Central Pollution Control Board, 2007-08.

AUTHORS PROFILE

Dr. **Sanjeev Gour** pursed Bachelor of Computer Science (Honors) and Master of Science (Electronics & Comm.) from Devi Ahilya University Indore and Master of Science Computer Science) from Barkatullah University Bhopal. He is completed his Ph.D. in Computer Science and currently working as a Associate Professor in Department of Computer Science in Career College, Bhopal. He was member of Board of Studies & Examination Committee of Computer Science in Barkatullah University, Bhopal. He was a member of Managing Committee in Computer Society of India (Bhopal Chapter) 2017-18. He has published more than 15 research papers in reputed international journals including Thomson Reuters & Scopus (SCI & Web of Science).His main research work focuses on Data Mining and Analysis. He has 15 years of teaching experience and 5 years of Research Experience.

Dr. Mamta Gour pursed **B**achelor of Science in Biology and Master of Science in Chemistry from Devi Ahilya University Indore. She is completed his Ph.D. in Chemistry Science from Barkatullah University Bhopal and currently working as a Guest Faculty post against Asst. Professor in Department of Chemistry in SNGGAPG (Nutan) College Bhopal. She has published more than five research papers in reputed international journals.

