Türkiye Jeoloji Bülteni
Geological Bulletin of Turkey
64 (2021) 79—92
DOI: 10.25288/tjb.1010636

K-Means Clustering and General Regression Neural Network Methods for Copper Mineralization probability in Chahar-Farsakh, Iran

Adel Shirazy[1]*, Ardeshir Hezarkhani[1], Aref Shirazi[1], Shayan Khakmardan[1,2], Reza Rooki[3]

[1]Amirkabir University of Technology, Tehran, Iran
[2]University of Lorraine, Nancy, France
[3]Birjand University of Technology, Birjand, Iran

• Geliş/Received: 16.10.2021 • Düzeltilmiş Metin Geliş/Revised Manuscript Received: 04.12.2021 • Kabul/Accepted: 07.12.2021
• Çevrimiçi Yayın/Available online: 10.01.2022 • Baskı/Printed: 20.01.2022
Araştırma Makalesi/Research Article *Türkiye Jeol. Bül. / Geol. Bull. Turkey*

Abstract: Due to the efficiency of data mining science for analyzing and reviewing extensive data, especially geochemical data, essential methods and techniques such as the hierarchical method, K-Means method, density-based methods, Cohennon method, and so forth, have been developed and utilized by numerous researchers for clustering. One of the most notable and widely used algorithms in the field of clustering is the K-Means algorithm. This algorithm divides the data into K clusters by emphasizing the distance criterion. This study focuses on applying this method according to lithogeochemical data taken from the 1:100,000 scale map of Chahar-Farsakh in South Khorasan province for the elements of copper, cobalt and nickel to the sampling coordinates. The optimal value of K was classified according to the desirability of the selection and the data, and thus the relationships between these elements in the range were determined. This was analyzed by changing the value of K from 3 to 15 criteria mentioned in each class to reveal the optimal K. According to the observations, the existence of a quadratic relationship with negative concavity between copper and cobalt elements, as well as a special exponential relationship between copper and nickel and a positive linear relationship between nickel and cobalt, were reported. Finally, considering the coordinates of the samples and the concentration of cobalt and nickel, the quantity of copper was predicted using a General Regression Neural Network (GRNN). The accuracy of this method was estimated to be 0.99 on training data and 0.76 on test data. Therefore, using the proposed method (K-means Clustering and GRNN) in this paper, it is possible to examine the extent of changes in other elements in the analysis. Also, it is possible to make deeper and broader explorations via determining the relationship between the elements.

Keywords: Behavioral Measurement, Chahar-Farsakh, Desirability, General Regression Neural Network, K-means, Prediction.

INTRODUCTION

One of the essential perspectives in data mining science for analyzing and examining large volumes of data and samples with different characteristics is the clustering perspective, which itself includes essential methods and techniques such as those studied by Anderberg (Ullman, 1984) and has been used by various researchers so far (Shirazi et al., 2018a, b; Shirazy et al., 2020a). Due to the expansion of industry and the need for more and more raw minerals, the importance of mineral exploration is felt more than ever in today's world. New methods in the statistical analysis of mineral data help identify better mineral reserves,

* Correspondence / Yazışma: adel.shirazy@aut.ac.ir

including geostatistical methods and methods based on artificial intelligence and machine learning (Shirazy et al., 2021a, b). Using advanced and intelligent methods, the accuracy of our performance in the field of mineral identification and exploration increases. Today, due to the advancement of technology and the rising level of technical knowledge, secondary mines have also been raised. Secondary mines have been mined in the past with a higher threshold grade, but today, low-grade minerals can be processed in them with the help of advanced technology (Alahgholi et al., 2018; Shirazy et al., 2019).

Each of the methods used in statistical studies has special strengths and applications. However, these methods may have weaknesses and limitations. One of the applied techniques in mineral data analysis (due to the large volume of data) is the clustering method. The clustering technique is used to classify data. Accordingly, it has applications in every branch of science. The application of clustering methods (Hajnajafi et al., 2021) (e.g., K-means) is performed to analyze exploratory data, such as in geochemistry and geophysics, and to categorize data to achieve a specific goal (Khayer et al., 2021; Shirazy et al., 2021a,b). One of the applications of the clustering method is in the analysis of geochemical data from the sampling of stream sediments. In this case, it is usually used to study the geochemical behavior of the elements and classify their concentration data. The clustering technique can also be used in remote sensing (satellite image analysis) to identify alterations and lithology based on the wavelengths of satellite imagery bands (Shirazi et al., 2018b,c ; Shirazy et al., 2021c; Khosravi et al., 2022).

The function of clustering analysis is to place data with similar characteristics in a specific group. Therefore, after applying clustering analysis to the data, instead of dealing with a large amount of data, we are faced with a series whose members are very similar in terms of features. The purpose of using clustering analysis is to present the classes whose members are most similar to each other and more different from the members of other classes. Existing behaviors can be identified based on the similarities that each member of a class has with each other. Identifying the behavior of the data (the behavior of each class relative to each other) is a great help in analyzing earth-related sciences. The clustering technique is one of the indirect analytical methods. This means that, in using this method, there is no need to obtain information about the internal structures of the members, and, in fact, by using clustering, the hidden patterns in the data are identified. Using the results of this method, the process of direct studies can be improved.

One of the important and practical analytical methods in clustering is the K-means method. This method classifies data based on the Euclidean distance. The distance of each datum from the center of the cluster should be as small as possible. To achieve this goal, start with a small number of classes and increase the number of classes as much as possible. As a result, after plotting the cluster utility function, the appropriate number of clusters are identified, and clustering is performed based on that.

The K-means clustering analysis technique is used to study the behavior of the data relative to each other. Among the applications of this method that other researchers to date have performed is the study of the impact of vegetation on the process of restoring water health, the geological classification of land, the identification of geochemical patterns and geochemical behavior of elements in mineral studies, and predicting the amount of carbon in intelligent systems, as well as greenhouse gas emissions and their impact on the urban environment.

In recent decades, the use of methods based on artificial intelligence, including artificial neural networks (ANNs), has become particularly

important. Artificial neural networks are programmed based on the processing capabilities of the human brain, including knowledge, control, speech, and prediction. Managing large volumes of data with complex relationships and many patterns, reasoning on ambiguous data, and providing fast and adequate responses are some of the capabilities of artificial neural networks (ANNs) (Shirazi et al., 2018d). In this study, by performing a string of codes in MATLAB software and using SPSS software, the capabilities of the mentioned methods were investigated. Based on the obtained results, it is possible to introduce a process to improve exploration operations in the studied area and reduce the cost of geochemical operations in neighboring areas.

