International Journal of Research in Advent Technology, Vol.7, No.6S, June 2019
E-ISSN: 2321-9637
Available online at www.ijrat.org

Clustering of Water Reservoirs based on Water Chemical Analysis

Umesh Gupta, Sonal Jain, Dharmendra Kumawat, Roopali Sharma

Abstract— As per Rajasthan Water Assessment report of International Finance Corporation, World Bank Group, Rajasthan is among regions with greatest climate sensitivity and lowest adaptive capability. The major groundwater measurable parameters are Fluoride, Nitrate, Total Hardness, Total Alkalinity, pH and Total Dissolved Solids (TDS). Ground water quality invariably changes with distance in any region. In most of the urban areas, water supply is dependent on surface water and ground water reservoirs. Data mining is being used significantly for water quality modelling. Some of such techniques such as clustering and classification, support vector regression, artificial neural network are being applied in various such applications. In this paper, a case study is undertaken to form clusters of 152 water reservoirs in Jaipur city based on their chemical analysis reports. K-means clustering algorithm is performed for clustering of the water reservoirs based on similar chemical and water quality parameters. Correlation analysis has also been performed on these parameters in each of these obtained clusters. It would help to device area specific strategy for ensuring water quality which is one of the major Sustainable Development Goal (SDG) defined United Nations.

Keywords- Chemical Analysis; Ground Water Quality; Data Mining; K-means clustering.

I. INTRODUCTION

Water, although one of the most inevitable natural resource, it's quality measurements are either ignored are not systematically measured. There are many threats that affects quality of water when it comes to usage of drinking purpose.

Jaipur is largely dependent on conjunctive water sources. A part of Jaipur city is being supplied water from Bisalpur Dam which is affecting water quality in the area. Water quality has been the prime focus for Government. With the over exploitation of ground water resources, surface water bodies are getting low [1]. Also the adverse climatic conditions are creating several imbalances between the recharge and draft. Kala, et. al. [2] observes poor sanitation condition for water quality beyond permissible limit as per IS: 10500 standards. It is always challenging to ensure water quality and therefore, Government is drafting effective policies (State Water Policy, 2010) and working on various projects.

Several researchers used Multivariate statistical techniques for water quality assessment in different regions. Hada, et. al. [3] used goal programing with multiple nonlinear regression to assess water quality in both pre-mansoon and post-mansoon season in Tonk district of Rajasthan and found this methodology to provide more accurate results. Shrestha and Kazama [4] used cluster analysis, principal component analysis and discriminant analysis for assessment of surface water quality of the Fuji river basin in Japan. Singh, et. al. [5] also applied these techniques in Gomtiriver in India and found them useful in evaluation of the complex water quality dataset. Iscen, et. al. [6] identified microbiological, organic-nutrient and physiochemical factors explaining variation in dataset of surface water quality in Uluabat lake in Turkey. Similar study was done by Gupta, et. al. [7] on Mumbai cost to identify pollution sources and the authors found organic pollution, natural pollution and nutrient pollution and seasonal effects of temperature as major pollution sources.

Cluster analysis is found by many researchers as a useful technique to divide the water resources based on similarity in water quality dataset. Emad, et. al. [8] used Hierarchical cluster analysis on the water quality dataset of Euphrates River and established the usefulness of this technique. This technique was also used by Grande, et. al. [9] and Fataei & Shiralipoor [10].

All these studies have signified the importance of applying statistical techniques for evaluation of water quality of different regions around the globe. The present paper is aimed to apply K-means clustering analysis to classify the water resources in Jaipur city based on their similar water quality parameters.

II. METHODOLOGY

A. Dataset

Jaipur city has been chosen due to its varied water conditions in which some area are almost dry, some areas are under AmanishahNala bed and some are affected by

Manuscript revised on May 9, 2019 and published on June 5, 2019
Umesh Gupta, *JK Lakshmipat University, Jaipur. Email ID: umeshgupta@jklu.edu.in*
Sonal Jain, *JK Lakshmipat University, Jaipur. Email ID: sonaljain@jklu.edu.in*
Dharmendra Kumawat, *BIT, Mesra, Jaipur. Email ID: kumawatdhamendra@gmail.com*
Roopali Sharma, *BIT, Mesra, Jaipur. Email ID: roopalisharma@bitmesra.ac.in*

doi: 10.32622/ijrat.76S201954

Industry. Though, major areas are supplied water through Bisalpur dam, yet there are many areas uncovered or the water quality is very poor due to various reasons. The water quality parameters ph, Alkalinity, TDS, Total hardness (calcium hardness and magnesium hardness), Chloride, Nitrate and Fluoride were measured in the labs of Public Health Engineering Department (PHED), Government of Rajasthan.

