International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2018 IJSRCSEIT | Volume 3 | Issue 5 | ISSN : 2456-3307

A Study on Ground Water Quality Prediction using R Tool

S. Vijay
Assistant Professor, Computer Science, Vivekananda College of Arts and Science for Women, Tiruchengode, Tamil Nadu India

ABSTRACT

Water pollution is the presence in groundwater of toxic chemicals and biological agents that exceed is naturally found in the water and may pose a threat to human health and/or the environment. Additionally, water pollution may consist of chemicals introduced into the water bodies as a result of various human activities. Any amount of those chemicals pollutes the water, regardless of the harm they may pose to human health and the environment. The objective of this paper is to assess the overall groundwater quality of the district based on Water Quality Index (WQI), and find out the factors leading to continuous deterioration in groundwater quality and to predict the water quality using different classification and clustering algorithms. The analysis of ground water dataset with various data mining techniques may yield useful outcome to the people to make correct decision for achieving and maintaining necessary level of water quality. This work is to build a quality water for people usage and as well as for drinking purposes using data mining techniques such classification and clustering to find suitable data models with high accuracy.

Keywords : Water Quality, Association rules, Naïve Bayes, ANN and SVM.

I. INTRODUCTION

Water pollution is the contamination of water bodies (like oceans, seas, lakes, rivers, aquifers and groundwater) usually caused due to human activities. Water pollution is any change in the physical, chemical or biological properties of water that will have a detrimental consequence any living organism. Drinking water, also called Potable Water, is the water that is considered safe enough for human and animal consumption. This water is generally used for drinking, cooking, washing, crop irrigation etc. These days' chemicals, bacteria and other pollutants are even affecting our drinking water. Water pollution not only affects water quality but also threats human health, economic development, and social prosperity.

Groundwater is the only alternative option for even the urban centre having well planned, designed and executed water supply systems like Vellore, during the period of water scarcity due to shortfall of rain or its non-occurrence. Nowadays, the groundwater potential and its quality level in the district is getting deteriorated due to population explosion, urbanization, industrialization, and also the failure of monsoon and improper management of rain water. The groundwater quality is normally characterized by different physico – chemical characteristics. These parameters change widely due to the type of pollution, seasonal fluctuations, groundwater extraction, etc., A continuous monitoring of groundwater becomes mandatory to minimize the groundwater pollution, to have control over pollution causing agents and to give corrective measures. Since the groundwater quality in different parts of the district is widely deviating from the prescribed standards, it also becomes necessary

to identify locations and categorize them according to pollution levels.

The specific contaminants leading to pollution in water include a wide spectrum of chemicals, pathogens, and physical changes such as elevated temperature and discoloration. While many of the chemicals and substances that are regulated may be naturally occurring (calcium, sodium, iron, manganese, etc.) the concentration usually determines what is a component of water and what is a contaminant. High concentrations of naturally occurring substances can have negative impacts on aquatic flora and fauna. Agricultural wastewater treatment for farms, and erosion control from construction sites can also help prevent water pollution. Nature-based solutions are another approach to prevent water pollution.

II. LITERATURE REVIEW

Seshadri et.al [1] proposed Association rules to predict ground water quality. The present investigation on groundwater quality with reference to F- concentration in rural area. A high rate of evapotranspiration, long-term contact of waters in the weathered zone by virtue of its low hydraulic conductivity and stagnation of water in the aquifer zone caused by intrusive bodies, intensive and long-term irrigation, and heavy use of fertilizers are the supplementary factors to further increase the F- content in the ground waters. In order to enable sustainable development of groundwater resources, it is necessary to delineate the safe and unsafe zones with reference to F- content.

Ruchi Gupta, Anil Kumar Misra developed a method to detect the quality of ground water isthe availability of salt rich geological formation in subsurface in Jhajjar district. The groundwater quality is totally unsuitable for domestic purposes as Water Quality Index is more than 100 for the region. Long term intake of fluoride above the permissible limit in drinking water is causing dental fluorosis diseases in the study areas.

Gorai, Hasni, Jased Iqbal proposed Fuzzy rule-based approach to predict ground water quality index to assess suitability for drinking purpose. The study suggests a robust decision-making tool for drinking water quality management in the form of the fuzzy water quality index (FWQI). The developed methodology demonstrates to determine a single index value to make assessment of drinking water quality more understandable especially in public consideration. The fuzzy model developed is applicable only for specific number of water quality parameters in specified range selected.

Mallika et.al proposed linear regression to predict ground water quality model for irrigation using data mining techniques. An effective data mining technique was used to predict water quality and thereby forecast the crop yield. It is decisively concluded that this may help the decision maker to predict the crop yield with respect to water quality before harvesting the crop.

