Agriculture, Environment & Society (2022) Volume 2, Issue 1, P. 77-84

Agriculture, Environment & Society

Journal homepage: www.aes.uoz.ac.ir

Providing a model for predicting and status of aquifer depth changes through tree and clustering algorithms

Seyed Hassan Mirhashemi [*a], Mehdi Panahi [b]

[a] M.Sc Graduate of Water Engineering, Faculty of agriculture, University of Zanjan, Zanjan, Iran
[b] Department of Water Engineering, Faculty of Agriculture, University of Zanjan, Zanjan, Iran

ARTICLE INFO

Article history:
Received: 25 March 2022
Accepted: 27 April 2022
Available online: 5 May 2022

Keywords:
Accuracy
Human and natural factors
Irrigation network
Temperature

ABSTRACT

The need for a model for effective planning and management of water resources, particularly groundwater, is especially critical in light of water scarcity and aquifers. Given the importance of various factors in determining the amount of drop, this study used human and natural factors to predict the amount of aquifer drop in Qazvin. To accomplish this, the K-Means clustering algorithm was used first, followed by the tree algorithms CART, CHAID, C5.0, and QUEST to determine the optimal ratio between different fields. Accuracy values of 0.90, 0.96, 0.94, and 0.92 were obtained for the aforementioned tree algorithms. The values obtained for the CHAID algorithm's sensitivity, transparency, accuracy, precision, false-positive rate, false-negative rate, F-measure, geometric mean, and error rate demonstrate that this algorithm outperforms other algorithms. The amount of water in the irrigation network is the most influential human factor in model production, while the amount of temperature is the most influential natural factor. The proposed model enables more accurate prediction of aquifer changes and can be used by managers and farmers to improve aquifer management.

Highlights

- Given water scarcity and aquifers, a model for effective water planning and management is critical.
- Given the importance of various factors in determining aquifer drop, this study used both human and natural factors in Qazvin.
- CHAID algorithm outperforms other algorithms in terms of sensitivity, transparency, accuracy and precision, geometric mean and error rate.
- The irrigation network's water supply is the most important human factor, while temperature is the most important natural factor.

1. Introduction

The development and progress of agriculture, as well as population growth, which led to an increasing need for water resources, created instabilities in traditional water resource management. A major part of the imbalance in water resources is due to the natural limitations of water resources. Managers in this sector are also faced with complex relationships and very diverse characteristics of the vast amount of data collected, which are difficult to analyze and manage by experimental and statistical methods, and in many basins, practically impossible.

In recent years, researchers have conducted various studies related to spatial variation and groundwater level estimation (Jang et al., 2013). The Standardized Precipitation Index (SPI) was used to investigate the effects of drought and rainfall on groundwater levels in three irrigated areas in the Marie-Darlin Basin, Australia. Their results showed a good correlation between the SPI index and groundwater level fluctuations in the region, and it can be used to determine the pattern of major droughts in Australia (Khan et al., 2008).

Prediction results showed that the CART data analysis tree model can provide a correlation between variables and can increase the accuracy of prediction by reducing additional information. By comparing the CART model with the PSO-SVR model, the CART model with better fit and better forecasting ability can be used to predict groundwater level drops (Zhao et al., 2016). The results showed that groundwater managers and decision-makers could support the implementation of programs to protect

[*] Corresponding author
Email: hassan.mirhashemi@yahoo.com (S.H. Mirhashemi)
http://dx.doi.org/10.22034/aes.2022.335733.1028

groundwater resources by using general and detailed information from the data analysis decision tree (Stumpp et al., 2016). Using data analysis and data of old aquifers in the Toluca valley, it was determined that data analysis of these aquifers is able to produce new knowledge. Using the data analysis algorithm, it is determined that water management underground is affected by social and economic factors such as GDP and population structure (Corona et al., 2016). Data analysis classification methods were used to classify water quality improvement factors. Cluster analysis tends to be categorized based on groundwater quality and pollution characteristics (Oorkavalan et al., 2016).

