Neutrosophic K-means for market segmentation

A. Romero Fernández[1], G. Alvarez Gómez[2], C. Gómez Armijos[3]

[1]Director de Investigación de la Universidad Regional Autónoma de los Andes (UNIANDES), Ecuador

[2]Rector de la Universidad Regional Autónoma de los Andes (UNIANDES), Ecuador

[3]Vicerrectora General de la Universidad Regional Autónoma de los Andes (UNIANDES), Ecuador

Email: dir.investigacion@uniandes.edu.ec; rectorado@uniandes.edu.ec; vicerrectorado@uniandes.edu.ec

Abstract

Markets may be broken down into subsets with the use of cluster analysis. Multivariate analytic methods are often used in traditional research. Due to their success in engineering, artificial neural systems have recently found use in business as well. When it comes to grouping observations with comparable traits or attributes, the K-means method is a common choice. It has various uses in marketing, but it finds particular success in cluster analyses of customer behavior. Several commercial packages include implementations of the K-means algorithm. Data mining statistical approaches like K-Means are useful for handling this data and analyzing it later on. For better results, this study combines the traditional K-Means technique with Neutrosophy, which accounts for the uncertainty inherent in such complicated data sets by factoring in the data's diversity and its inherent volatility as a result of proximity between the bounds of the separate segments as well as the members who make up each.

Keywords: Neutrosophic; K-Means; Clustering; Market Segmentation;

1. Introduction

Since Wendal Smith originally introduced the notion of market segmentation in 1956 it has evolved into one of the cornerstones of marketing. A company may improve its profitability even in the face of diverse marketplaces by adopting a strategy of market segmentation. Several market segmentation strategies have been presented as potential solutions to this issue for decades. However, the distance between academically focused research on categorization and the cases involved must be reduced if market segmentation research is to improve[1]–[4].

Segmentation's overarching goal is to classify data into clusters of comparable elements. People, plant species, components, and signals are all examples of such things. Multivariate analysis is one such technique that is widely used. Artificial neural networks (ANNs), which are distributed and simultaneous data processing systems, have lately been used to tackle marketing difficulties as a result of improvements in both computing power and affordability[5]–[7].

Therefore, the purpose of this research is to have a conversation about the prospect of combining ANN with multivariate analysis. Punj and Steward propose a two-step process. In this approach, a hierarchical technique, such as Ward's minimal variance technique, is supplemented by a non-hierarchical technique, like the K-means technique. Self-organizing feature maps are used in this work to establish the initial clustering and number of clusters, and the K-means algorithm is then

Doi: https://doi.org/10.54216/IJNS.190123
Received: March 23, 2022 Accepted: September 12, 2022

used to discover the optimal solution. Numerical simulations and data from real-world problems are utilized to test the effectiveness of the suggested approach. The suggested two-stage approach improves upon the standard two-stage approach (Ward's minimal variance technique accompanied by the K-means method), as shown by the simulation results. The suggested technique is superior based on Wilk's Lambda and analysis of variance, as shown by both theoretical and practical results. Moreover, when compared to the modified and traditional two-stage techniques, the outcomes of applying ego image features alone are determined to be the poorest[8]–[11].

In the 21st century, commercialization will be a thing of the past. Changes have been made in marketing strategy, with a shift away from broad campaigns and toward more specific ones. Due to the high volume of requests from various clients, it is hard to meet everyone's expectations. The company must segment the market and choose the most lucrative segment to increase customer satisfaction. Wendel Smith is credited with developing the concept of market segmentation. It is the initial step in the multi-step process that constitutes targeted marketing and begins with "positioning."[8]–[15]

As Kotler explained, there are three steps—the survey, the analysis, and the profiling—involved in ensuring the various sectors of a market. Similar procedures were explored by Wind. According to Kolter, the following five qualities are necessary for efficient market segmentation: 1. quantifiability; 2. significance; 3. availability; 4. distinguishability; 5. actionability. A successful clustering result should meet the following six requirements, albeit there are many alternative clustering algorithms available. There must be sufficient resemblance among each group while also being (1) homogenous within, (2) heterogeneous among, (3) significant, (4) operational, (5) available, and (6) reliable[16]–[18].

