

MULTIVARIATE DATA MINING TECHNIQUES FOR ASSESSING WATER POTABILITY

Durga Karthik^{1,*} and K.Vijayarekha²

¹School of CSE, SRC, SASTRA University, Kumbakonam, India.

²School of EEE, SASTRA University, Thanjavur

*E-mail: durgakarthik@src.sastra.edu

ABSTRACT

Water is required for all living organisms of which 1.7% is available as ground water. Water has no calories or any nutrients, but essential for various metabolic activities in our body. Chemical and physical parameters can be tested for identifying the potability of ground water. Electrical conductivity, pH, alkalinity, Total Alkalinity, TDS, Calcium, Magnesium, Sodium, Potassium, Chloride, and Sulphate of the ground water from three districts: Thanjavur, Nagapattinam and Ariyalur of south Tamilnadu were analyzed. Our aim was to check, if the ground water from the above areas are potable or not. As multivariate are present, Principal Component Analysis with Data mining technique using JRIP rules was employed for classifying the ground water.

Keywords: Multivariate, Potability, Ground Water, Data mining, JRIP, PCA, classification

©2014 RASĀYAN. All rights reserved

INTRODUCTION

‘Aapaha Punanthu Prithveem’ in Sanskrit means let abundant water purifies the Earth. Water is required for all activities on earth. It is a known fact that the existence of life on earth is due to the presence of water. Due to increase in population, industrialization all the natural resources are polluted, nowadays ground water from most of the regions cannot be directly used for cooking or drinking. The natural water contains essential nutrients but when purified using recent methods, the TDS of water is lost. A method is required for quick identification of potability of ground water using the chemical and physical properties. According to BIS (Indian standards) 34 parameters can be tested for drinking water quality. River water tested with various chemical parameters such as pH, EC, Alkalinity, Cl, Biological oxygen demand, permanganate value etc., and with various physical parameters like temperature, color¹⁻². Fuzzy logic can be applied to identify water quality index of the river water³. Ground water from tube wells with a correlation matrix can be used to detect the physical and chemical interrelation properties of water⁴. Groundwater depletion using satellite and data mining techniques is employed to identify the water quality.⁵⁻⁶

As water quality parameters are more, multivariate with PCA can be applied for parameter reduction and detecting quality of water⁷⁻¹¹. An authentication and quick method to determine the potability of water can help to identify water from regions that can be consumed directly.

EXPERIMENTAL

Study Area

Ground water from various locations of Thanjavur, Ariyalur and Nagapattinam, the Cauvery Delta Districts of Tamilnadu, India were taken for testing. River Cauvery flows through most of the regions of the above districts and major occupation is agriculture.

Sample Collection

Ground water was collected from the various regions in sterilized bottles of one liter capacity. The locations are ulkottai, Kadarankondan, Kangaikondacholapuram, Thaluthalimedu, Meikavalputhur, Ayuthakalam of Ariyalur District, Sirkali, Mayiladuthurai, Nagai, Vetharanyam, Sirkali,

Tharangai, of Nagapattinam District, various samples from Kumbakonam Taluk and Nachiyarkovil village of Thanjavur District were collected.

Parameters Tested

pH was measured with digital pHmeter, Electrical Conductivity through Electrical Conductivity meter for testing. All other parameters such as alkalinity, Total Alkalinity, TDS, Calcium, Magnesium, Sodium, Potassium, Chloride, Sulphate were tested using standard methods.

RESULTS AND DISCUSSION

All 26 samples were tested for the 11 parameters and the results are in Table-1. Indian standards (IS: 10500) was adopted for identifying the quality of water.

Table-1: Ground water analysis for potability from three southern districts of Tamilnadu, India