It was identified areas of high potential groundwater using a CART1 data analysis algorithm method and RF and BRT methods, and in mapping the springs of the Koohrang basin using BRT, CART, and RF models, the accuracy of the models was 0.8103, 0.7870, and 0.7119, respectively. Therefore, the BRT model had the best performance in mapping groundwater resources, followed by the CART and RF models in the second and third ranks. According to the results, the accuracy of all three models is more than 70%. Therefore, all three models can be used by planners and engineers in water and land resource management and planning in the study area (Naqbi et al., 2016). Using data from the Meteorological Station of the Laboratory of Energy and Environmental Physics, the Department of Physics at Patrice University in Greece, data analysis techniques were used to estimate maximum, minimum, and average temperatures. It was concluded that the data analysis regression algorithm makes it possible to predict the maximum, minimum, and average temperatures with satisfactory accuracy. Also, a hybrid data analysis technique was developed for estimating daily values of mean temperature and achieved the same result (Kotsiantis et al., 2008). In India, groundwater resources are declining (Bonsour et al., 2017). Improper use of alluvial aquifers is one of the main causes of subsidence (Novinpour, 2017). Natural factors affect water access (Konapala et al., 2020). Mirhashemi et al., Used data mining methods to predict aquifer depth changes (Mirhashemi et al., 2020).

In the field of water resources management, we are faced with a huge amount of spatial and temporal data such that it is practically impossible to use experimental and statistical methods for converting such data into applied knowledge. Data analysis is a powerful technique for managing and organizing information as well as extracting useful knowledge from a large amount of data. In this paper, it was assumed that this method could be used for better aquifer management. Therefore, due to the need for a strong and appropriate algorithm in this field and the capabilities of data analysis algorithms regarding aquifer management, it is necessary to use this method. Data analysis is the process of recognizing valid, new, inherently useful, and understandable patterns of data, as well as automatically searching large data sources for patterns and dependencies that simple, routine statistical analysis cannot perform.

2. Materials and methods

2.1. Study area

The Qazvin plain, with an area of about 450,000 hectares, is located in the range of longitudes of 49 degrees and 25 minutes to 50 degrees and 35 minutes east and latitudes of 35 degrees and 25 minutes to 36 degrees and 25 minutes north. This plain is composed of a wide alluvial plain composed of sediments from surface currents of the surrounding mountains (Mohammadi et al., 2011). The total aquifer nutrition of the Qazvin plain is 1259.46 million cubic meters. The total discharge factor of the Qazvin plain aquifer is 1458.66 million cubic meters. Accordingly, the share of discharge in the agricultural sector is about 1352.92 million cubic meters, of which about 857.3 million hectares is the share of the agricultural sector. Due to the limited surface water resources and the seasonality of these resources, most of the irrigation water is extracted from groundwater sources. In the current situation, the harvest has caused an annual drop of 1.5 meters in the surface of the aquifers and up to about 25 centimeters per year of subsidence in this area. Considering the importance of the Qazvin plain as a potential agricultural area, on the one hand, and the problem of severe water drop in this area, on the other hand, it seems necessary to pay attention to the sustainability of groundwater resources in the production of agricultural products and the choice of cultivation pattern in this area. (Barikani et al., 2011). The Qazvin province has six counties: Abik, Avaj, Alborz, Buinzahra, Takestan, and Qazvin. Of these six counties, parts of Abik, Alborz, Buinzahra, Takestan, and Qazvin are in the Qazvin plain. Due to the different behavior of the Qazvin plain aquifer in different parts of the plain, in this study, only part of the Qazvin plain aquifer, which is within the agricultural area of Qazvin County, was studied (Figure 1).

Figure 2 shows the location of the irrigation network in the agricultural area of Qazvin County. According to Figure 2, the area outside the irrigation network is about 757.7587 hectares, and the area within the irrigation network is about 19908.908 hectares.

Out of 174 authorized wells used in the agricultural area of Qazvin city, 138 wells are of the agricultural exploitation type, 23 wells are of the integrated exploitation type, and 13 wells are of the multi-purpose exploitation type. Of the 174 wells available, 134 are within the irrigation network. The average depth of wells is 123 meters, the maximum depth of wells is 200 meters, and the average discharge of wells is 36 liters per second, and the maximum discharge of wells is 90 liters per second (Figure 3).