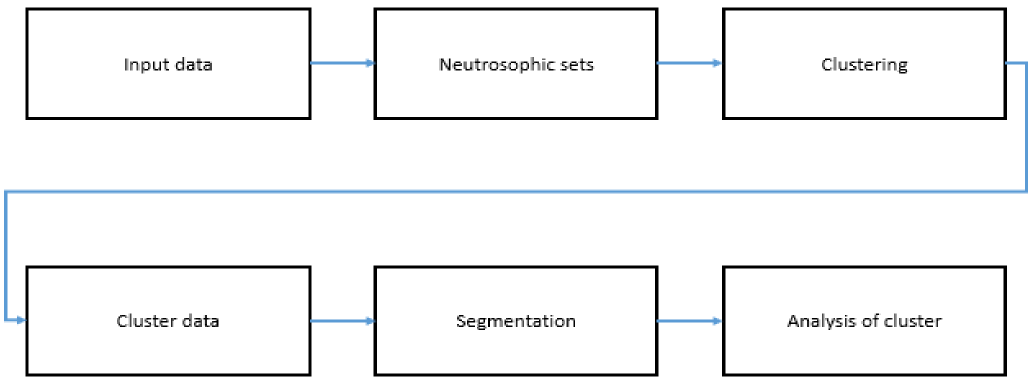

Figure 1: The proposed model.

2. K-Means Clustering

We employ the Neutrosophic variant of the K-Means mathematical formula for our research since it accounts for the uncertainties and uncertainty context that are built into our forecasts. Decision-making based on the perception of the linguitic concepts offered by Neutrosophy is efficient, therefore using a Neutrosophic K-Means that uses the classical method is suitable[19]–[21], [21]. Figure 1 shows the proposed model.

According to the literature, the K-Means approach is widely used in Data Mining for the simple reason that it may be utilized to efficiently handle and categorize massive datasets using clustering or grouping. The goal of k-means clustering is to group items that are comparable or have common

Doi: https://doi.org/10.54216/IJNS.190123
Received: March 23, 2022 Accepted: September 12, 2022

properties. To classify 'n' observations into 'k's clusters, where k is a predetermined or user-defined constant, K-means clustering is an unstructured machine learning approach. The key concept is to identify k cluster centroids. The steps of the K-Means algorithm are as follows:
settling on a value for k, the cluster size, Second, cluster the data based on a randomization process, And so on, until the evolution of clusters ceases: The "centroid" of a cluster is found by averaging all the points inside it. Each observation is ascribed to the cluster whose centroid it most closely resembles[22]–[25].
Scaling the factors before grouping the data is crucial in K-means, and using a scatter diagram or a presentation of the information to figure out how many cluster centers to set again for the k variable in the model is also recommended[26]–[28].

Professor Florentin Smarandache coined the term "neutrosophy" to describe his study of the history, definition, and use of neutralities. Its inclusion ensures that the inherent ambiguity in government decisions is taken into account including the unknowns surrounding the criteria by which experts would evaluate language and nonnumerical words. Judgment, segmentation techniques, and machine learning are just a few of the many domains where logic and neutrosophic sets, an extension of Zadeh's fuzzy logic and needs to set, have proven useful[29]–[32].

References indicate that Neutrosophic K-Means is a neutrosophic data gathering approach for clustering and that it is an extension of the standard K-Means. To automate the manipulation of databases given by the Ecuador seismic network, this variation is helpful. The method enables the discovery of regular behavior patterns or correlations between multiple variables, the organization and segmentation of massive data sets, and their subsequent application to new sets of information[33]–[37].
Since it is challenging to determine the boundaries that separate them from the groups they correspond to, this analysis takes the variety and variability of the data into account, which may lead to erroneous results and the presence of inconsistencies[38]–[41]. Taking into account the information, the procedure entails giving each piece of data a score representing its level of involvement in each cluster (therefore allowing it to partly belong to much more than a group) and displaying the results graphically. The foregoing are revealed to be:

Definition 1:

Suppose that X is the dataset and x_i is a component such that $x_i \in X$

Definition 2:

$$\begin{gathered}A\ partition\ P\ =\ C1, C2, \dots, Cc\ of\ the\ data\ set\\ X\ is\ said\ to\ be\ soft\ if\ and\ only\ if\\ the\ following\ is\ true:\\ (xi\ X, Cj\ P)Cj(xi)1\ y\ (xi\ X, Cj\ P) such\ that\ Cj\ (xi\ ) >\ 0.\\ Specifically, xi's\ degree\ of\ membership\ in\ cluster\ Cj\\ is\ denoted\ by\ the\ expression\ Cc(xi)xi.\end{gathered}$$