In this study, the other data can be predicted with high accuracy by combining the K-means clustering method and the General Regression Neural Network (GRNN) based on the behavior of the available data (Geochemical Data). The K-means clustering method was used to investigate and observe the behavior of Cu and other geochemical data. The artificial neural network was then used to create an estimator model based on Co and Ni data for Cu estimation (Shirazi, et al., 2018e). The main advantage of the combined method based on K-means clustering and ANN is the possibility of estimating data (Shirazy et al., 2021d, e; Shirazy et al., 2020A) based on each other with high validity in this method without the need for the simultaneous measuring of Cu and Co, Ni and their locations data. In this way, in addition to saving time, financial costs are also significantly reduced.

MATERIAL and METHODOLOGY

First of all, it should be noted that in this study, 637 ICP-MS analyses of stream sediments were used. These data, collected by the Geological Survey of Iran, are highly accurate with careful examination of duplicate samples.

In this view, it is assumed that the dataset M consists of m samples, as $\{x_1, x_2,\ldots, x_m\}$, and that each sample is defined as a vector in the M dataset. This vector represents different characteristics for that sample. It is assumed that these samples are classified in K class as C_1, C_2, to C_K (Menard, 1995). It should be noted that for this classification, a series of primary hypotheses must be considered. These hypotheses are then summarized in equations (1) to (3) (Ghannadpour et al., 2013; Pelleg & Moore, 2000).

$$C_i \neq Æ \; for \; i = 1,..., K \quad (1)$$

$$C_i \cap C_{j=} Æ \; for \; i \neq j \quad (2)$$

$$\bigcup_{I=1}^{k} C_i = M \quad (3)$$

According to Equations (1) to (3), the first hypothesis implies that each of the K categories must have a member (sample), and an empty category must not exist. Regarding the second Equation (hypothesis), the samples should only be in one category, and the categories must not have a standard sample. The third hypothesis states that all samples should be categorized, and there should be no sample without a class or category.

Among the most widely-used clustering techniques in data mining is the K-Means method, which has been extensively studied by various researchers and which tries to cluster a number of samples with a certain number of categories (K) so that the sum of the Euclidean distances of each specimen from the center of the category to which it is assigned should be minimized. In several studies, the behavior of the elements is measured in relation to each other (Tarkian & Stribrny, 1999; Xu, 2012). To gain a proper analysis from the behavior of existing data and the analysis relative to each other, numerous studies are found that, with the help of clustering-based methods, try to classify the existing analyses of a study.

One of these studies is Frederick's radiometric study of copper in 1984, using the K-Means method. Recent studies (Cheung, 2003; Murthy & Chowdhury, 1996; Krishna & Murty, 1999) suggest that the K-Means method has been studied using advanced ultra-innovative algorithms, and outstanding results have been obtained. In another study, a meta-innovative method called the GK method is proposed for clustering, including a combination of genetic algorithms and the K-Means method (Yaghini & Gereilinia, 2013). In this method, an operator called the K-Means operator is used, which is derived from the K-Means algorithm, and it is used instead of a hybrid open operator in the GKM algorithm. The results of using this algorithm for different values of K from 4 to 10 are much better and more acceptable than the results of the K-Means method.

One of the problems of the K-Means method is dealing with large volume datasets and large K values, which causes problems in reaching the appropriate answer. According to Cheung (2003), it has been shown that this method is usually trapped at local optimal points. Therefore, in studies with large datasets, more advanced methods should be used to produce a reliable answer (Shirazy et al., 2020b). However, the above problem is not the case in the present study due to the appropriate volume of data (Shirazy et al., 2018a, b).

K-Means Algorithm

The K-means algorithm method starts with a specific value for K (number of categories) and tries to categorize a set of specific samples in this K group so that the hypotheses expressed in Equations (4) and (5) are observed (Hezarkhani & Ghannadpour, 2015). It should also be noted that the criterion for assigning each sample to a class in this method is the minimum Euclidean distance of each sample from the center of each class or category. The main steps that take place in this algorithm (Hamerly & Elkan, 2003) are summarized as follows:

- Form K class or batch as $\{C_1, C_2, \ldots C_k\}$ to cluster m sample from set M.
- Calculate the vector Z_j based on Equation 4, representing each Cj category's center or representative.

$$z_j = \frac{\sum_{x \in c_j} x}{\#c_j} \; for \; j = 1. \ldots . k \qquad (4)$$

In this regard, x represents the vector of an instance that is a member of C_j, and $\#C_j$ represents the number of samples that are members of the C_j class. It should be noted that Equation (4) is used to calculate the center of each class during the solvation process, and usually, at the beginning of the algorithm, K samples are randomly selected and considered as the center of each batch.

- Calculate the objective function resulting from the classification $\{C_1, C_2, \ldots C_k\}$, based on Equation (5), which calculates the sum of the distances of the samples from the center of the categories.
- Minimize the objective function of Equation (5) and find the appropriate classification on the dataset M with the number of K categories.

$$f(C_1.C_2.\cdots.C_k) = \sum_{j-1}^{k} \sum_{x \in c_j} |X - z_j|^2 \qquad (5)$$

General Regression Neural Network

Artificial neural networks, or more simply Neural networks, are cutting-edge systems and computational approaches for machine learning, knowledge presentation, and, eventually, the application of gained information to maximize the output reactions of complex systems. The basic principle behind such networks is inspired in part by how the human brain system processes data and information in order to learn and produce knowledge. The critical element of this idea is to create new structures for the information processing system (Schalkoff, 1997).

The system comprises a huge number of extremely linked processing components known as neurons, which collaborate to solve problems and send information via synapses. In these networks, if one cell is damaged, other cells can compensate for its absence and participate in its reconstruction. These networks are able to learn. For example, by injecting burn to tactile nerve cells, the cells learn not to go to the hot surfaces, and with this algorithm, the system learns to correct its error (Yegnanarayana, 2009). Learning in these systems is adaptive. In other words, using parables, the weight of the synapses adjusts in such a manner that the system generates the right response when fresh inputs are provided (Dayhoff & DeLeo, 2001). Artificial neural networks have many applications, especially in geological and mining engineering, and also it can be considered a successful estimation method in various fields of earth sciences. The schematic structure of a single-layer artificial neural network can be seen in Figure 1.

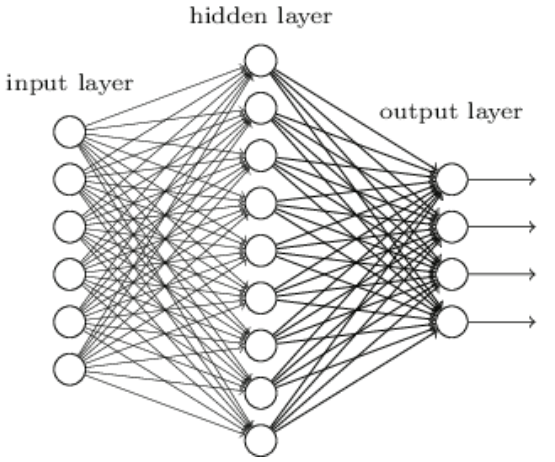

Figure 1. Schematic representation of a single-layer artificial neural network (Abraham, 2005).