2.2. Models and data used

The method of work in this study was predictive data analysis. Clustering is the non-regulatory process of grouping similar elements into clusters. Classification can be performed based on clustering if category or class information is used to evaluate the obtained clusters. This approach is based on the "cluster to batch" evaluation procedure and finds a mapping with minimal error from clusters to classes (Lopez et al., 2012).

In this research, first, the K-Means clustering algorithm was used to obtain the best ratio between different variables. Then the decision tree algorithms were used to obtain the best ratio between different clusters resulting from the implementation of the K-Means algorithm.

To train a decision tree, a class of variables must have an output field and one or more input variables. Input fields, human and natural variables affecting aquifer depth changes were selected, and the result of clustering with the K-Means algorithm was considered as the output variable and prediction target.

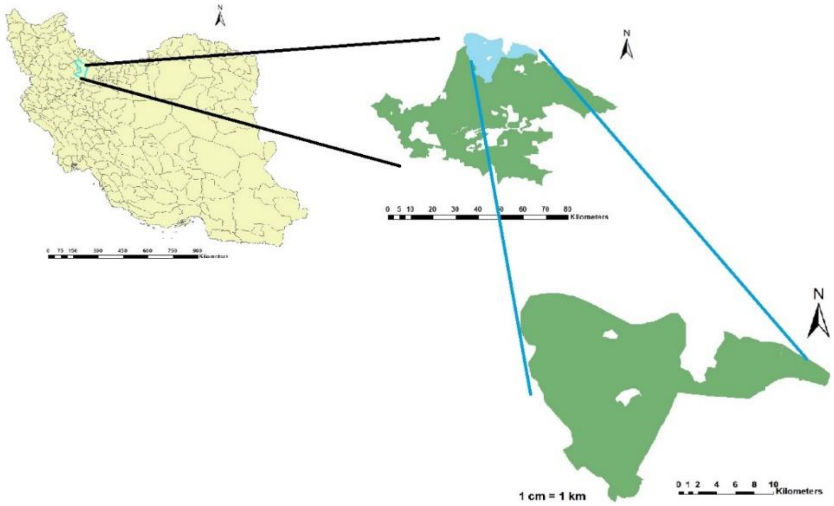

Figure 1. A- Location of Qazvin plain among the plains of Iran B- Location of Qazvin cities in Qazvin plain C- Agricultural area of Qazvin city

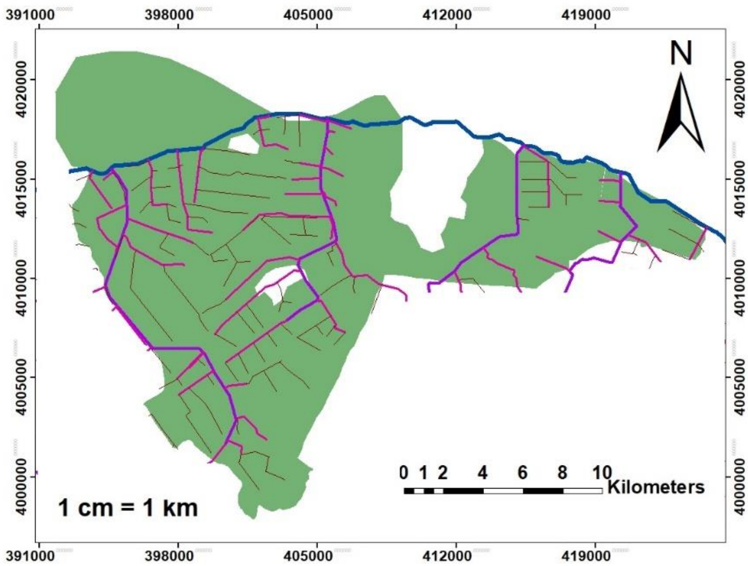

Figure 2. Location of irrigation network in the agricultural area of Qazvin County

