مجلة البحوث المالية والتجارية
المجلد (26) – العدد الثاني – إبريل 2025

Enhancing Accuracy of Neutrosophic –C means and Fuzzy C means with Application on insurance companies retention limit

Neveen Mohsen Zaki

Department of applied statistics and insurance, Faculty of commerce, Mansoura university, Mansoura, Egypt, neveen_mohsen@mans.edu.eg

Alaa A. Abdelmegaly

Higher institute of advanced management sciences and computers, Al-Buhayrah, Egypt, bintmasr880@yahoo.com

Doaa A. Abdo

Applied statistics and insurance department, Faculty of commerce, Mansoura university, Mansoura, Egypt, doaaashour@mans.edu.eg

تاريخ الإرسال	2024-12-16
تاريخ القبول	2025-02-16
رابط المجلة:	https://jsst.journals.ekb.eg/

Abstract:

This study proposed two Methods of clustering including neutrosophic C- means and Fuzzy C-means with application on two insurance companies, contained, misr insurance company and suez canal insurance company. This paper aims to handle the drawbacks of FCM method of minimizing the intra-cluster variance and the sensitive noise which resulted in un accuracy findings by incorporating neutrosophic theory into clustering means. The neutrosophic C- means proposed as an alternative technique for FCM to handle the problems of fuzzy C- means. The both introduced approaches were applied on two insurance companies using two variables included retention limit and reinsurance commission rate. The findings of this study showed that the neutrosophic C- means get results more accuracy than the FCM according to neutrosophic accuracy measures included falsity, truth and indeterminacy. As using Neutrosophic C- means increasing accuracy from 65.4% according to FCM to 77%.

Key words: FCM, Neutrosophic C- Means; Falsity; Truth; Indeterminacy; Retention limit, Reinsurance commission rate.

1. Introduction:

On computer version, the traditional FCM algorithm is developed on the basis of fuzzy theory(Hong et al., 2016). However, the fuzzy theory has certain limitations (Miyamoto etal., 2008). The lack of expressive ability of uncertain information makes the FCM unable to handle clustered boundary pixels and outliers. Although, the fuzzy C means algorithm one of the most popular fuzzy clustering where the degrees of membership of the data are assessed via iterative minimization of a cost function subject to the constrain that the sum of the membership degrees over the clusters for each data be equal to 1.

The FCM have several drawbacks, it attempts minimizing intra-cluster variance, and has the same problems as K- means algorithm; the minimum is a local minimum, and the results depend greatly on the initializations (Menard et al., 2000). In addition to that, the FCM algorithm is very sensitive to the presence of noise. The membership of noise points might significantly high. This algorithm cannot highly likely and equally highly. Unlikely, and is sensitive to the selection of distance metric. To solve these problems of FCM, a new clustering algorithm, neutrosophic C means (NCM), is proposed for uncertain data clustering which is derived from fuzzy C- means and the neutrosophic set framework. Smarandache(1998) provided the neutrosophic set theory.

(Guo and senger ,2015) improved FCM on the basis of the neutrosophic theory, and introduced the neutrosophic C- means clustering algorithm (NCM). The algorithm includes the degree of membership in addition to the uncertainty and opposition. (Castellano et al., 2024) incorporates neutrosophic theory to manage the uncertainties in sensory evaluation and applies neutrosophic C- Means clustering to analyze the data obtained from both expert assessments and instrumental measurements. (Qiu etal., 2024) introduced interval – valued data to formulate a novel objective function and provide iterative procedures for updating cluster prototype and neutrosophic partition.(Abdelhafeez etal., 2023) analyzed pointedly the breast cancer data set cluster ability through applying the widely neutrosophic cluster C-Means, they conducted a superiority of

neutrosophic C-Means in segregation similar breast cancer instances into clusters. (Abdelhafeez and Mohamed, 2023) analyzed a comparative study between neutrosophic C-Means and Fuzzy C-Means and the superiority in segregation similar breast instances into clusters

The rest of this paper is designed as follows: in section2, we proposed Fuzzy C-Means. In section 3we introduced Neutrosophic C- Means. Finally, in section 4, we presented numerical analysis of proposed methods under two insurance companies.

2. Fuzzy C-Means:

The foregoing fuzzy system allows us to convert and embed empirical qualitative knowledge into reasoning systems capable of performing approximate pattern matching and interpolation. However, these systems cannot adapt or learn because they are unable to extract knowledge from existing data. One approach for overcoming this limitation is to use a fuzzy clustering method such as the fuzzy c-means algorithm (Shapiro, 2005).