III. DATA MINING TECHNIQUES

Data Mining is the process of turning raw data into appropriate and meaning information. Various researchers have studied and work on data mining techniques to evaluate and classify the water quality.

A. ANN (Artificial Neural Network)

ANN is a classification model, which is grouped by interconnected nodes. An ANN is based on a collection of connected units or nodes called artificial neurons which loosely model the neurons in a biological brain. Each connection, like the synapses in a biological brain, can transmit a signal from one artificial neuron to another. An artificial neuron that receives a signal can process it and then signal additional artificial neurons connected to it. The original goal of the ANN

approach was to solve problems in the same way that a human brain would. However, over time, attention moved to performing specific tasks, leading to deviations from biology. ANNs have been used on a variety of tasks, including computervision, speech recognition, machine translation, social network filtering, playing board and video games and medical diagnosis.

B. Back Propogation Neural Network (BPNN)

ANN consists of interconnected processing units. Each unit is known as neuron. Each neuron will receive an input from another neuron. Weights are assigned to each neuron. These kinds of weights regulate the nature as well as strength and power of the significance involving the interlocked neurons. The respective signals named as indicators tend to be refined from each and every input and then further processed via a weighted sum to the inputs. The BPNN algorithm criteria looks for the error with the method called as steepest descent. The united weights are modified by simply moving on the way to the negative gradient of the energy function by providing emphasis at each and every iteration for evaluating the network performance.

C. Naïve Bayes

Naïve Bayes is a classification technique, which is based on probability theories, which entirely demonstrate the characteristics of water quality assessment. Bayes model is easy to use for very large datasets. In other terms, a Naive Bayes assumed that the value of a distinct feature does not related to the presence or absence of any other feature, given in the class variable.

D. Support Vector Machine

SVMs are supervised data mining algorithm which can be used for either classification or regression challenges. However, it is frequently used in classification problems. SVMs have become the method of choice to solve difficult classification problems in a wide range of application domains. SVMs built on the fundamental of minimization of structural risk. In linear classification, SVMs produce a hyper plane that splits the training data into two groups with maximum-margin. A maximum-margin hyper plane is a hyper plane which separates two of points and is at equal distance from the two. Mathematically, SVMs learn the sign function f (x) = sng (wx + b), where w is a weighted vector in Rn. SVMs find the hyper plane y = wx + b by separating the space Rn into two half-spaces with the maximum-margin. Linear SVMs can be generalized for nonlinear problems. To do so, the data is mapped into another space H and we perform the linear SVMs algorithm over this new space.

E. Decision Tree

Decision tree is one of the predictive modelling technique used in data mining. It aids to divide the larger dataset into smaller dataset indicating a parent-child relationship. Each internal node defined as inner node is labelled with an input feature. The inner nodes which exhibit many types of attribute test, bifurcations exhibit the test outcomes and leaf nodes particularly exhibit the category of a specific type[4]. Decision tree can handle both numerical and categorical data. It is well suited with large datasets. Higher accuracy in decision tree classification technique depicts that the technique can simulate. It is able to optimize variety of input data such as nominal, numeric and textual. It is a successful supervised learning approach which has the capability of extracting the information from vast amount of data based on decision rules.

IV. WATER QUALITY INDEX

WQI is computed to reduce the large amount of water quality data to a single numerical value that expresses the overall water quality at a certain location and time based on several water quality parameters. It is also defined as a rating reflecting the composite influence of different water quality parameters on the overall quality of water.

According to the concept of indices to represent gradation in water quality was first proposed by Horton (1965). The main objective of water quality index is to turn complex water quality data into information that is understandable and useable by the public. Water Quality Index based on some very important parameters can provide a simple indicator of water quality.

V. MATERIAL AND METHODS

Study Area

Vellore is the second most populous district in TamilNadu and had population of 906,745 as per 2011 census. In terms of urbanization level, Vellore district ranks 8thplace among the other districts in Tamil Nadu. Vellore is a major transit point for travellers, a hub for medical tourism and is emerging as a tourism hot spot. This place is known for its extreme climatic conditions. Vellore has an arid and dry climate, reaching high temperatures during summer. The city experiences wet winters and dry summers and has an elevation of about 224 meters with the north-east monsoon the highest contributor to rainfall. The mean maximum and minimum temperatures during summer and winter varies between 38.3°C and 18.95°C. The District lies between 12º15'23''to 13º12'32'' N Latitude, 78º24'16'' to 79º54'56''E Longitude and has an aerial extent of 6077 sq.km.