The general regression neural network algorithm can be considered a normalized radial network (Demuth & Beale, 1993) with one hidden neuron for each training unit. This network of single-pass parallel learning algorithms with a parallel structure can generate continuous outputs (Specht, 1991). These networks are based on the probability density function; one of these networks' salient features is the rapid training time and modelling of nonlinear functions. This network provides smooth changes from observational data to other data, even with scattered data in multi-dimensional measurement space. The algorithmic form of this network can be used for any regression problem where there are no assumptions for linear judgment. This network does not have the parameters of the error propagation network but instead does not have the error smoothing factor; instead, the smoothing factor is obtained by considering the mean squared error (Artun et al., 2005), as shown in Figure 2. The structure of this network is similar to the general structure of the radial network; there is only a slight difference in the second layer.

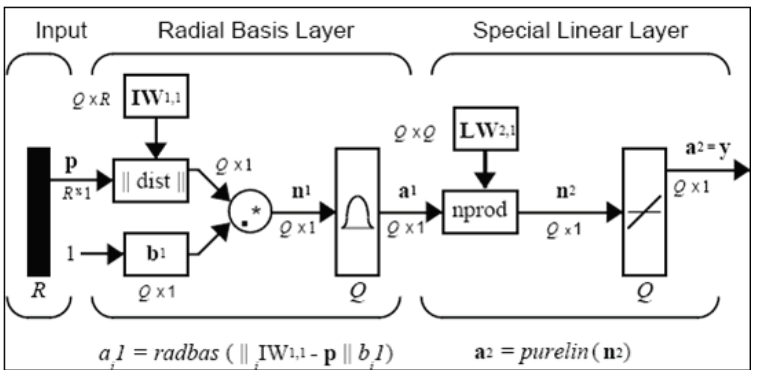

Figure 2. General regression neural network structure (Demuth & Beale, 1993).

Geographical Location and Geology

The 1:100,000 map of Chahar-Farsakh from South Khorasan province is located between longitude 59.3° to 60° and latitude 31.5° to 32°. This map has the prominent elements of Kuh-e Bozorg, Kuh-e Sorkh, Kuh-e Shisheh and Kuh-e Sarbisheh, and is so named due to the location of Chahar Farsakh village on this map.

This map, prepared by the Geological Survey of Iran, code 7953, is a suitable area for metal mineralization (Figure 3).

The study area is located 24 km northwest of Nehbandan city. This area is a part of the intrusive suits in the eastern part of Lut structural zone. The studied granitoid bodies (Chahar-Farsakh) include two types of acidic and intermediate–basic. In addition, the lithological composition of this mass is granite, granodiorite, diorite, pegmatite–aplite, and the gabbroic suit, belonging to the melange ophiolite complex of the region. In this area, sedimentary sequences are composed mainly of shale, sandstone and conglomerate. Infiltrations in the region have caused shale-sandstone metamorphism and metamorphic rocks with andalusite schist composition. The typical texture in acidic rocks, consisting of granite, granodiorite, and pegmatite-aplite, is granitic, perthite, aplite and pueiclitic. The usual texture in the studied diorite is granitic, and gabbro has a granular hypidiomorphic texture. Granite, pegmatite and aplite are composed of quartz, K-feldspar and albite minerals. Granodiorite comprises quartz, the albite-oligoclase type of plagioclase and K-feldspar, while diorite is composed of plagioclase and hornblende, and gabbro is made up mainly of plagioclase and augite type pyroxene. Common alterations in acidic rocks are the conversion of plagioclase to sericite. In basaltic rocks up to the middle rocks of the region, pyroxene is converted to hornblende and plagioclase to sericite (Ghorbani, 2013a, 2013b; Moritz, 2016).

K-Means Studies

The relationship between molybdenum and copper has been investigated using the K-Means method. By examining the relationships and changes in these elements, a broad view of the field and exploration can be obtained (Heil, Häring, Marschner, & Stumpe, 2019).

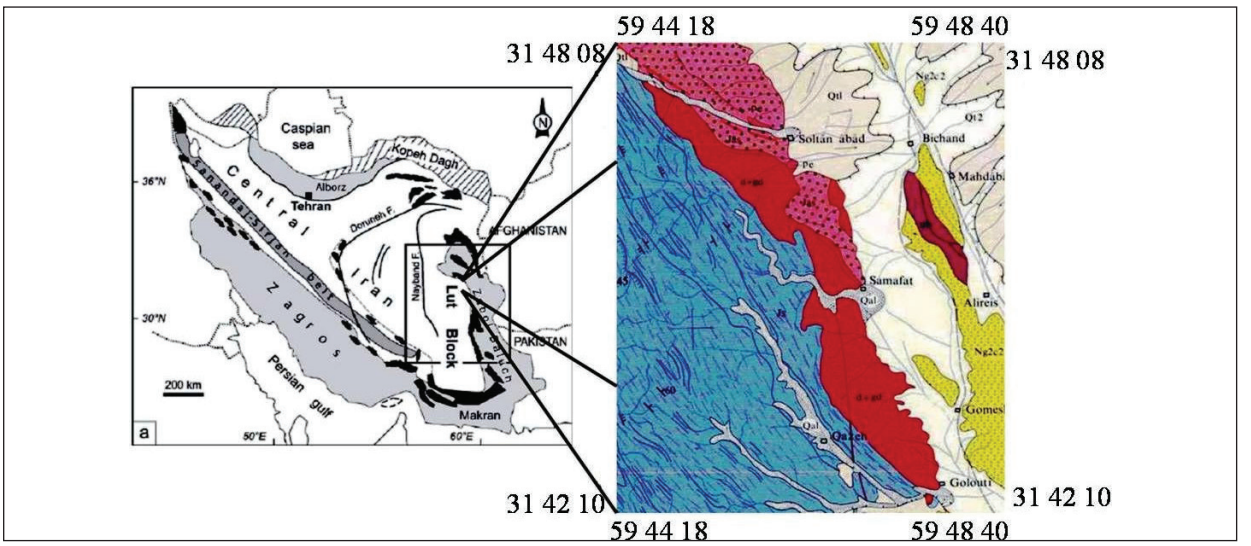

Figure 3. Simplified map of Chahar-Farsakh with its location on the structural map of Iran.