Fuzzy C-Means (FCM) is a popular clustering technique that extends the traditional hard clustering methods, such as K-Means, by allowing data points to belong to multiple clusters simultaneously. This flexibility is achieved by assigning membership degrees to each data point in each cluster, rather than assigning them to a single, definitive cluster(Ikotun etal., 2023).

Dunn and Bezdek proposed the fuzzy C-means algorithm (FCM). The main concept is to identify the proper membership and clustering center in order to determine the objective function's minimal value(Lu etal., 2019):

$$J_m(U,V) = \sum_{i=1}^{n} \sum_{j=1}^{c} \mu\mu_{ij}^{m} d_{ij}^{2}(x_i, v_j) \qquad (1)$$

$$\mu_{ij} = \left(\sum_{r=1}^{c} \left[\frac{d_{ij}}{d_{it}} \right]^{2-1} \right)^{-1} \qquad (2)$$

$$v_j = \frac{\sum_{i=1}^{n} \mu_{im}^{m} x_i}{\sum_{i=1}^{n} \mu_{ij}^{n}} \qquad (3)$$

where $J(U,V)$ denotes the square sum of the weighted distance from the pixel to the cluster center of the region, $U = (\mu_{iv})_{\text{rxc}}$ denotes the degree of membership matrix, C is the number of clusters of the image, μ_{ij} is the value of the sample point x_i belonging to j^{th} Class. m represents a fuzzy exponent, at a typical value of 2. When m = 1, the fuzzy clustering degenerates to hard clustering (HCM), $V = (v_1, v_2, \dots, v_c)$ is the matrix of clustering center values, v_j Represents j^{th} Cluster Center, and $d_{ij}^{2}(x_i, v_j) = \left\| x_i - v_j \right\|$ represents the Euclidean distance between sample point x_i and the cluster center v_j(Lu etal., 2019).

The algorithm first determines the number of clustering and initialize the membership matrix. Then the Clustering Center and Membership matrix are updated repeatedly through formula (2) and (3). When the objective function is less than a certain threshold, all kinds of clustering centers and membership degrees are obtained(Lu etal., 2019).

The essence of the c-means algorithm is that it produces reasonable centers for clusters of data, in the sense that the centers capture the essential feature of the cluster, and then groups data vectors around cluster centers that are reasonably close to them. A flowchart of the c-means algorithm is depicted in Figure NO.().

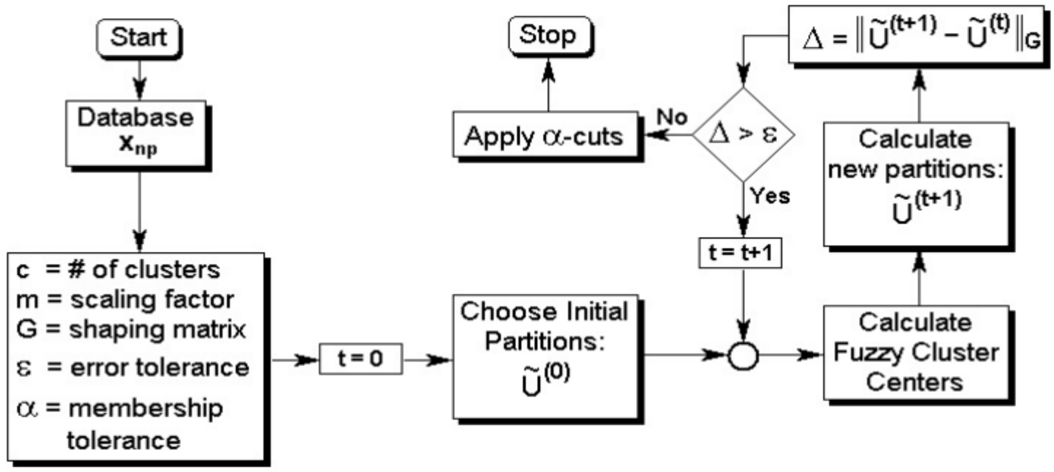

Figure1: A flowchart of the c-means algorithm

Source: (Shapiro, 2005).