In prediction algorithms, the goal is to predict a particular attribute based on another attribute. The predictable property is called a dependent variable, and the rest of the variables are called independent (Tavousi et al., 2015). Decision tree algorithms were used for modeling. In order to validate the models, the data was divided into two parts: training and test data. Models were constructed using training data, and the models were tested on test data. The percentage of samples of test data whose objective feature was correctly identified by the model expresses the accuracy of the model (Gupta, 2011). For all models used, 70% of the data was randomly selected as training data and the remaining 30% was tested as test data. For modeling, CART, CHAID, C5.0, and QUEST tree algorithms were used for modeling. The CHAID and QUEST decision trees solve classification problems, while the CART and C4.5 trees are used to solve both regression and classification problems (Pham, 2006). The C5.0 algorithm is a new version of the C4.5 algorithm that uses less memory than C4.5 when generating a set of rules (Pandya and Pandya, 2015). The experimental option chosen for the tree algorithms was the Cross-Validation decision with ten repetitions (10-Fold Cross Validation) because experiments have shown that the best choice for obtaining the most accurate estimate is section validation of the ten sectors (Ameri et al., 2013). There are various indicators, such as transparency, sensitivity, accuracy, and precision, for evaluating classification methods. It is also possible to calculate the error rate or incorrect classification based on the accuracy index. The criterion of classification error, or error rate, is exactly the opposite of the criterion of accuracy, and its minimum value (zero) is when the best performance is achieved. Also, its highest value (one) is when the lowest efficiency is achieved (Alizadeh et al., 2014; Han and Kamber, 2006). To evaluate the models and select the best model, the indicators of sensitivity or True Positive Rate (TPR), transparency (TNR), accuracy (ACC), positive predictive value or precision (PPV), false-positive rate (FPR), false-negative rate (FNR), F-measure (FM), geometric mean (GM), and error rate (ER) were used. The mentioned indicators are defined in equations (1) to (9), respectively.

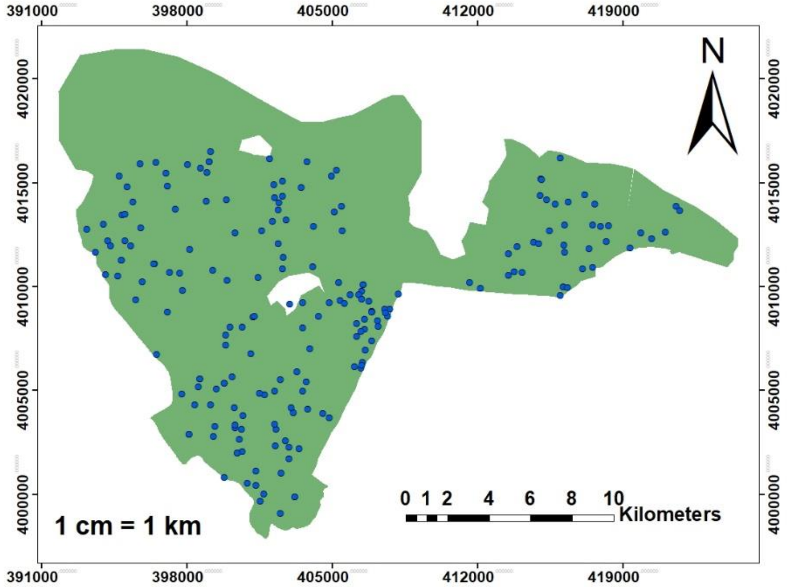

Figure 3 - Location of agricultural wells within the irrigation network of Qazvin County

In real problems, the classification accuracy criterion is by no means a good criterion for evaluating the performance of classification algorithms because, in relation to classification accuracy, the value of records in different categories is considered to be the same. Therefore, other criteria are used when dealing with unbalanced categories or when the value of a category is different from that of another category. In real problems, other criteria such as sensitivity and false positive rate are of particular importance. These criteria, which pay more attention to positive categorization, explain the classifier's ability to recognize a positive category. The sensitivity criterion shows how accurate the positive category is, and the false positive rate criterion expresses the false alarm rate with respect to the negative category (Seliya and Khoshgoftaar, 2011). The desired indices are calculated according to relations (1) to (9) (Han, 2000).