PHED has been assessing water quality and chemical parameters in various parts of city on regular basis and it releases its data. This data has been referred for this study. This study is covering 54 different areas for the sampling. Total 152 samples were collected from Central Water Reservoirs (CWR) and Surface Water Reservoirs (SWR) of these areas. The sampling process was carried out during September-November, 2014. Measuring and analysis was done upon 8 physical, chemical, microbiological parameters. The descriptive statistics of data is provided in Table 1.

Table 1: Descriptive Statistics of dataset

	pH	Total Alkalinity	Total Dissolved Solids	Total Hardness	Chloride	Nitrate	Fluoride
count	152	152	152	152	152	152	152
mean	7.87	115.59	346	97.30	45.26	9.22	0.28618
std	0.11	79.89	216.75	41.56	30.54	10.26	0.123466
min	7.3	60	176	40	20	2	0.2
25%	7.8	80	248	77.5	30	2	0.3
50%	7.9	100	280	90	40	5	0.3
75%	7.9	110	344	110	40	11	0.3075
max	8.3	720	1440	270	220	45	1.3

B. Data Scaling and Normalization

The dataset provided by the laboratory has different scales. For data analysis further, it is required to standardize or normalize the dataset before processing it. In the present study, Min-Max scaling is used to scale the data to fix the data range between 0 and 1. This scaling is done by following equation.

$$X_{ms} = \frac{X - X_{min}}{X_{max} - X_{min}}$$

1) This process is considered as an alternative to z-score normalization.

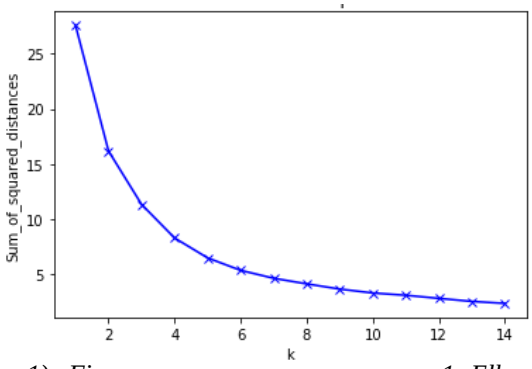
C. Clustering

Clustering is a fundamental technique of unsupervised learning in statistics and machine learning [11]. In this paper, K-means clustering has been used for making clusters of water reservoirs based on similar water quality characteristics. This clustering approach has been used for water quality analysis of the Haihe River in Chinaby Zou, et. al. [12].

$$Minimize\ J = \sum_{j=1}^{k}\sum_{i=1}^{n}\left\|x_i - c_j\right\|^2$$

Where, the Euclidian distance $\left\|x_i - c_j\right\|^2$ is the distance of n data points from their respective cluster centers and objective to find out cluster centers c_j for minimum distance.

Elbow criterion method has been applied to determine optimal number of K. The plot of sum of squared distances for each K is presented in Figure 1 that suggests the optimal value of K = 3.



1) Figure 1: Elbow Method - SSD plot for K

Sum of squared distance =

$$\sum_{j=1}^{k}\sum_{x_i \in Cluster\,j}\left\|x_i - \bar{x}_j\right\|^2$$

Where, $\bar{x}_j$ is the sample mean in cluster j.

D. Pearson Correlation

Correlation between all water quality parameters is obtained through heat map for each cluster. The advantage of heat map is to distinguish strongly correlated and poorly correlated parameters clearly.

doi: 10.32622/ijrat.76S201954

III. RESULTS AND DISCUSSION

A. Quality Assessment

Obtained average values of various physio-chemical parameters along with their permissible limits have been discussed in Table 2.