District	Location	EC	PH	Alkalinity	Total Alka	TDS	ca mg/l	Mg mg/l	Na mg/l	K mg/l	cl mg/l	so4 mg/l
Ariyalur	Ulkottai	610	7.12	0	166	170	40	17	55	3	80	8
	Kadarank	730	7.17	0	260	220	60	17	60	2	72	8
	Kangaikor	550	7.2	0	140	160	48	10	52	3	82	8
	Thaluthali	450	7.13	0	120	134	34	12	38	1	44	15
	Meikaval	610	7.17	0	184	174	42	17	60	2	56	15
	Ayuthakal	470	7.08	0	184	234	60	20	57	3	116	8
Nagai	Sirkali	103	8.8	536.8	60	10.96	49.6	33.07	70	135	454.4	425
	Mailadud	198	8.7	658.8	96	185.6	89.6	84.63	240	89	2364.3	925
	Tharangai	95	8.9	549	60	128.8	68.8	21.4	150	23	724.2	250
	Nagai	128	8.7	402.6	36	98.4	62.4	35.99	150	43	1341.9	1150
	Vetharany	218	8.7	402.6	48	208	160	55.44	210	120	3067.2	1100
	Sirkali	120	8.7	231.8	192	270.4	78.4	48.64	120	5	1384.5	1260
	Mailadud	57	8.6	402.6	120	176	56	35.99	52	2	418.9	0
	Tharangai	374	8.6	341.6	48	105.6	57.6	62.25	550	56	8065.6	1700
	Nagai	214	8.7	573.4	36	140	104	112.84	260	78	2421.1	900
	Vetharany	280	8.7	646.6	96	220.8	124.8	90.47	330	130	4260	2200
Tanjore	Kumbakon	100	8.6	744.2	60	104.8	44.8	35.02	56	360	2243.6	1675
		200	8.4	183	24	65.6	41.6	15.56	38	7	454.4	100
		200	8.4	317.2	36	60	24	10.7	100	1	298.2	0
		200	8.3	451.4	0	78.4	78.4	56.42	90	23	752.6	350
		500	8.4	305	0	24	24	12.65	56	5	447.3	0
	Natchial K	132	8.5	536.8	60	1292	1232	54.48	130	74	1477	625
		52	8.4	329.4	24	520	496	20.45	60	12	539.6	125
		32	8.4	427	24	712	688	27.25	90	4	582.2	125
		35	8	707.6	0	1664	1664	56.42	100	15	1185.7	775
		35	8.4	244	0	464	464	12.65	24	3	255.6	0

A chart given in Fig.1 shows the results of the parameters indicated by lines. As 11 parameters are tested, the 5 important components (EC, pH, Alkalinity, Total Alkalinity, and TDS) are shown separately and the other parameters are shown in the next chart.

The above data were input to PCA using SPSS and the output principal components from the above tool were identified as pH, Total Hardness, Sulphate and chlorides. Using Indian standards, the permissible range for above four parameter ranges were identified. A test set was developed to create a classification rule using JRIP rule miner in WEKA tool. The rules developed for classification using JRIP is given below:

(Cl mg/l less than equal to 999) and (SO₄ mg/l less than equal to 399) and (pH more than equal to 7.1) and (pH less than equal to 8.4); any three= yes (4.0/0.0) (Total Hardness as CaCO₃ mg/l less than equal to

599) and (SO₄ mg/l less than equal to 399) and (pH more than equal to 7.1) and (pH less than equal to 8.4); any three = yes (2.0/0.0); any three= no (18.0/0.0)

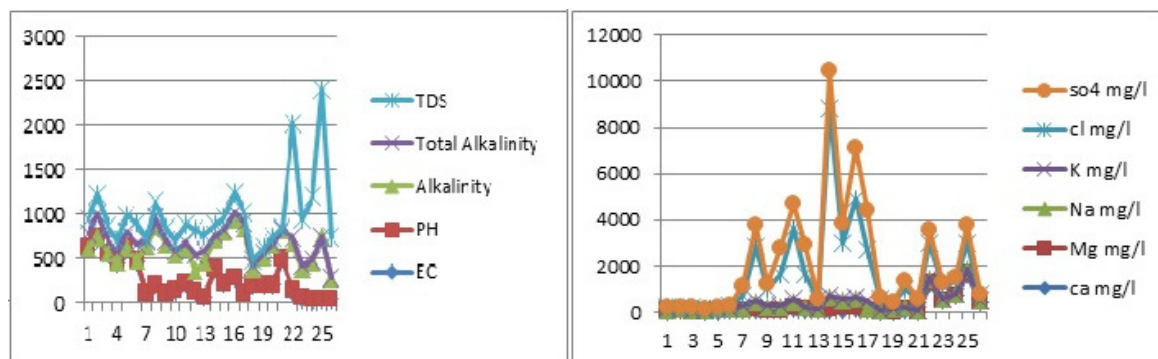


Fig.-1: Water quality parameters of ground water for three southern districts.

Table-2: Classification of Water using JRIP rules.