We have

$$TPR = \frac{TP}{TP + FN} \tag{1}$$

$$TNR = \frac{TN}{FP + TN} \tag{2}$$

$$ACC = \frac{TP + TN}{TP + FN + FP + TN} \tag{3}$$

$$PPV = \frac{TP}{TP + FP} \tag{4}$$

$$FPR = \frac{FP}{FP + TN} = 1 - TNR \tag{5}$$

$$FNR = \frac{FN}{TP + FN} = 1 - TPR \tag{6}$$

$$FM = \frac{2 \times TP}{TP + FP + FN} \tag{7}$$

$$\mathrm{GM} = \sqrt{\mathrm{TPR} \times \mathrm{TNR}} \tag{8}$$

$$ER = \frac{FP + FN}{TP + FN + FP + TN} = 1 - Accuracy \tag{9}$$

where TP is the number of positively labeled data correctly classified, FP is the number of negatively labeled data positively classified as positive, FN is the number of positively labeled data classified as incorrectly negative, and TN is the number of negatively labeled data that is properly categorized.

The data set used in this research is information about the plain of Qazvin County during the period of 2001-2005 on a monthly basis. The independent variables used, including human and natural factors, were selected. The amount of water output from agricultural wells (million cubic meters), the volume of irrigation network water (million cubic meters), the amount of agricultural water consumption (million cubic meters) as human factors, and the amount of rainfall (million cubic meters), temperature (Celsius), humidity (percentage) and evapotranspiration (mm / day) as natural factors were selected. Data from the amount of change in aquifer depth (meters) was also introduced to the K-Means algorithm as target data.CART, CHAID, C5.0, and QUEST tree algorithms were selected as the target variables to determine the best ratio between different clusters resulting from the output of the K-Means algorithm.

Qazvin county has 34 piezometers within the Qazvin plain. Considering the distribution of piezometers in different parts of the aquifer, to approximate the amount of utilization and effect of each piezometer from different parameters such as rainfall, water supply, and consumption, the Thyssen method was calculated to determine the range of effect of each piezometer. To achieve this, ArcGIS 10 software was used. In this regard, using GIS software, the first piezometer information layers and borders of Qazvin County in the Qazvin plain were prepared.

Table 1 contains a summary of monthly statistical characteristics related to the aquifer data of the agricultural area of Qazvin County within the irrigation network for a period of 15 years from 2001 to 2015.

Table 1. Statistical specifications related to groundwater depth within the irrigation network of Qazvin County

Scope of Study	Average Total (M)	Average Drop Values (M)	Average Rise Values (M)	Maximum Drop (M)	Maximum Rise (M)
Inside The Network	-0.16	-0.59	0.51	-4.61	3.3
Number	3780	2309	1471	1	1

3. Results and discussion

According to the results, it was found that the highest amount of aquifer loss occurs in the growing season and peak months of agricultural water consumption, with a temperature above 25 degrees and a moisture content of about 40%. The amount of water output from wells is directly related to the amount of drop, and with the monthly volume of more than one million cubic meters of agricultural wells, the probability of an aquifer drop will be higher. The clustering algorithm divided the studied data into 6 clusters.