Table 2: Quality Assessment of Physio-chemical parameters

Physio-chemical Parameter	Permissible Limit	Discussion as per Table 1
pH	6.5 – 8.5	Within permissible limit
Total Hardness	200	Average is within limit but the quartile and quartile range is showing 25% of the water reservoirs sensitive
Total Dissolved Solids (TDS)	500	
Total Alkalinity	200	
Chloride	250	Within permissible limit
Fluoride	1.0	Quartile range is showing that within 25% of the water reservoirs are sensitive
Nitrate	45	Within permissible limit

B. K Means Cluster Analysis:

The result of K means cluster analysis divides all the 152 reservoirs of 54 areas into 3 different clusters named as Cluster 0, 1 and 2. This is based on the assumption that the clusters should be homogeneous within the clusters and heterogeneous from the other clusters. Heat map representing correlation between various physio-chemical parameters are provided in Figure 2.

Cluster 0:

This cluster contains water reservoirs from only 7 areas of Jaipur which are Sec. 26 Pratap Nagar, Shastri Nagar Housing Board, Bhatta Basti and J.K. Loan Hospital. Heat map for this cluster representing correlation between various physio-chemical parameters. Significant correlation is found between pH, Total Alkalinity and Total Dissolved Solids and moderately strong correlation is found between Total Hardness, Chloride and Nitrate.

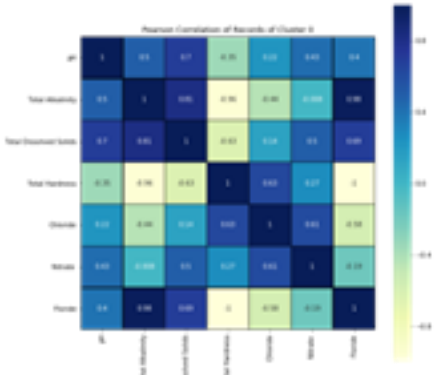


Figure 2: Heat Maps for Cluster 0, 1 and 2

Cluster 2:

This cluster contains water reservoirs from 22 areas of Jaipur which are Sec. 10Pratap Nagar, LaxmanDoongri, Jyoti Nagar, ChitrakootYojna, Vaishali Nagar, Khatipura, Hanuman Nagar, Shanti Nagar, Hasanpura, Durgapura and Taro Ki Koot. Heat map for this cluster representing correlation between various physio-chemical parameters. There is a high correlation found between Nitrate and Total Hardness. This heat map is also highlighting the parameters having weak correlation as well.

Cluster 1:

This cluster contains water reservoirs from all the remaining 123 areas of Jaipur Heat map for this cluster representing correlation between various physio-chemical parameters. There is a high correlation found between Nitrate and Total Hardness. This heat map is also highlighting the parameters having weak correlation as well.

It is clearly observed that these clusters have been classified irrespective of the geographical boundaries. However, having 123 areas in the same cluster is may be due to piped water distribution through Bisalpur project. It is not necessary that one particular area will have similarity in terms of physio-chemical parameters rather small areas in different corners may have such similarity and those can lie in the same cluster.

The moderate to high correlation between TDS and

doi: 10.32622/ijrat.76S201954

Nitrate shows that the ground water quality is being deteriorated. Presently the surface water quantities, as compared to underground is being used more, dilutes the impact of high TDS. In a case the ground water use increases, the impact of TDS would call alarming state of affairs. High nitrate indicates poor sanitation arrangements by the utilities. The high nitrates may results in methanoglobimania among infants, which is a high alarming symptom.

Presently water treatment for high TDS and multi water quality problems, RO Plants have been recommended, but due to high wastages of water this solution should be adopted with limited use. The clustering as per the variation in water quality suggests that the cluster centric approach for water treatment may be more cost effective. The study proposes techno-economic solution which may reduce the burden on resources, and also reduces waste water.

The quality of water is not geographic area specific, but as the study suggests, they are depending on the reservoirs. Presently, chemical water treatment is not being adopted; instead such poor sources are being adopted. But high technology use can be preferred for specific clusters.