In the present research work, the study area in Vellore district have large industrials profiles such as textiles, leather tanneries and small-scale dying industries. The effluents of the leather industries, usage of the chemical fertilizers for agriculture and small-scale dying industries fall heavily on the quality of the drinking water. The impact is felt very much on the drinking water sources, which are available for the people.

5.1 Water Quality Index Calculation

In this study, we used 5 groundwater quality variables [TDS, Hardness, Magnesium, Sulfate, Nitrate] the quality of the water in Deep Tube well located in and around the Vellore district was studied by determining the physico-chemical parameters of the water quality. These studies generate a baseline regarding the quality of the water, which is used for drinking and other household purposes. To calculate the water quality index, the method proposed by Horton (1965), and followed by many researchers was used. The following formula is used to calculate the water quality index (WQI):

$$WQI = \sum q_i w_i / \sum w_i$$

In which qi is the rating scale and wi is the related unit weight for the respective water quality parameter.

DATA PROCESSING

In order to predict the ground water quality, the data related to physical and chemical parameters of ground
water were collected from TWAD (Tamilnadu Water Supply and Drainage Department) board from the year 2015 to 2016 for all 11 locations and data model is developed for each parameter.

VI. RESULTS AND DISCUSSION

The various physic-chemical parameters of groundwater samples were analyzed and the descriptive statistics of the analyzed parameters are given in Table 1. The results are compared to the World Health Organization recommended maximum permissible limits and BIS standards.

Table 1. The analytical results showing quality of ground water in the study area

Sample Station	Type	TDS [mg/l]	Hardness[mg/l]	Magnesium [mg/l]	Nitrate [mg/l]	Sulphate[mg/l]
ANAICUT	Deep Tubewell	3150	880	154	53	420
ARAKKONAM	Deep Tubewell	2349	880	103	56	461
ARCOT	Deep Tubewell	2067	720	106	49	604
GUDIYATHAM	Deep Tubewell	2520	950	118	49	410
KANIYAMBADI	Deep Tubewell	3080	960	106	52	405
KAVERIPAKKAM	Deep Tubewell	2426	695	113	50	512
PERANAMBATTU	Deep Tubewell	2961	840	120	53	436
THIMIRI	Deep Tubewell	2188	720	110	48	408
TIRUPATHUR	Deep Tubewell	2092	900	108	54	443
VELLORE	Deep Tubewell	2443	840	101	58	517
WALAJAPET	Deep Tubewell	2468	630	106	55	478

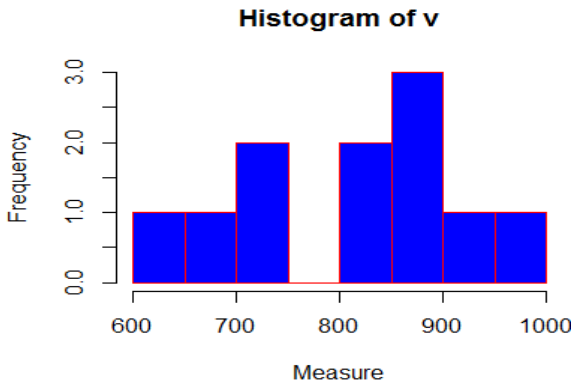
WQI of ground waters collected at 11 different locations at Vellore district were calculated using the proposed methods Based on the attribute database of the water quality, the mapping of physicochemical analysis has been carried out and the extent of estimated parameters of pollution. The groundwater quality is totally unsuitable for domestic purposes as per Water Quality Index.

VII. RESULTS AND DISCUSSION

7.1 Findings and Interpretations

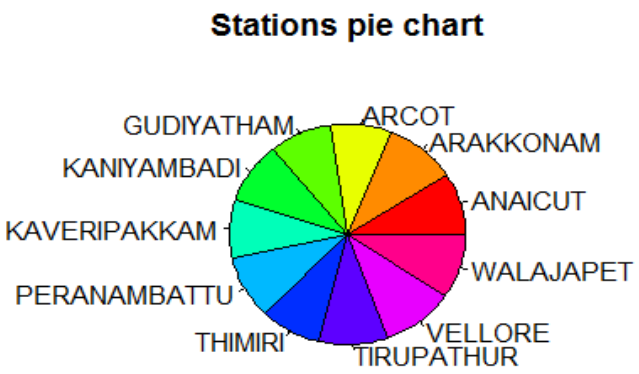
Histogram:

```
v<-
c(3150,2349,2067,2520,3080,2426,2961,2188,2092,2443,2468)
>hist(v,xlab = "Measure",col = "blue",border = "red")
```