As indicated, the database consists of the n x p matrix, Xup, where n indicates the number of patterns and p denotes the number of features. The algorithm seeks to segregate these n patterns into c, 2 c ≤ n - 1, clusters, where the within clusters variances are minimized and the between clusters variances are maximized. To this end, the algorithm is initialized by resetting the counter, t, to zero, and choosing: c, the number of clusters; m, the exponential weight, which acts to reduce the influence of noise in the data because it limits the influence of small values of membership functions; G, a symmetric, positive-definite (all its principal minors have strictly positive determinants), p x p shaping matrix, which represents the relative importance of the elements of the data set and the correlation between them, examples of which are the identity and covariance matrixes; s, the tolerance, which controls the stopping rule; and a, the membership tolerance, which defines the relevant portion of the membership functions(Shapiro, 2005).

Given the database and the initialized values, the counter, t, is set to zero. The next step is to choose the initial partition (membership matrix), which may be based on a best guess or experience. Next, the fuzzy cluster centers are computed, which, in effect, are elements that capture the essential feature of the cluster. Using these fuzzy cluster centers, a new (updated) partition, U+1), is calculated. The partitions

are compared using the matrix norm G (+)- and if the difference exceeds s, the counter, t, is increased and the process continues. If the difference does not exceed s, the process stops. As part of this final step, a- cuts are applied to clarify the results and make interpretation easier, that is, all membership function values less than a are set to zero and the function is renormalized(Shapiro, 2005).The Fuzzy C-Means Works as follow(Abdullah etal., 2024; Al-Janabee & Al-Sarray, 2022):

1. **Initialization:**
   - **Number of Clusters (C): The user specifies the desired number of clusters.**
   - **Membership Matrix (U): Initially, membership values for each data point in each cluster are assigned randomly.**
   - **Cluster Centers (V): Initial cluster centers are chosen randomly.**
2. **Membership Update:**
   - **Calculate the distance between each data point and each cluster center using a distance metric (e.g., Euclidean distance).**

3. Neutrosophic C-means

3.1 Neutrosophic theory

To handle the problems of the classical fuzzy theory and improve its capability of processing and expressing uncertain information, the neutrosophic theory which is a generalization of other extended theories (Smarandache, 1998). The neutrosophic theory works out the un solved problems in addition to that represent non- deterministic issues when applying the fuzzy theory.

The main idea of neutrosophic theory is that a very view point has a degree of truth, uncertainty, and falsity. Hence, T,I and F have been introduced as Neutrosophic theory, which represent the authenticity, uncertainty and absurdity of events respectively. These neutral elements are named true, indeterminate and false values.

3.2 Neutrosophic C-Means clustering algorithm.

In clustering, the degree of every group can be described only by traditional fuzzy clustering methods. But, when there is samples on the boundary region between various groups, it is difficult to determine

which group they belong to and what partitions they join in. So, to handle these difficulties, (Guo et al., 2015) improved the FCM based on the neutrosophic theory and presented the Neutrosophic C-means clustering algorithm (NCM). Anew unique set A has been introduced, which refers to the union of the determinant clusters and indeterminate clusters.

$$Let\ A = C_j \cup B \cup R, j = 1, 2, 3, \dots, C \qquad (5)$$

Where:

C_j: is an indeterminate cluster,

B: represents the clusters in boundary regions.

R is related to noisy data.

$\cup$: is the union operation.

$B\ and\ R$ are two kinds of indeterminate clusters.

T: is defined b as the determinant clusters degree.

I: is the boundary clusters .

F: is the degree belonging to the noisy data set.

By taking the clustering with indeterminacy, a new objective function and membership are defined as:-

$$J(T,I,F,C)=\sum_{i=1}^{N}\sum_{K=1}^{C}(w_1T_{ik})^m\|x_i-v_k\|^2+\sum_{i=1}^{N}\sum_{K=1}^{\binom{C}{2}}(w_2I_{2ik})^m\|x_i-\overline{v_{2k}}\|^2+\sum_{i=1}^{N}\sum_{K=1}^{\binom{C}{3}}(w_3I_{3ik})^m\|x_i-\overline{v_{3k}}\|^2+\ldots\ldots+\sum_{i=1}^{N}\sum_{K=1}^{\binom{C}{C}}(w_CI_{Cik})^m\|x_i-\overline{C_{ck}}\|^2+\sum_{i=1}^{N}(\overline{w_{c+1}}+F_i)^m \quad (6)$$

Where:

w_i: is the weight factor.

δ: is used to control the number of objects considered as outliers.