Table 2. Results of clustering

Variables	First Cluster	Second Cluster	Third Cluster	Fourth Cluster	Fifth Cluster	Sixth Cluster
Volume of well outlet water	0.3±0.1	1.02±0.05	0.02	0.5±1	0.03	0.3±0.61
Evaporation and transpiration	0.62±2.8	0.9±4.9	0.31±1	0.13±5	0.1±1.1	0.1±3.79
Vol. of water in the whole net.	0.1±0.2	0.3±0.61	0.01	0.3±0.4	0.03±0.05	0.3±0.65
Temperature	3.5±17	1±24	2±4	0.6±26	2.5±6	1±18.31
Humidity	6±47	3±41	7±66	3±40	7±63	6±52
Agricultural water volume	0.18±0.32	0.6±1.08	0.03	0.5±1.01	0.03	0.3±0.65
Rainfall volume	0.1±0.3	0.1±0.58	0.2±0.47	0.03±0.06	0.3±0.06	0.3±0.57
Aquifer depth changes	0.01±0.2	0.01±0.1	0.01±0.2	0.2±0.1	0.2±0.5	0.1±0.15
Month						
October	33	25	0	0	0	0
June	0	100	0	0	0	40.3
January	0	0	25.5	0	35.5	0
July	0	0		49.7	0	0
November	35.6	0	0	0	51.6	0
May	0	85.8	0	0	0	100

Quantitative variables as "standard deviation ± mean" and nominal variables as "percentage" are reported

Table 3. Value of indicators for the models produced

	CART	CHAID	C5.0	QUEST
Sensitivity	0.90	0.96	0.93	0.90
Transparency	0.92	0.98	0.95	0.93
False-Positive Rate	0.03	0.02	0.03	0.08
False-Negative Rate	0.08	0.07	0.14	0.17
Precision	0.89	0.98	0.91	0.94
Accuracy	0.90	0.96	0.94	0.92
F-Measure	1.7	1.9	1.8	1.7
Error Rate	0.07	0.05	0.06	0.07
Geometric Mean	0.89	0.95	0.92	0.90

Table 2 shows the frequency of variables in clustering. Each cluster determines the amount of changes in the aquifer drop so that the first cluster (a drop of about 0.2 m), the second cluster (a drop of about 0.1 m), the third cluster (a rise of about 0.2 m), the fourth cluster (a drop of about 0.5 meters), the fifth cluster (a rise of about 0.5 meters), and the sixth cluster (a drop of about 0.15 meters). Also, we calculate the amount of water output from the well, the amount of network water, agricultural water, and rainfall in millions of cubic meters, the temperature in degrees Celsius, humidity in percent, evapotranspiration in millimeters per day, and the amount of aquifer depth changes in meters (the amount of drop with a negative sign).

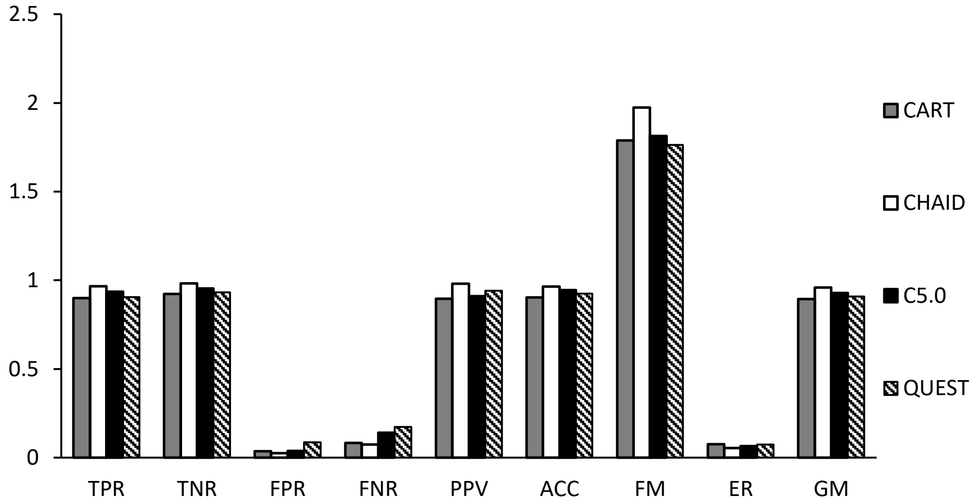

Figure 4. Chart of indicators for generated models

Table 3 shows the values obtained for the sensitivity, transparency, accuracy, precision, false-positive rate, false-negative rate, F-measure, geometric mean, and error rate for the four tree algorithms.