District	Location	PH	TH mg/l	cl mg/l	so ₄ mg/l	potable
Ariyalur	Ulkottai	7.12	170	80	8	yes
	Kadarank	7.17	220	72	8	yes
	Kamgaiko	7.2	160	82	8	yes
	Thaluthali	7.13	134	44	15	yes
	Meikaval	7.17	174	56	15	yes
	Ayuthakal	7.08	234	116	8	yes
Nagai	Sirkali	8.3	500	354.4	390	yes
	Mailadud	8.7	185.6	2364.3	925	no
	Tharangai	8.9	128.8	724.2	250	no
	Nagai	8.7	98.4	1341.9	1150	no
	Vetharany	8.7	208	3067.2	1100	no
	Sirkali	8.7	270.4	1384.5	1260	no
	Mailadud	8.6	176	418.9	0	no
	Tharangai	8.6	105.6	8065.6	1700	no
	Nagai	8.7	140	2421.1	900	no
	Vetharany	8.7	220.8	4260	2200	no
Tanjore	Mathur	8.6	104.8	2243.6	1675	no
	Village,	8.4	65.6	454.4	100	yes
	Kumbako	8.4	60	298.2	0	yes
	nam	8.3	78.4	752.6	350	yes
	Taluk	8.4	24	447.3	0	yes
	Natchial	8.5	1292	1477	625	no
	Koil	8.4	520	539.6	125	yes
	Village ,	8.4	712	582.2	125	yes
Kumbako	8	1664	1185.7	775	no	
nam	8.4	464	255.6	0	yes	

The above rule is interpreted as follows-

Classification rule 1: If chloride is less than 1000 and sulphate is less than 400 with pH between 7.1 to 8.4, the water is classified as potable.

Classification rule 2: If total hardness is less than 600 and sulphate is less than 400 with pH between 7.1 to 8.4, the water is classified as potable

Classification rule 3: If the above two rules are not satisfied then the water is classified as not potable.

The above rules were used for classification of the original data and the results are shown below in Table-2.

According to JRIP the summary of classification is given below in Table-3.

Table-3: Classification summary for Water Data.

Correctly Classified Instances	24	92.3077 %
Incorrectly Classified Instances	2	7.6923 %
Kappa statistic	0.8301	

The true positive, false positive rate along with precision and recall for the class along with F measure and ROC area that shows the effect of conditions on accuracy of the classes is given in Table-4.

Table-4: Detailed Accuracy By Class

TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
0.889	0.059	0.889	0.889	0.889	0.889	Yes
0.941	0.111	0.941	0.941	0.941	0.889	No
Ave 0.923	0.093	0.923	0.923	0.923	0.889	

CONCLUSION

The results indicate that Ariyalur and Thanjavur districts water regions are generally potable with ph, Total Hardness, chloride and sulphate with in the permissible limit. Regions in Nagapattinum District were affected by Tsunami, hence the chloride values are much higher and pH values are above 8.4. To conclude the ground water from Nagapattinum district cannot be directly used for drinking, so should be appropriately treated before consumption. Data mining techniques can be employed for quicker classification of water potability.

ACKNOWLEDGEMENT

We thank the support and suggestions offered by Professor V.Monika, Department of chemistry, SRC/SASTRA UNIVERSITY, Kumbakonam.

REFERENCES

1. Naziremazlum, Ademozer ,Suleymanmazlum, *Tr.J.of Engineering and Environmental Sciences*, **23**,19,(1999)
2. T. Kowalkowski, R. Zbytniewski, J. Szpejna, B. Buszewski, *Water Research*, **40**(4),744 (2006).
3. D. Juan, H. González, F. Luis, S. Carvajal, M. Francisco, Toro B, *Dyna*, February, 50 (2012).
4. Chadetrik Rout, Arabindha Sharma, *Int.J. of Environmental Sciences*, **2**,2 (2011)
5. Matthew Rodell, Isabella Velicogna , James S. Famiglietti, *Nature* , **20**, 460 (2009).
6. K. W. Chau, N. Muttil, *Journal of Hydroinformatics*, **9**(4),305(2007)
7. D. Karthik, K. Vijayarekha, K. Manjula, IEEE-International Conference on Advances in Engineering, Science and Management, ICAESM-2012, art.no. 6215610, 272, (2012).
8. D. Karthik, K. Vijayarekha, *Research Journal of Applied Sciences, Engineering and Technology*, **4** (24), 5357(2012)
9. D. Karthik, K. Vijayarekha, S. Sekar, *Polish Journal of Environmental Studies*, **23** (2), 573(2014).
10. D. Karthik, K. Vijayarekha, V. Manickkam, *Rasayan Journal of Chemistry*, **7**(1), 99(2014)
11. D. Karthik, K. Vijayarekha, V. Manickkam, *World Applied Sciences Journal* ,**29**, 60,(2014)

[RJC-1144/2014]