Definition 3:

$$\begin{gathered}It\ is\ argued\ that\ a\ certain\ soft\ partition\ exists\\ when\ the\ total\ of\ a\ point's\ degrees\ of\ membership\ in\\ all\ clusters\ are\ 1\\ \sum_j uc(x_i) = 1\end{gathered}$$

Definition 4:

The partitions that allow for this extra criterion are called constrained soft partitions. By extending the goal function J in two ways, the Neutrosophic K-Means method can generate a restricted smooth partition.

$$\begin{gathered}\forall xi\ \in\ X, \exists Cj\ \in\ P\ such\ that\ \mu Cj\ (xi\ ) >\ 0\\ where\ the\ degrees\ of\ neutrosophic\ membership\\ of\ each\ data\ in\ each\ cluster\ are\ incorporated\ or;\end{gathered}$$

Doi: https://doi.org/10.54216/IJNS.190123
Received: March 23, 2022 Accepted: September 12, 2022

Definition 5:

The membership function is expanded by adding a new argument that acts as an exponent weight. Jm

$$uc_1(X_1) = \frac{1}{\sum_{j=1}^{2}\left[\frac{x_i - v1^2}{x_i - v\_j^{\ 2}}\right]^2}$$

In this case, P is a fuzzy partitioning of X produced by C1, C2,..., Ck, and m is a weight that establishes how much the incomplete members of a cluster influence the outcome.
Referring to the fact that both the classical approach and its neutrosophic outgrowth aim to find a separation by having to search for rapid prototyping vi such that they minimize the optimal solution Jm, and that the latter must similarly search for membership functions C1that minimize Jm, this statement emphasizes the similarities between the two approaches.
The approach is supplemented by the establishment of equation 3 to compute the baseline membership function parameters of each cluster:

$$J_m(p,v) = \sum_{j=1}^{k}\sum_{j_{k\in x}}(uc_j(x_k))^m \left\|x_k - v_j\right\|^2$$

$$v_1 = \frac{\sum_{k=1}^{n}(uc_j(x_k))^2 x_k}{\sum_{k=1}^{n}(uc_j(x_k))^2}$$

The KNCM algorithm, based on the aforementioned equations, may be formulated as follows.

I. First, set the values for T(0), I(0), and F(0) to zero;
II. next, set the values for C(0), m(0), (1), (2), and (3) to zero;
III. finally, choose the kernel function and its parameters in step (3);
IV. Step 4: Determine the vectors c(k) that represent the centers of each cluster at each k-step;
V. Step 5: Determine the maximum value of ci by summing the coordinates of the centers of the clusters whose greatest and second-largest values of Tij satisfy Eq. (3).

VI. Sixth, to change T (k) to T (k+1), to change I (k) to I (k+1), and Eq. (17) to change F (k) to F (k+1);
VII. Seventh, if |T (k+1) T (k)| , end; otherwise, go back to Step 4; Eighth, if |T (k+1) T (k)| >, continue;
VIII. Eighth, place each piece of information into the category that has the highest value of TM = [T, I, F]:
IX. If k = arg max (TMij), where j = 1, 2,, C + 2, then x I is in the kth class.

3. Results

The participants in this survey are those who have shopped at 3C shops in Kaohsiung City. This is a lie; the shops sell nothing but 3C goods. There are seven 3C establishments included in the city's annual report.

Data for this investigation was gathered mostly via in-person interviews, using a quota sampling technique that may be described as follows. There are a total of seven 3C shops in Kaohsiung. There are 35 copies of the survey for each location. On the weekends, we do the interviews. In all, 245 surveys are taken. All questions are conducted through in-person interviews, so there is no room for error. However, five copies were disregarded because their responses were not full. Therefore, only 240 surveys will be considered. A 98 percent rate of return is achieved.