The values of the indicators presented in Table 3 show that the CHAID algorithm has produced the best model. Sensitivity, transparency, accuracy, positive predictive value, F-measure, and geometric mean index are the highest values for this model. The higher the value of these indicators, the more classifications are used in the right place. The false-positive rate, false-negative rate, and error rate have the lowest values for this model. Low values of these indicators confirm the occurrence of fewer errors in the classification of samples. Figure 4 shows a better comparison of the model index. Among the algorithms used, the best results are related to the CHAID algorithm, with an accuracy of 0.98 and a precision of 0.96. Table 4 shows the rules created by the CHAID tree algorithm.

Table 4. Some rules extracted from the CHAID tree algorithm

Number	Rules
1	If in April and October, the percentage of humidity is less than 60%, then, in the label of the "first cluster" category, the drop in the aquifer will be about 0.2 meters.
2	If in September, the total monthly volume of water, in the irrigation network is less than 0.76 million cubic meters, then, in the label of the "first cluster" category, the drop in the aquifer will be about 0.2 meters.
3	If the temperature is more than 14.6 in June, then, in the label of the "second cluster" category, the drop in the aquifer will be about 0.1 meters.
4	If the temperature is less than 14.6 in December, January and February, then, in the label of the "third cluster", the elevation in the aquifer will be about 0.2 meters above.
5	If the amount of evapotranspiration is more than 2 mm per day in July and August, then, in the "Cluster Four" category, the drop in the aquifer will be about 0.5 meters.
6	If in March the total monthly volume of water in the irrigation network is more than 0.07 million cubic meters and the amount of evapotranspiration is more than one millimeter per day, then, in the label of the category "fifth cluster", the rise in the aquifer will be about 0.5 meters.
7	If the amount of evapotranspiration is more than 2 mm per day in May, then, in the label of the "sixth cluster" category, the drop in the aquifer will be about 0.15 meters.

4. Conclusion

According to Rule One in Table 4, the percentage of moisture has the opposite effect on the amount of drop. That is, if the amount of moisture increases, the amount of moisture decreases. In the second rule, it is specified that in September, the amount of water in the network has the opposite effect on the amount of drop and that the more water in the network, the amount of drop decreases. If the temperature in June is higher than 14.6, then in the label of the category "second cluster," a drop in the aquifer is of about 0.1 meters. Also, in December, January, and February, if the temperature is less than 14.6, then, in the "third cluster" category, the elevation in the aquifer will be about 0.2 meters. As it is clear from the third and fourth rules, the amount of temperature is directly related to the amount of drop in the aquifer, and as the temperature increases, the amount of drop decreases more, and as the temperature decreases, the amount of drop decreases. In rules 4 and 5, it was found that the amount of evapotranspiration is directly related to the amount of drop, and the amount of drop decreases and increases with more or less the amount of evaporation and transpiration, respectively. According to rule 6, it is clear that the effect of the monthly volume of the total water in the irrigation network is greater than the effect of evaporation and transpiration on the drop. According to the results, it is clear that the most influential human factor on the amount of changes in aquifer depth is the amount of monthly volume of total water in the irrigation network, and the most natural factor on the amount of changes in aquifer depth is the amount of temperature.

References

Alizadeh, S., Teymourpour, B., Ghazanfari, M., 2014. Data mining and knowledge discovery. Iran University of Science and Technology (In Persian).

Ameri, H., Alizade, S., Barzegari, A., 2013. Knowledge extraction of diabetics' data by decision tree method. Journal of Health Administration, 16(53), 58-72.

Barikani, A., Ahmadian, M., Khalilian, S., 2011. Sustainable optimal utilization of groundwater resources in agriculture: a case study of qazvin plain agriculture. Journal of Economics and Agricultural Development (Agricultural Science and Technology), 25, 262-253 (In Persian).

Bonsor, H.C., MacDonald, A.M., Ahmed, K.M., Burgess, W.G., Basharat, M., Calow, R.C., Dixit, A., Foster, S.S.D., Gopal, K., Lapworth, D.J., Moench, M., 2017. Hydrogeological typologies of the Indo-Gangetic basin alluvial aquifer, South Asia. Hydrogeology Journal, 25(5), 1377-1406.