These analyses can be made possible by other classifications, such as lead and zinc, gold and silver, and so forth, because these elements have an almost high correlation compared to other elements. In order to determine the appropriate K to determine the number of classes, the number of categories was changed from k=2 to k=10, and the resulting groupings were analyzed. Then, to evaluate the determined groups for different values of k, the desirability relationship was determined in accordance with Equation (6), and the performed classifications were evaluated accordingly (Murthy & Chowdhury, 1996).

$$S(i) = \frac{Min\,(Aveg\_Between(i.k)) - Aveg\_Within(i)}{Max\,[Aveg\_Within(i).Min(Aveg\_Between(i.k))]} \qquad (6)$$

The profile of the classes and the degree of desirability of each sample for the optimal performed classification for the values of k=3 to k=5 of the two elements, such as copper and cobalt, are shown in the Figure 4.

As shown in Figure 4 and Figure 5, the trend of desirability change is shown for 3 to 15 classes; the best class number for the above two elements is four classes. After dividing into four classes, the centers of the categories were drawn in the diagram of Figure 6, and a trend line with the formula $y = -0.0029x^2 + 0.5183x - 4.5085$ was fitted to it. This curve shows the behavior of copper to cobalt. The behavior is in the form of a negative concavity curve that increases with the amount of copper up to 90 ppm, and also the amount of cobalt increases and then proceeds to a decreasing trend.

Spearman correlation coefficient between these two elements is 0.755. However, it can be seen that the K-Means method is completely superior to other behavioral methods.

The profile of the classes and the desirability value of each sample for the best classification performed for the values of K = 5 to K = 6 of the two elements copper and nickel are shown in the diagrams of Figure 7. For classification up to K = 15, the degree of desirability is shown in Figure 8.

According to the curve fitted to the centers of the categories shown in Figure 9, it can be seen that the unexpected behavior of copper versus nickel has a function.

$y = -0.0027x^4 + 0.7125x^3 - 54.827x^2 + 1609.7x - 15694$

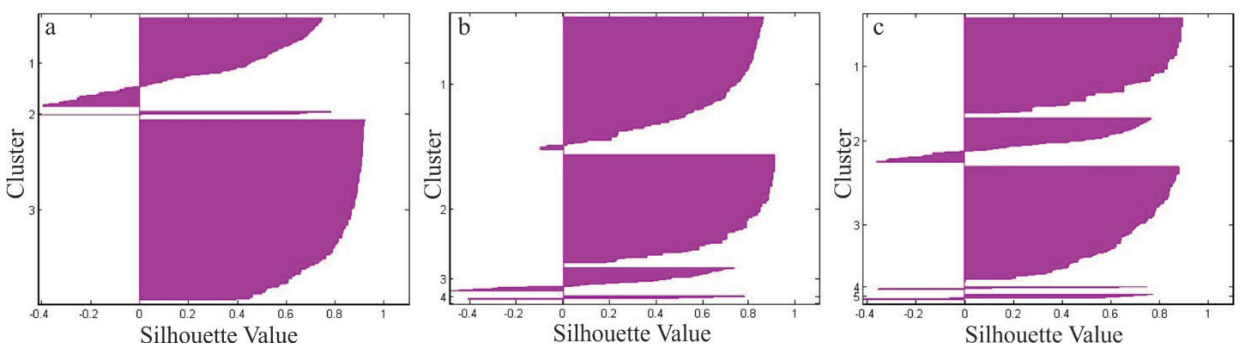

Figure 4. Cluster profile and desirability values of 3 to 5 classes for copper and cobalt. **a)** Classification of three classes with average value of 0.6498, **b)** Classification of four classes with average value of 0.6525, **c)** Category of five classes with average value of 0.6247.

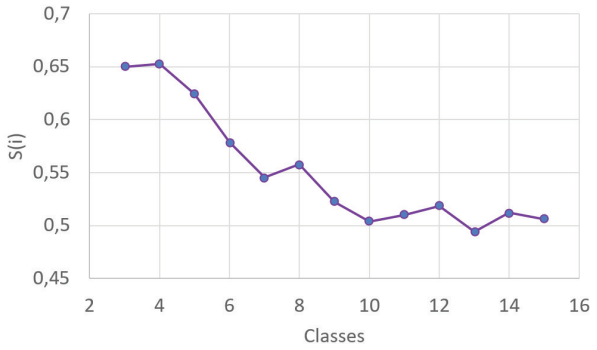

Figure 5. Validation criterion S(i) based on the number of clusters, copper vs cobalt.

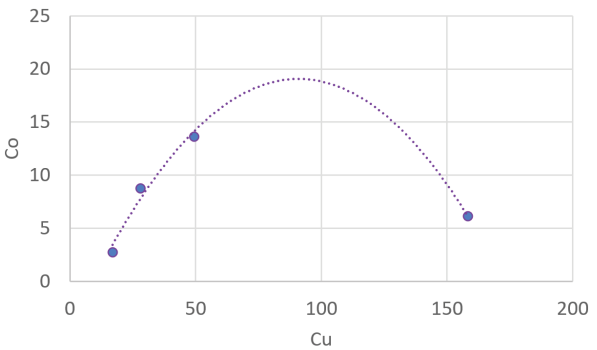

Figure 6. Best-fitted curve to centers of categories for copper and cobalt.

This indicates a very high rise in nickel with copper. When copper reaches about 130 ppm, a drop in nickel concentration begins.

Since the behavior of nickel and cobalt can also show other perspectives of behavior measurement, Figure 10 shows the desirability diagram of 3 to 6 classes. In order to identify the maximum desirability, the desirability value for these two elements was calculated up to 15 classes (Figure 11). By examining this diagram, it can be inferred that the maximum desirability has occurred in 3 classes for the two elements of nickel and cobalt. After calculating the category centers, due to the optimality of the three classes, three category centers were also selected, the centers were drawn, and the trendline fitted to justify the changes. This trendline has the function of y = 9.7136x - 29.426 and a confidence level of 96% (Figure 12). The linear structure of this behavior can be a reasonable justification for the accuracy of the Spearman correlation coefficient, which is estimated to be 0.58. This means that this positive linear correlation indicates a gradual increase in nickel with cobalt.