Doi: https://doi.org/10.54216/IJNS.190123
Received: March 23, 2022 Accepted: September 12, 2022

Data are entered into SPSS for analysis once the usual "verify, check, remove, and coding" processes have been completed. According to the data collected, the majority of responders (65%) are male. 34.2% of respondents are between the ages of 21 and 30, and 92.5% had completed some level of postsecondary education beyond high school. There is an equitable distribution of living spaces, and 32.5 percent of the average household income is between NT$250,000 and NT$35,000. Of the people who filled out the survey, 29.2% work in business and 22.5% are students.

Neutrosophic K-means is the result of applying neutrosophic approaches to the traditional Kmeans clustering algorithm.
The approaches were implemented using Orange (Version 3.27.1), a data mining program with various advantages since it allows trying out alternative processes and applying many different widgets until the desired result is achieved. This is when neutrosophic K-means come in handy.
To visualize the Silhouette numbers for each cluster, the dataset is imported using a File widget, analyzed with the k-Means widget, Neutrosophy is applied with a Python script, and the results are presented via a Scatter Plot and a Silhouette Plot. A Data Structure was also included at the process's conclusion to facilitate further analysis by filtering, counting, sorting, etc.
The k-Means widget in Orange was used to figure out the number of groups (k) since this widget allows for iteratively adjusting the number of clusters to discover the optimal clustering settings given the Silhouette Scoring values. The option with the greatest Silhouette score, k = 2 clusters in this example (with a score of 0.822), is the best choice.
In order to incorporate the Neutrosophic component, a Python script was written to compute the initial membership values of both clusters using the formula, and then iteratively refine those computations using the formula until the expanded objective function was minimized.

The numbers of segments are 8, 117, 33, and 82 when using only the K-means approach and a segmentation level of 4. That the outliers affect the outcomes. There is significant inequality in the group's total size. Due to the Wilk's Lambda value of 0.16943 being greater than that of the suggested two-stage technique, it seems that the K-means technique alone with an initial random point may yield the poorest result.

Figure 2 represents the examined data set. As shown, the Shape value was computed for two distinct sets of data, yielding two distinct clusters, one having 800 rows while the other had 200 rows. Figure 2 shows that there is a substantial quantity of data with high Shape score values in cluster 3, the blue cluster, indicating that these observations have a strong affiliation with the cluster where they are placed. Figure 3 shows the clustering of data into three groups. Table 1 shows the data of clustering.

Doi: https://doi.org/10.54216/IJNS.190123
Received: March 23, 2022 Accepted: September 12, 2022