When the clustering number C is greater than 3, the objective function is very complex .After simplification, the objective function is rewritten as:

$$J(T,I,F,C)=\sum_{K=1}^{C}(w_1T_{ik})^m\,\|x_i-v_k\|^2+\sum_{K=1}^{C}(w_1T_{ik})^m\,\|x_i-v_k\|^2+\sum_{i}^{N}(w_2I_i)^m\left\|x_i-\overline{v_{\imath max}{}^2}\right\|^2+\sum_{i}^{N}(\overline{w_{c+1}}F_i)^2\delta^2 \quad (7)$$

4. Numerical Analysis:-

4.1 Data Description

The following figure shows the trend of data on the retention limit and Reinsurance commission rate issued in the companies under study during the period. The most important variables included in the model are:

- The retention limit in insurance companies is simply the portion of the insurance risk for which the insurance company itself bears direct responsibility, without resorting to reinsurance companies. In other words, it is the maximum loss that the company can bear from any one insurance claim or from a group of claims during a given period of time(Mahasneh, 2024).
- Reinsurance commission rate issued.

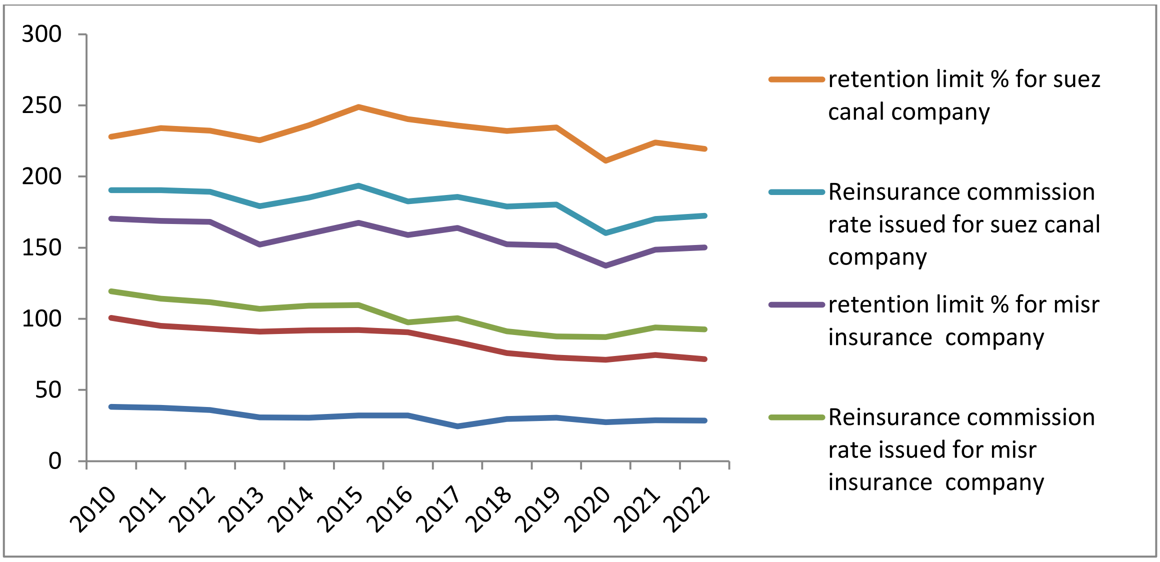

Figure 2: The time series for misr and sueuz canal Insurance companies during the period from 2010 to 2022
Source: Statistical book

4.2 Numerical Application

This study depended on the data of two insurance companies contained Misr insurance company and Suez Canal company. Each company has two variables retention limit and reinsurance commission rate.

Table1: Original Data of insurance companies

No.	Company	retention limit %	Reinsurance commission rate
1	Misr Insurance	51	18.6
2	Misr Insurance	54.6	19.1
3	Misr Insurance	56.5	18.7
4	Misr Insurance	45.2	16.1
5	Misr Insurance	50.6	17.3
6	Misr Insurance	57.7	17.5
7	Misr Insurance	61.4	7

8	Misr Insurance	63.4	16.9
9	Misr Insurance	61	15.4
10	Misr Insurance	63.9	14.7
11	Misr Insurance	50	16
12	Misr Insurance	54.8	19.3
13	Misr Insurance	57.7	20.8
14	Suez Canal	37.6	20.1
15	Suez Canal	43.6	21.6
16	Suez Canal	42.9	21.2
17	Suez Canal	46.2	27
18	Suez Canal	50.9	25.5
19	Suez Canal	55.3	26.2
20	Suez Canal	57.8	23.6
21	Suez Canal	50.2	21.8
22	Suez Canal	53.1	26.5
23	Suez Canal	54.3	28.8
24	Suez Canal	50.9	23
25	Suez Canal	53.7	21.6
26	Suez Canal	47	22.2