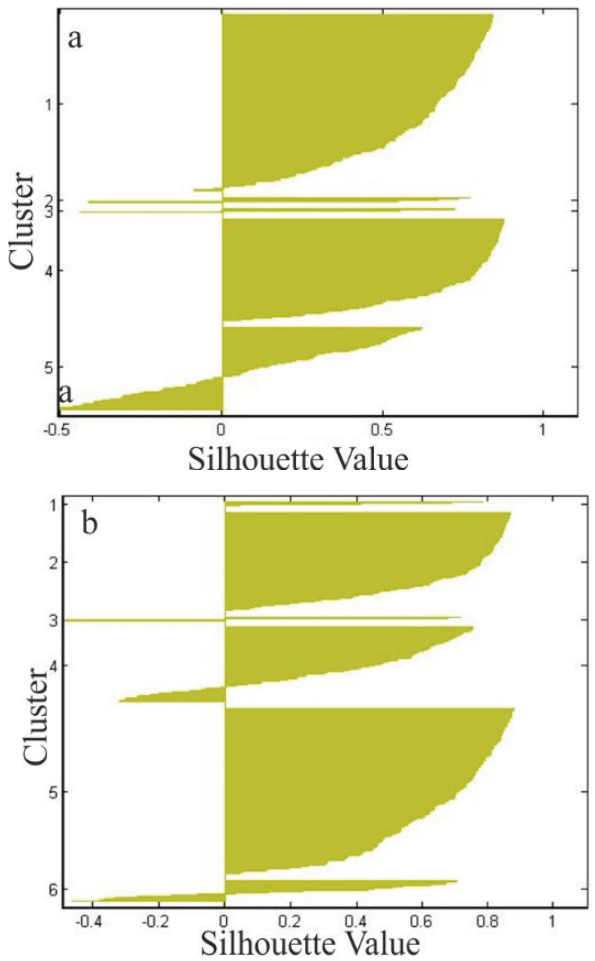

Figure 7. Cluster profile and desirability values with 5 and 6 classes of copper and nickel. **a)** Category of five classes with average value of 0.5104, **b)** Classification of six classes with average value of 0.5635.

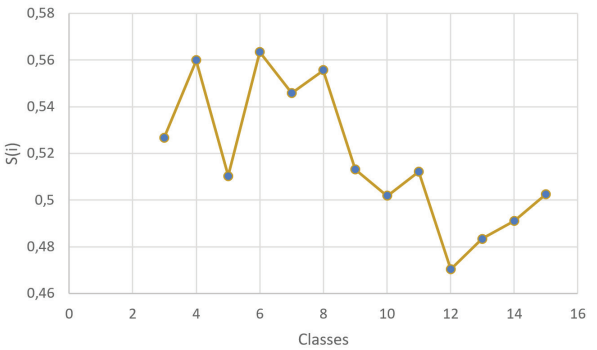

Figure 8. Validation criterion S(i) based on the number of clusters, copper vs nickel.

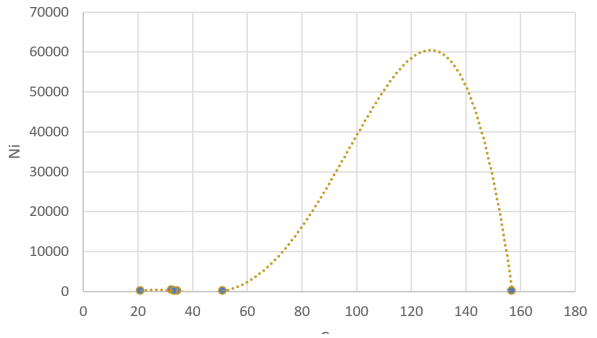

Figure 9. Best-fitted curve in center of categories for copper and nickel.

One of the essential points in the field of behavioral measurement is to pay attention to the structural discussion; in other words, the sampling position can affect the behavior of that sample compared to another sample. Based on this, the behavior of three elements, copper, cobalt and nickel, along with their harvest position, was investigated. This evaluation resulted in the graphs of Figure 13 as desirability graphs in 3 to 6 classes. According to the previous procedure, the desirability value for 15 classes is given in Figure 14, and, as can be seen, four classes of maximum desirability have been calculated. The centers of the obtained four classes are illustrated in Table 1.

Table 1. Category centers for copper, cobalt and nickel with coordinates in 4 classes.

X	Y	Cu	Co	Ni
750592.4	3530995	31.7093	7.55814	35.90116
771446.3	3523393	25.86792	7.698113	38.18239
771694.9	3498900	25.38994	5.805031	35.40252
753990.5	3507895	23.55782	5.931973	31.63946

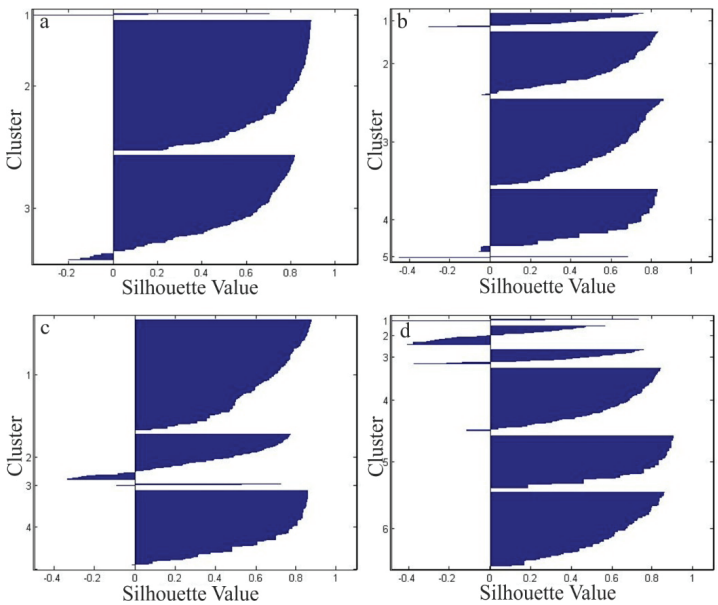

Figure 10. Cluster profile and utility values with 5 and 6 grades of nickel and cobalt. **a)** Classification of three classes with average value of 0.6486, **b)** Category of five classes with average value of 0.5841, **c)** Classification of four classes with average value of 0.5921, **d)** Classification of six classes with average value of 0.5937.

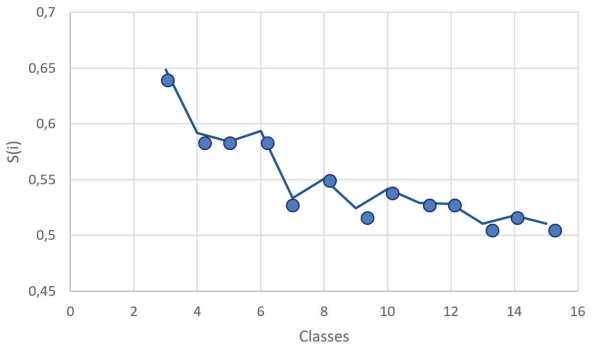

Figure 11. Validation criterion S(i) based on the number of clusters, nickel vs cobalt.

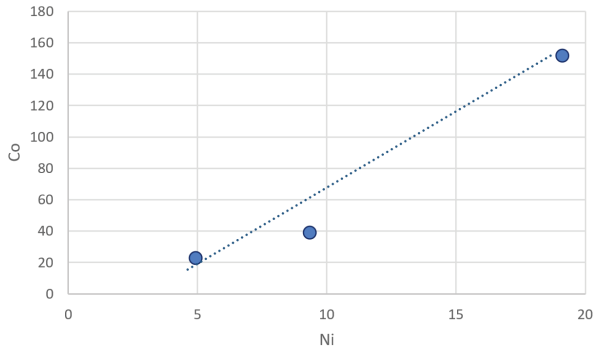

Figure 12. Best-fitted trendline in center of the classes for nickel and cobalt.