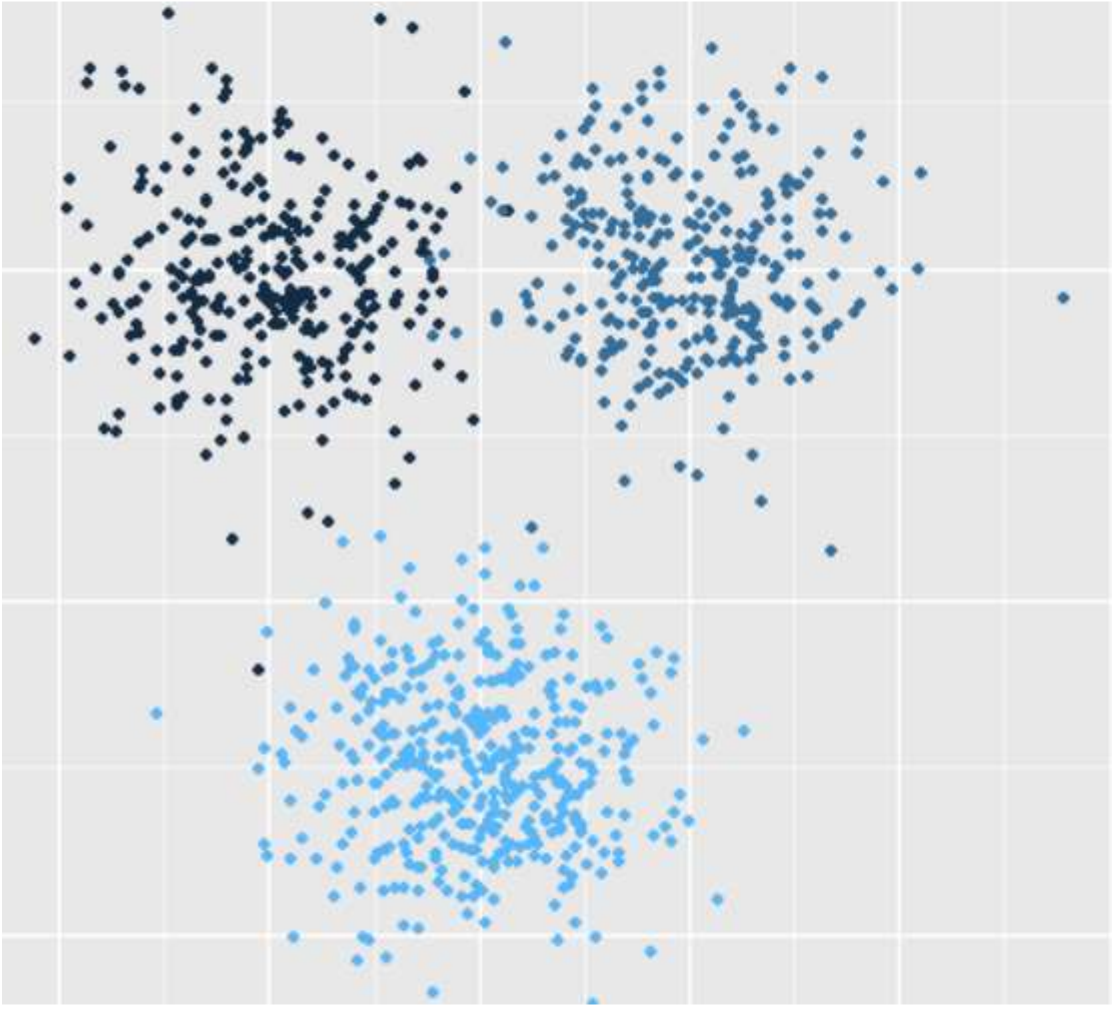

Figure 2: K=3.

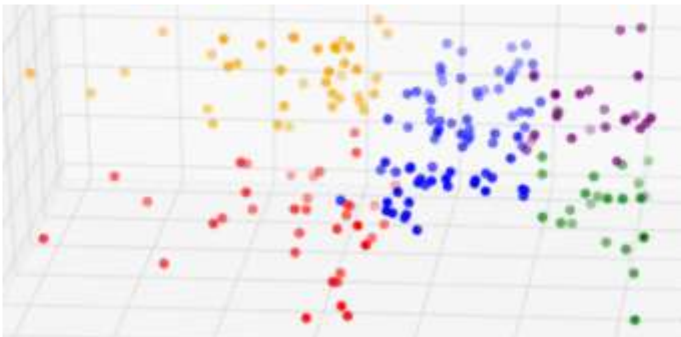

Figure 3: K=2.

Table 1: The neutrosophic k-Means clustering outcomes.

	Class 1	Class 2	Class 3
Class 1	200	180	0
Class 2	100	20	220
Class 3	0	60	30

Doi: https://doi.org/10.54216/IJNS.190123
Received: March 23, 2022 Accepted: September 12, 2022

6. Conclusion

In this study, we performed a data mining assessment on a dataset containing information about market segmentation using the K-means neutrosophic clustering algorithm. Our goal was to identify patterns in the data of the sets of clusters obtained so that we could use them to predict the behavior of these market segmentation in the future.

In conclusion, the results demonstrated the superiority of Neutrosophic K-Means as a Neutrosophic extension of the classical method, as it is better able to handle data that partially belong to more than one cluster, and it calculates both the prototypes of the cluster and the membership functions of the data within each cluster in addition to the prototypes themselves. When the data can be broken down into distinct categories, Neutrosophic K-Means may help by creating a partition that preserves these properties while still being somewhat smooth.

Data categorization, medical, bioinformatics, economics, and many more may all benefit from this strategy.

References

[1] W. R. Smith, "Product differentiation and market segmentation as alternative marketing strategies," *Journal of Marketing*, vol. 21, no. 1, pp. 3–8, 1956.

[2] P. Arabie, "Cluster analysis in marketing research," *Advanced methods of marketing research*, pp. 160–189, 1994.

[3] M. Wedel and W. A. Kamakura, *Market segmentation: Conceptual and methodological foundations*. Springer Science & Business Media, 2000.

[4] P. Kotler and G. Armstrong, *Marketing management, analysis, planning, implementation, and control, Philip Kotler*. London: Prentice-Hall International, 1994.