López-Corona, O., Fuentes, O.E., Morales-Casique, E., Longoria, P.P., Moran, T.G., 2016. Data Mining of Historic Hydrogeological and Socioeconomic Data Bases of the Toluca Valley, Mexico. Journal of Water Resource and Protection, 8(4), p.522.

Gupta, G.K., 2011. Introduction to Data Mining with Case Studies. (2nd ed) Prentice Hall of India

Han, J., Kamber, M., 2006. Data Mining: Concepts and Techniques. 2nd ed.Morgan Kaufman.

Jang, C.S., Chen, S.K., Kuo, Y.M., 2013. Applying indicator-based geostatistical approaches to determine potential zones of groundwater recharge based on borehole data. Catena, 101, 178-187.

Khan, S., Gabriel, H.F., Rana, T., 2008. Standard precipitation index to track drought and assess impact of rainfall on watertables in irrigation areas. Irrigation and Drainage Systems, 22(2), 159-177.

Konapala, G., Mishra, A.K., Wada, Y., Mann, M.E., 2020. Climate change will affect global water availability through compounding changes in seasonal precipitation and evaporation. Nature Communications, 11(1), 1-10.

Kotsiantis, S., Kostoulas, A., Lykoudis, S., Argiriou, A., Menagias, K., 2008. Using data mining techniques for estimating minimum, maximum and average daily temperature values. International Journal of Mathematical, Physical and Engineering Sciences, 1(1), 16-20.

Lopez, M.I., Luna, J.M., Romero, C., Ventura, S., 2012. Classification via clustering for predicting final marks based on student participation in forums. International Educational Data Mining Society.

Mirhashemi, S.H., Haghighat jou, P., Mirzaei, F., Panahi, M., 2020. The study of environmental and human factors affecting aquifer depth changes using tree algorithm. International Journal of Environmental Science and Technology, 17(3), 1825-1834.

Mohammadi, M., Mohammadi Qalehuni, M., Ebrahimi, K., 2011. Temporal and spatial changes of groundwater quality in Qazvin plain. Iranian Journal of Water Research, 8, 52-41. (In Persian).

Novinpour, E.A., 2017. A study of the relationship between the exploitation and subsidence of Salmas. Urban Management, 15(45), 319-326.

Oorkavalan, G., Chidambaram, S.M., Mariappan, V., Kandaswamy, G., Natarajan, S., 2016. RETRACTED: Cluster analysis to assess groundwater quality in Erode District, Tamil Nadu, India. Circuits and Systems, 7(6), 877-890.

Pandya, R., Pandya, J., 2015. C5. 0 algorithm to improved decision tree with feature selection and reduced error pruning. International Journal of Computer Applications, 117(16), 18-21.

Pham H, editor., 2006. Springer Handbook of Engineering Statistics. Germany: Springer.

Seliya, N., Khoshgoftaar, T.M., 2011. The use of decision trees for cost-sensitive classification: an empirical study in software quality prediction. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 1(5), 448-459.

Sotoudehnia, A., Sotoudehnia, S., 2016. Investigating the role of taleghan reservoir dam construction on sedimentation of Qazvin plain irrigation network, 15th Iran hydraulic conference, Faculty of Engineering, Imam Khomeini International University, Qazvin December 15-17, 2016 (In Persian).

Stumpp, C., Żurek, A.J., Wachniew, P., Gargini, A., Gemitzi, A., Filippini, M., Witczak, S., 2016. A decision tree tool supporting the assessment of groundwater vulnerability. Environmental Earth Sciences, 75(13), 1-7.

Tavousi, A., Sepehri, M.M., Malakoutian, T., Khatibi, T., 2015. Data mining approach in prediction of erythropoietin dosage in hemodialysis patients. Journal of Mazandaran University of Medical Sciences, 25(129), 26-35.

Zhao, Y., Li, Y., Zhang, L., Wang, Q., 2016. Groundwater level prediction of landslide based on classification and regression tree. Geodesy and Geodynamics, 7(5), 348-355.