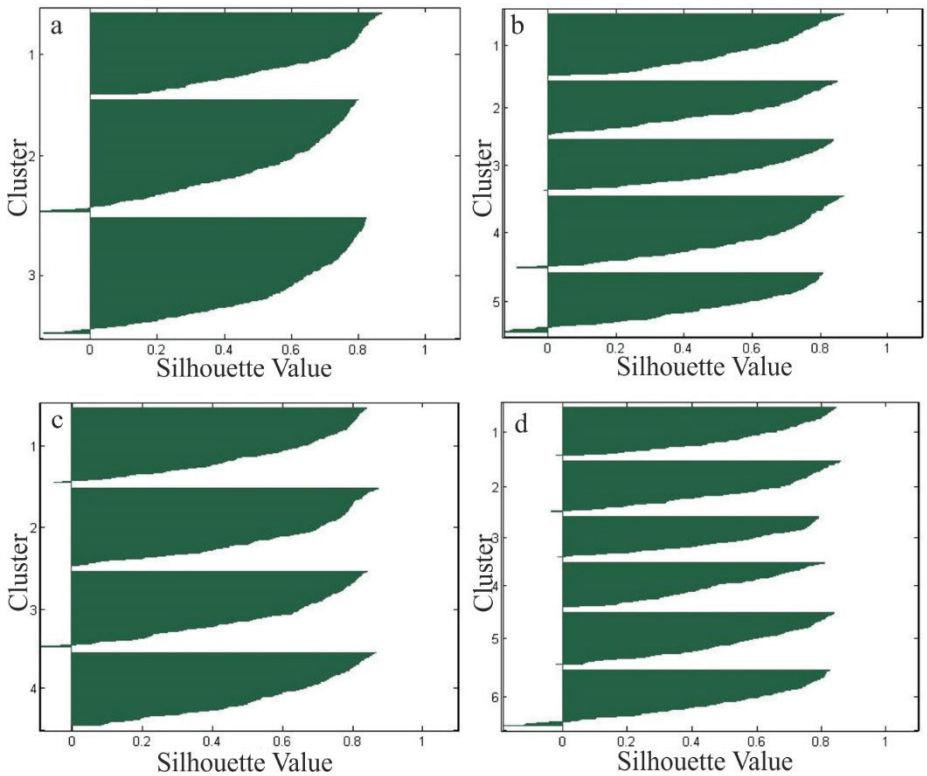

Figure 13. Cluster profile and utility values with 5 and 6 grades of nickel and cobalt. **a)** Classification of three classes with average value of 0.5679, **b)** Category of five classes with average value of 0.5642, **c)** Category of four classes with average value of 0.5769, **d)** Classification of six classes with average value of 0.5228.

Estimation Evaluation with General Regression Neural Network

In this study, 637 data were used for the train (449 data) and test (188 data) of GRNN. Input parameters of GRNN are the locations of samples (X, Y) and the copper grade was used as the output of the network. Due to the need to determine the optimal radius to estimate the best value in this method, different values from 0 to 1 were selected experimentally. The optimal value of 0.015 was selected for the impact radius.

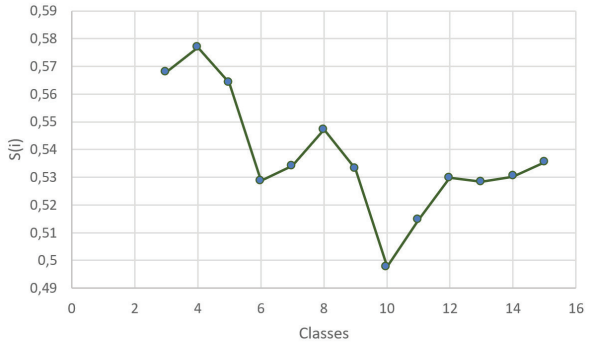

Figure 14. S(i) validation criterion based on number of clusters, copper, cobalt and nickel with sampled coordinates.

Figures 15 and 16 show the copper grade estimation trendline of actual point values in the training and test data, respectively.

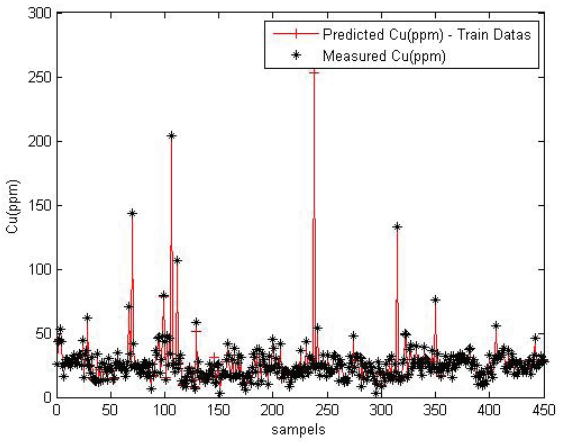

Figure 15. Copper estimation line with Lab-values in the training data.

Confirmation of the estimation accuracy was made by putting regression of predicted values against measured values in each coordinate for two categories of train and test data, as given in Figures 17 and 18, respectively.

The accuracy (R) of these estimates was 0.99 in the train data and 0.76 for test data.

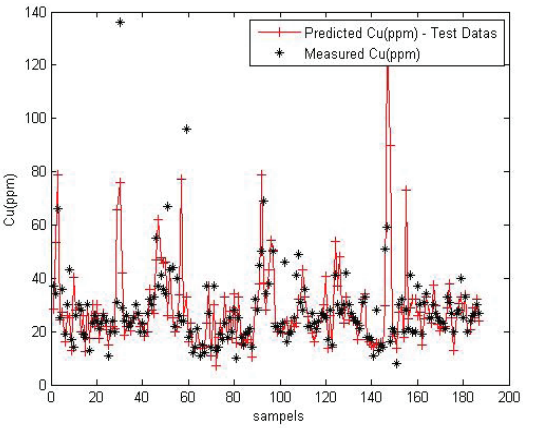

Figure 16. Copper estimation line with Lab-values in the test data.