[5] R. Grover and V. Srinivasan, "A simultaneous approach to market segmentation and market structuring," *Journal of Marketing Research*, vol. 24, no. 2, pp. 139–153, 1987.

[6] Y. J. Wind and D. R. Bell, "Market segmentation," in *The marketing book*, Routledge, 2008, pp. 260–282.

[7] L. Quinn, "Market segmentation in managerial practice: a qualitative examination," *Journal of Marketing Management*, vol. 25, no. 3–4, pp. 253–272, 2009.

[8] L. Quinn and S. Dibb, "Evaluating market-segmentation research priorities: Targeting re-emancipation," *Journal of Marketing Management*, vol. 26, no. 13–14, pp. 1239–1255, 2010.

[9] Y. J. Wind, "Advances in Marketing Research and Practice," *Marketing Research and Modeling: Progress and Prospects: A Tribute to Paul E. Green*, vol. 14, p. 301, 2013.

[10] Sireesha Rodda , Vaibhav Kovela , Sanjay Dokula, Instance Segmentation and Labeling of Teeth from Dental X-Ray using Region Based Convolutional Neural Network, Journal of Neutrosophic and Fuzzy Systems, Vol. 2 , No. 2 , (2022) : 20-30

[11] J. G. Barnes and R. McTavish, "Segmenting industrial markets by buyer sophistication," *European Journal of Marketing*, 1983.

[12] D. C. Marschner, "Basic marketing: a managerial approach," *Journal of Marketing (pre-1986)*, vol. 36, no. 000001, p. 106, 1972.

[13] E. J. McCarthy and W. D. Perreault, *Basic marketing: a managerial approach*. Irwin, 1972.

[14] E. J. McCarthy, S. J. Shapiro, and W. D. Perreault, *Basic marketing*. Irwin-Dorsey Toronto, ON, Canada, 1979.

[15] R. D. Hisrich and V. Ramadani, *Entrepreneurial marketing: a practical managerial approach*. Edward Elgar Publishing, 2018.

[16] D. Krimpmann, "Development of a Managerial Approach for a New IT Organisation Design Framework (ITODF) Based on Digitisation Trends.," 2018.

[17] A. Androniceanu, O. M. Sabie, and A. Pegulescu, "An integrated approach of the human resources motivation and the quality of health services," *Theoretical and Empirical Researches in Urban Management*, vol. 15, no. 1, pp. 42–53, 2020.

[18] A. Alrida Basheer , Katy D. Ahmad , Rozina Ali, On Some Open Problems About n-Cyclic Refined Neutrosophic Rings And Number Theory, Journal of Neutrosophic and Fuzzy Systems, Vol. 3 , No. 2 , (2022) : 37-42

[19] M. N. Qureshi and M. V. Ahamad, "An improved method for image segmentation using K-means clustering with neutrosophic logic," *Procedia computer science*, vol. 132, pp. 534–540, 2018.

[20] B. B. Yousif, M. M. Ata, N. Fawzy, and M. Obaya, "Toward an optimized neutrosophic K-means with genetic algorithm for automatic vehicle license plate recognition (ONKM-AVLPR)," *IEEE Access*, vol. 8, pp. 49285–49312, 2020.

Doi: https://doi.org/10.54216/IJNS.190123

Received: March 23, 2022 Accepted: September 12, 2022

[21] N. V. Q. Arnaiz, N. G. Arias, and L. C. C. Muñoz, *Neutrosophic K-means Based Method for Handling Unlabeled Data*, vol. 37. Infinite Study, 2020.

[22] E. B. N. Arias, B. M. G. Nuñez, L. N. Fernández, and J. M. R. Pupo, "CRISP-DM y K-means neutrosofía en el análisis de factores de riesgo de pérdida de audición en niños.," *Revista Asociación Latinoamericana de Ciencias Neutrosóficas. ISSN 2574-1101*, vol. 16, pp. 73–81, 2021.

[23] A. Sengur and Y. Guo, "Color texture image segmentation based on neutrosophic set and wavelet transformation," *Computer Vision and Image Understanding*, vol. 115, no. 8, pp. 1134–1144, 2011.