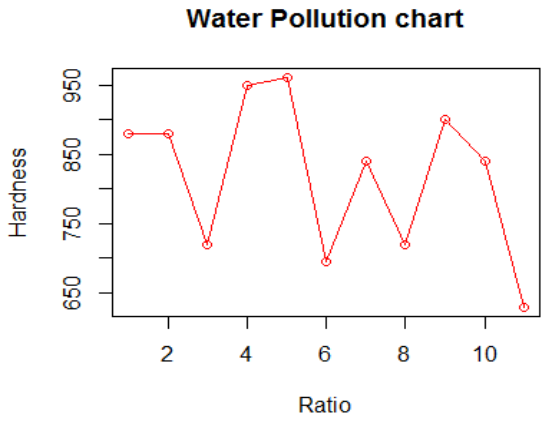
Pie Chart:

```
x <- c(53,56,49,49,52,50,53,48,54,58,55)
labels <-
c("ANAICUT","ARAKKONAM","ARCOT","GUDIYA
THAM","KANIYAMBADI","KAVERIPAKKAM",
"PERANAMBATTU","THIMIRI","TIRUPATHUR","
VELLORE","WALAJAPET")
```



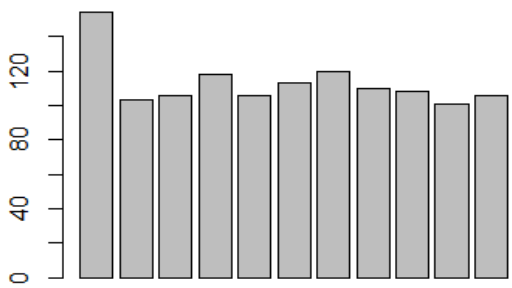
Line Chart:

```
v <- c(880,880,720,950,960,695,840,720,900,840,630)
plot(v,type = "o")
plot(v,type = "o", col = "red", xlab = "Ratio", ylab =
"Hardness", main = "Water Pollution chart")
```



Bar Chart:

```
H <- c(154,103,106,118,106,113,120,110,108,101,106)
barplot (H)
```



VII.CONCLUSION

This paper presents an evaluation for predicting water quality by applying numerous data mining techniques and methods at many different locations. Many existing evaluation methods are studied. Various algorithms have been reviewed for predicting the water quality and hence made a comparison. There are numerous water borne diseases like cholera, diarrhea, dysentery etc., which are transmitted by drinking contaminated water. Research is being conducted all over the world to develop more and more techniques which can generate pure at low cost.

VIII.REFERENCES

[1] Das Gupta, M.; Purohit, K.M.; Jayita Datta.Assessment off drinking water quality of river Brahmani. Journal of f Environment and Pollution. 2001, 8, 285-291.

[2] Subba Rao, N., Srinivas Rao, G., Venkateswara Rao, S., Madhusudhana Reddy, P. & John Devadas, D. (1999) Environmental control of groundwater quality in a tribal region of Andhra Pradesh, India. *Indian J. Geol.* 71, 299–304.

[3] Subba Rao, N. & John Devadas, D. (2003)Fluoride incidence in groundwaters in a part of Peninsular India. *Environ. Geol.* (in press).

[4] Niketa Gandhi, lesia Armstrong "Applying Data Mining Techniques to Predict Yield of Rice in Humid Subtropical climatic zone of India" IEEE Paper 2016.

[5] Shoba G, Dr. Shobha G." Water Quality Prediction Using Data Mining techniques: A Survey" IJECS 2014.

[6] Samson S. and Elangovan K "Hydro chemical analysis and estimation of ground water quality in namakkal District ", tamilnadu,,International journal of

Ecotoxicology and environmental pollution 2011.

[7] A. Ambica, Ground water quality haracteristics study by using water quality index in Tambaram area, Chennai, Tamil Nadu,Middle-East J. Sci. Res. 20 (11) (2014) 1396–1401.

[8] K.M. Anwar, V. Aggarwal, Analysis of groundwater quality of Aligarh city, (India): using water quality index, Curr. World Environ. 9 (3) (2014) 851–857.

[9] I.N. Balan, M. Shivakumar, P.D.M. Kumar, An assessment of ground water quality using water quality index in Chennai, Tamil Nadu, India, Chronicles Young Sci. 3 (2) (2012) 146–150.

[10] Abbasi SA (2002) Water quality indices, state of the art report, National Institute of Hydrology, Scientific contribution no. INCOH/SAR-25/2002, INCOH, Roorkee p 73.