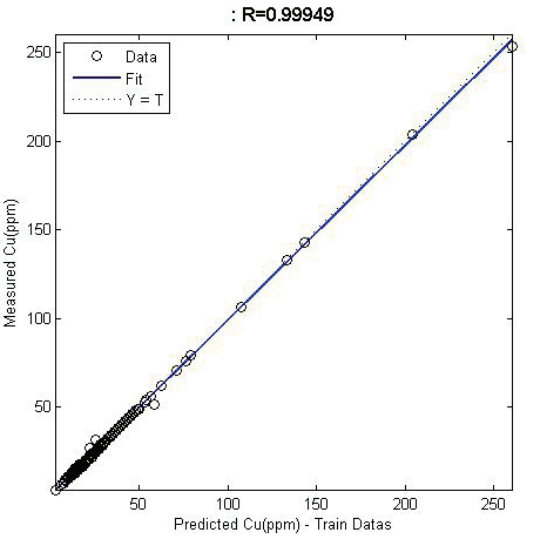

Figure 17. Regression of predicted data versus Lab-data (Train Data).

For the final step, the predicted map of the Chahar Farsakh area was generated and is shown in Figure 19. According to Figure 19, it can be understood that the accumulation of copper in this area is in the north and northwest. It is recommended that researchers study these areas for further investigation.

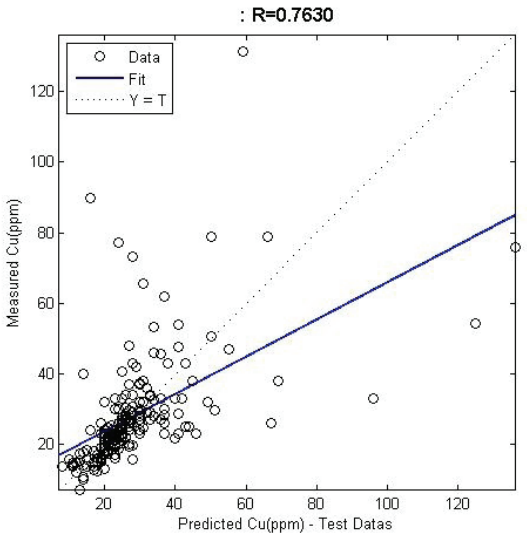

Figure 18. Regression of predicted data versus Lab-data (Test Data).

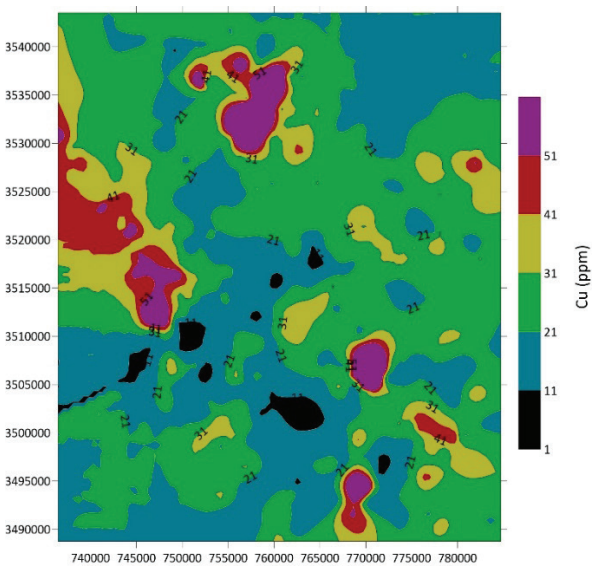

Figure 19. Predicted Cu in Chahar Farsakh area.

CONCLUSION

In this study, by using the K-Means method, lithogeochemical samples were analyzed in the Chahar Farsakh region, and the relationships between copper, cobalt and nickel and sampling coordinates were investigated. Using the mentioned method indicated a quadratic relationship with a negative concavity between copper and cobalt elements, a special exponential relationship between copper and nickel, and a positive linear relationship between nickel and cobalt. An appropriate criterion was presented to determine the desired K, which is, in fact, the desired number of classes for classification. By changing K from 3 to 15, a geometric relationship obtained the best results between the studied elements. The behavior of these elements towards each other and towards the recorded coordinates was obtained. On the other hand, the behavior of copper was predicted by the method of general regression neural network, which is a kind of radial network. This prediction resulted in finding a radius of 0.015 as the optimal radius of the network. Besides, the confidence level (accuracy) resulted in 0.99 for train data and 0.76 for test data. Consequently, the K-Means method can be used to investigate the relationships between other elements such as gold, lead, zinc, and silver due to the desirability of the results of this study. These relationships can be used to study the geochemical situation of the region, which improves the decision-making ability of miners and geologists.

ORCID

Adel Shirazy https://orcid.org/0000-0001-7756-3205
Ardeshir Hezarkhani https://orcid.org/0000-0002-1149-3440
Aref Shirazi https://orcid.org/0000-0001-7623-301X
Shayan Khakmardan https://orcid.org/0000-0003-4359-9538
Reza Rooki https://orcid.org/0000-0002-4761-5518

REFERENCES

Abraham, A. (2005). Artificial neural networks. *Handbook of measuring system design*.

Alahgholi, S., Shirazy, A. & Shirazi, A. (2018). Geostatistical studies and anomalous elements detection, Bardaskan Area, Iran. *Open Journal of Geology*, *8*(7), 697-710.

Artun, E., Mohaghegh, S. D., Toro, J., Wilson, T. & Sanchez, A. (2005). *Reservoir characterization*

using intelligent seismic inversion. SPE Eastern Regional Meeting.

Cheung, Y.-M. (2003). k∗-Means: A new generalized k-means clustering algorithm. *Pattern Recognition Letters*, *24*(15), 2883-2893.

Dayhoff, J. E. & DeLeo, J. M. (2001). Artificial neural networks: opening the black box. *Cancer: Interdisciplinary International Journal of the American Cancer Society*, *91*(S8), 1615-1635.

Demuth, H. & Beale, M. (1993). *Neural Network Toolbox For Use with Matlab Users' Guide Version 3.0*.

Ghannadpour, S. S., Hezarkhani, A. & Farahbakhsh, E. (2013). An investigation of Pb geochemical behavior respect to those of Fe and Zn based on k-Means clustering method. *Journal of Tethys*, *1*(4), 291-302.

Ghorbani, M. (2013a). *Economic geology of Iran* (Vol. 581). Springer.

Ghorbani, M. (2013b). A summary of geology of Iran. In: *The economic geology of Iran* (pp. 45-64). Springer.

Hajnajafi, G., Jafarirad, A., Afzal, P. & Sheikh-Zakariaee, S.-J. (2021). Geological interpretation using multivariate K-means and robust factor analysis in Dezak area, SW Iran. *Environmental Earth Sciences*, *80*(1), 1-13.

Hamerly, G. & Elkan, C. (2003). Learning the k in k-means. *Advances in neural information processing systems*, *16*, 281-288.

Heil, J., Häring, V., Marschner, B. & Stumpe, B. (2019). Advantages of fuzzy k-means over k-means clustering in the classification of diffuse reflectance soil spectra: A case study with West African soils. *Geoderma*, *337*, 11-21.