[24] E. Karabatak, A. Sengur, and Y. Guo, "Modified neutrosophic approach to color image segmentation," *Journal of Electronic Imaging*, vol. 22, no. 1, p. 13005, 2013.

[25] Y. Guo and A. Sengur, "NCM: Neutrosophic c-means clustering algorithm," *Pattern Recognition*, vol. 48, no. 8, pp. 2710–2724, 2015.

[26] Y. Yuan, Y. Ren, X. Liu, and J. Wang, *Approach to image segmentation based on interval neutrosophic set*. Infinite Study, 2020.

[27] Ahmad Salama , Malath Al Aswad , Rasha Dalla , Rozina Ali, A Study of Neutrosophic Differential Equations By Using the One-Dimensional Geometric AH-Isometry Of NeutrosophicLaplace Transformation, Journal of Neutrosophic and Fuzzy Systems, Vol. 4 , No. 2 , (2022) : 08-25

[28] Y. Akbulut, A. Şengür, Y. Guo, and K. Polat, "KNCM: Kernel neutrosophic c-means clustering," *Applied Soft Computing*, vol. 52, pp. 714–724, 2017.

[29] A. R. Hawas, A. S. Ashour, and Y. Guo, "Neutrosophic set in medical image clustering," in *Neutrosophic Set in Medical Image Analysis*, Elsevier, 2019, pp. 167–187.

[30] J. Mohan, V. Krishnaveni, and Y. Huo, "Automated brain tumor segmentation on MR images based on neutrosophic set approach," in *2015 2nd international conference on electronics and communication systems (ICECS)*, 2015, pp. 1078–1083.

[31] A. S. Ashour, A. R. Hawas, Y. Guo, and M. A. Wahba, "A novel optimized neutrosophic k-means using genetic algorithm for skin lesion detection in dermoscopy images," *Signal, Image and Video Processing*, vol. 12, no. 7, pp. 1311–1318, 2018.

[32] A. A. Salama, M. Eisa, S. A. ELhafeez, and M. M. Lotfy, "Review of recommender systems algorithms utilized in social networks based e-Learning systems & neutrosophic system," *Neutrosophic Sets and Systems*, vol. 8, pp. 32–40, 2015.

[33] Y. Guo and H.-D. Cheng, "New neutrosophic approach to image segmentation," *Pattern Recognition*, vol. 42, no. 5, pp. 587–595, 2009.

[34] Y. Guo, R. Xia, A. Şengür, and K. Polat, "A novel image segmentation approach based on neutrosophic c-means clustering and indeterminacy filtering," *Neural Computing and Applications*, vol. 28, no. 10, pp. 3009–3019, 2017.

[35] Y. Guo and A. Şengür, "A novel image segmentation algorithm based on neutrosophic similarity clustering," *Applied Soft Computing*, vol. 25, pp. 391–398, 2014.

[36] H. D. Cheng, Y. Guo, and Y. Zhang, "A novel image segmentation approach based on neutrosophic set and improved fuzzy c-means algorithm," *New Mathematics and Natural Computation*, vol. 7, no. 01, pp. 155–171, 2011.

[37] V. WB, I. Kandasamy, F. Smarandache, V. Devvrat, and S. Ghildiyal, "Study of imaginative play in children using single-valued refined neutrosophic sets," *Symmetry*, vol. 12, no. 3, p. 402, 2020.

[38] A. Q. Ansari, P. Sharma, and M. Tripathi, "Automatic seizure detection using a neutrosophic classifier," *Physical and Engineering Sciences in Medicine*, vol. 43, no. 3, pp. 1019–1028, 2020.

[39] I. V. Pustokhina *et al.*, "Automatic vehicle license plate recognition using optimal K-means with convolutional neural network for intelligent transportation systems," *Ieee Access*, vol. 8, pp. 92907–92917, 2020.

[40] A. Elhassouny, S. Idbrahim, and F. Smarandache, *Machine learning in Neutrosophic Environment: A Survey*. Infinite Study, 2019.

[41] N. Akhtar, N. Agarwal, and A. Burjwal, "K-mean algorithm for image segmentation using neutrosophy," in *2014 International conference on advances in computing, communications, and informatics (ICACCI)*, 2014, pp. 2417–2421.

Doi: https://doi.org/10.54216/IJNS.190123
Received: March 23, 2022 Accepted: September 12, 2022