IV. CONCLUSION

There is a wide diversity in water quality in different regions in Jaipur. State Government of Rajasthan arranges drinking water supply through Bisalpur Dam and local tube-wells in various regions of Jaipur city. However, there is a large area dependent upon water mix of surface water and ground water through clear water reservoirs scattered at various locations in the district. Therefore, water quality varies with distance within Jaipur city. The aim of this paper was to divide the Jaipur region in the clusters that have equally likely water conditions. The authors used Python programming on the data set of 152 water reservoirs that was collected in the year 2014. This tool provides accurate clustering of a given data. K-means clustering algorithms on the chemical analysis report of various parts in Jaipur have been applied successfully and 8 clusters have been formed based on similar water quality and chemical parameters.

The clustering pattern shows that the areas under Pratap Nagar and nearby areas, Housing Board Colony, Bhatta Basti, Tilak Nagar and JK Loan Hospital are varying significantly as compared to other areas. One cluster of 7 specific areas is presently showing that they are sensitive and specific water treatment is required to be focused in these 7 selected areas.

The sustainability lies in optimal use of resources whether it is in terms of technology or cost associated with it. Based on cluster analysis, it is recommended to do further specific water treatment on the regions having equally likely water quality. K-means clustering technique is suitable to define these clusters. This innovative approach will help in policy formulation in water quality sector in particular and towards public health in general.

ACKNOWLEDGMENT

The authors would like to acknowledge Public Health Engineering Department for providing data for this study.

REFERENCES

[1] R. K. Tatawat and C.P. Singh Chandel, “Quality of Ground Water of Jaipur City, Rajasthan (India) and Its Sustainability for Domestic and Irrigation Purpose”, App. Eco and Env. Res., 6 (2): 79-88 2008.

[2] Jayshri Kala, Gunwant Sharma, Sudhir Kumar and Satish Pipralia, “Assessment of Ground Water Quality in Jaipur City (Rajasthan, India). International Journal of Enhanced Research in Science Technology & Engineering” 3(7): 89-101, 2014.

[3] Devendra Singh Hada, Umesh Gupta, and S. C. Sharma, “Seasonal Evaluation of Hydro-geochemical parameters using Goal Programming with Multiple Nonlinear Regression”, Gen. Math. Notes, 25 (2): 137-147, 2014.

[4] S. Shrestha and F. Kazama, “Assessment of surface water quality using multivariate statistical techniques: A case study of the Fuji river basin, Japan”, Environmental Modelling & Software 22: 464–475, 2007.

[5] K. Sing, A. Malik, D. Mohan and S. Sinha, “Multivariate Statistical Techniques for the Evaluation of Spatial and temporal Variations in Water Quality of Gomti River (India)—A Case Study. Water Research” 38 (18): 3980-3992, 2004

[6] C. Iscen, Ö Emiroglu, S. Iihan, N. Arslan, V. Yilmaz and S. Ahiska, “Application of Multivariate Statistical Techniques in the Assessment of Surface Water Quality in Uluabat Lake, Turkey”, Environmental Monitoring and Assessment, 144 (1-3): 269-276, 2008.

[7] I. Gupta, S. Dhage and R. Kumar, “Study of Variations in Water Quality of Mumbai Coast through Multivariate Analysis Techniques”, Indian Journal of Marine Sciences, 38(2): 170-177, 2009.

[8] A. M. Salah Emad, M. Turki Ahmed M. Al-Othman Eethar, “Assessment of Water Quality of Euphrates River Using Cluster Analysis”, Journal of Environmental Protection, 3(12): 1629-1633, 2012.

[9] J. Grande, J. Borrego, J. Morales and M. De La Torre, “A Description of Show Metal Pollution Occurs in the Tinto— Odiel Rias (Huelva-Spain) through the Application of Cluster Analysis. Marine Pollution Bulletin”, 46(4): 475-480, 2003.

[10] E. Fataei, and S. Shiralipoor, “Evaluation of Surface Water Quality Using Cluster Analysis: A Case Study”, World Journal of Fish and Marine Sciences, 3: 366-370, 2011.

[11] J. A. Hartigan, “*Clustering Algorithms*”, Wiley: New York, NY, USA, 1975.

[12] Hui Zou, Zhihong Zou and Xiaojing Wang, “An Enhanced K-Means Algorithm for Water Quality Analysis of The Haihe River in China”, International Journal of Environmental Research and Public Health, 12: 2015.

doi: 10.32622/ijrat.76S201954