Hezarkhani, A. & Ghannadpour, S. S. (2015). *Geochemical behavior investigation based on K-means clustering: basics, concepts and case study*. LAP (Lambert Academic Publishing).

Khosravi, V., Shirazi, A., Shirazy, A., Hezarkhani, A., & Pour, A. B. (2022). Hybrid Fuzzy-Analytic Hierarchy Process (AHP) Model for Porphyry Copper Prospecting in Simorgh Area, Eastern Lut Block of Iran. *Mining, 2*(1), 1-12.

Khayer, K., Shirazy, A., Shirazi, A., Ansari, A., Nazerian, H. & Hezarkhani, A. (2021). Determination of Archie's Tortuosity Factor from Stoneley Waves in Carbonate Reservoirs. *International Journal of Science and Engineering Applications (IJSEA)*, *10*, 107-110.

Krishna, K. & Murty, M. N. (1999). Genetic K-means algorithm. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, *29*(3), 433-439.

Menard, S. (1995). An introduction to logistic regression diagnostics. *Applied logistic regression analysis*, 58-79.

Moritz, R. (2016). The economic geology of Iran: mineral deposits and natural resources (M. Ghorbani). In: Society of Economic Geologists.

Murthy, C. A., & Chowdhury, N. (1996). In search of optimal clusters using genetic algorithms. *Pattern Recognition Letters*, *17*(8), 825-832.

Pelleg, D. & Moore, A. W. (2000). X-*means: Extending k-means with efficient estimation of the number of clusters*. Icml, Proceedings of the Seventeenth International Conference on Machine Learning.

Schalkoff, R. J. (1997). *Artificial neural networks*. McGraw-Hill Higher Education.

Shirazi, A., Hezarkhani, A., Shirazy, A. & Shahrood, I. (2018a). Exploration Geochemistry Data-Application for Cu Anomaly Separation Based On Classical and Modern Statistical Methods in South Khorasan, Iran. *International Journal of Science and Engineering Applications*, *7*, 39-44.

Shirazi, A., Hezarkhani, A., Shirazy, A. & Shahrood, I. (2018b). Remote sensing studies for mapping of iron oxide regions, South of Kerman, Iran. *International Journal of Science and Engineering Applications*, *7*(4), 45-51.

Shirazi, A., Shirazy, A. & Karami, J. (2018c). Remote sensing to identify copper alterations and promising regions, Sarbishe, South Khorasan, Iran. *International Journal of Geology and Earth Sciences*, *4*(2), 36-52.

Shirazi, A., Shirazy, A., Saki, S. & Hezarkhani, A. (2018d). Geostatistics studies and geochemical modeling based on core data, sheytoor iron deposit, Iran. *Journal of Geological Resource and Engineering*, *6*, 124-133.

Shirazi, A., Shirazy, A., Saki, S. & Hezarkhani, A. (2018e). Introducing a software for innovative neuro-fuzzy clustering method named NFCMR. *Global Journal of Computer Sciences: theory and research*, *8*(2), 62-69.

Shirazy, A., Shirazi, A., Heidarlaki, S. & Ziaii, M. (2018a). Exploratory Remote Sensing Studies to Determine the Mineralization Zones around the Zarshuran Gold Mine. *International Journal of Science and Engineering Applications*, *7*(9), 274-279.

Shirazy, A., Shirazi, A. & Hezarkhani, A. (2018b). Predicting gold grade in Tarq 1: 100,000 geochemical map using the behavior of gold, Arsenic and Antimony by K-means method. *Journal of Mineral Resources Engineering*, *2*(4), 11-23.

Shirazy, A., Shirazi, A., Ferdossi, M. H. & Ziaii, M. (2019). Geochemical and geostatistical studies for estimating gold grade in tarq prospect area by k-means clustering method. *Open Journal of Geology*, *9*(6), 306-326.

Shirazy, A., Ziaii, M. & Hezarkhani, A. (2020a). Geochemical Behavior Investigation Based on K-means and Artificial Neural Network Prediction for Copper, in Kivi region, Ardabil province, Iran. *Iranian Journal of Mining Engineering*, *14*(45), 96-112.

Shirazy, A., Ziaii, M., Hezarkhani, A. & Timkin, T. (2020b). Geostatistical and remote sensing studies to identify high metallogenic potential regions in the Kivi area of Iran. *Minerals*, *10*(10), 869.

Shirazy, A., Ziaii, M., & Hezarkhani, A. (2021a). Geochemical behavior investigation based on k-means and artificial neural network prediction for titanium and zinc, Kivi region, Iran. *Bulletin of the Tomsk Polytechnic University. Geo Assets Engineering*, *332*(3), 113-125.

Shirazy, A., Hezarkhani, A., Timkin, T. & Shirazi, A. (2021b). Investigation of Magneto-/Radio-Metric Behavior in Order to Identify an Estimator Model Using K-Means Clustering and Artificial Neural Network (ANN)(Iron Ore Deposit, Yazd, IRAN). *Minerals*, *11*(12), 1304.

Shirazy, A., Shirazi, A. & Nazerian, H. (2021c). Application of Remote Sensing in Earth Sciences–A Review. *International Journal of Science and Engineering Applications*, *10*, 45-51.

Shirazy, A., Shirazi, A., Nazerian, H. & Hezarkhani, A. (2021d). Investigation of Geochemical Sections in Exploratory Boreholes of Mesgaran Copper Deposit in Iran. *International Journal for Research in Applied Science and Engineering Technology (IJRASET)*, *9*(8), 2364-2368.

Shirazy, A., Shirazi, A., Nazerian, H. & Khayer, K. (2021e). Geophysical study: Estimation of deposit depth using gravimetric data and Euler method (Jalalabad iron mine, kerman province of IRAN). Open Journal of Geology, 11, 340-355.

Specht, D. F. (1991). A general regression neural network. *IEEE transactions on neural networks*, *2*(6), 568-576.

Tarkian, M. & Stribrny, B. (1999). Platinum-group elements in porphyry copper deposits: a reconnaissance study. *Mineralogy and Petrology*, *65*(3-4), 161-183.

Ullman, J. (1984). A review of: "Pattern Recognition: A Statistical Approach". By PA Devuver and J. Kittler. (London: Prentice Hall International, 1982.) [pp. 448.]. *Remote Sensing*, *5*(2), 464-464.

Xu, Z. (2012). Priority weight intervals derived from intuitionistic multiplicative preference relations. *IEEE Transactions on Fuzzy Systems*, *21*(4), 642-654.

Yaghini, M., & Gereilinia, N. (2013). Genetic TKM: A hybrid clustering method based on genetic algorithm, tabu search and k-means. *International Journal of Applied Metaheuristic Computing (IJAMC)*, *4*(1), 67-77.

Yegnanarayana, B. (2009). *Artificial neural networks*. PHI Learning Pvt. Ltd.