Table 1 displayed the original data of misr insurance company and suez Canal insurance company with two variables represented reinsurance commission rate and retention limit. The highest value of retention limit was 63.9 % with reinsurance commission rate 14.7 and the less value of retention limit represented45.2 with reinsurance commission rate 16.1 according to misr insurance company. For suez canal company, the retention limit has the highest value with 57.8 versus 23.6 for reinsurance commission rate, while the less value for the same variable was 37.6 with 20.1 reinsurance commission rate. This proved that there were an inverse relationship between two proposed variables.

Table 2: Descriptive statistics of insurance companies variables

Measure	Retention limit	Reinsurance Commission rate
Minimum	37.60	7.00
1st quartile	50.05	17.45
Median	53.40	20.45
Mean	52.74	20.25
3rd Quartile	57.40	22.80
Maximum	63.90	28.80

Table 2 proposed the descriptive statistics of the study variables, included min., first quartile, median, mean, third quartile and maximum value. The retention limit mean was 52.74 and the reinsurance commission rate was 20.25 this emphasized that the values of retention limit located around the mean greater than the values of reinsurance commission rate.

Table 3: Errors of classification for FCM and Neutrosophic C-M.

Observation No.	FCM			Neutrosophic C-M		
	Cluster1	Cluster2	sum	Cluster1	Cluster2	sum
1		×	1		×	1
4		×	1		×	1
5		×	1		×	1
11		×	1		×	1
19	×		1			
20	×		1	×		1
22	×		1			
23	×		1			
25	×		1	×		1
sum	5	4	9	2	4	6

Table 3 introduced the classification errors for both FCM and NCM for two clusters. For FCM, cluster one had two errors while cluster two had four errors. According to Neutrosophic C-M, there were one error of first cluster and four errors for the second cluster, which means that using Neutrosophic C-M reduced the errors related to the classification and enhanced the accuracy of models used.

Table 4: Accuracy of classification for FCM and Neutrosophic C-M.

cluster	FCM		Neutrosophic CM	
	T	F	T	F
1	8	5	9	4
2	9	4	11	2
sum	17	9	20	6

Table 4 presented the FCM and NFCM classification accuracy. As accuracy equals the number of true by the total number of true and false. The accuracy take the form:-

$$Accuracy = \frac{T}{T+F} \qquad (8)$$

The accuracy of FCM = 17/26 ≅ 65.4%, while, the accuracy of neutrosophic C-Means = 20/26 ≅ 77%. The results of accuracy showed that neutrosophic C- means was more accuracy than Fuzzy C- means . As a result, using neutrosophic enhancing the performance of the data under this study by increasing the accuracy of Fuzzy C- means.

Table 5 : Truth, Falsity and Indeterminacy measures

No	T		F		I
	Cluster 1	Cluster 2	Cluster 1	Cluster 2	Cluster 1- Cluster 2
1	0.716	0.284	0.284	0.716	0.432
2	0.191	0.809	0.809	0.191	0.612
3	0.026	0.974	0.974	0.026	0.947
4	0.795	0.205	0.205	0.795	0.590
5	0.666	0.334	0.334	0.666	0.331
6	0.005	0.995	0.995	0.005	0.991
7	0.258	0.742	0.742	0.258	0.485
8	0.111	0.889	0.889	0.111	0.777
9	0.078	0.922	0.922	0.078	0.843
10	0.140	0.860	0.860	0.140	0.721
11	0.637	0.363	0.363	0.637	0.274
12	0.175	0.825	0.825	0.175	0.650
13	0.069	0.931	0.931	0.069	0.861

14	0.782	0.218	0.218	0.782	0.565
15	0.911	0.089	0.089	0.911	0.821
16	0.891	0.109	0.109	0.891	0.782
17	0.877	0.123	0.123	0.877	0.755
18	0.834	0.166	0.166	0.834	0.669
19	0.506	0.494	0.494	0.506	0.013
20	0.235	0.765	0.765	0.235	0.530
21	0.947	0.053	0.053	0.947	0.893
22	0.670	0.330	0.330	0.670	0.340
23	0.596	0.404	0.404	0.596	0.192
24	0.894	0.106	0.106	0.894	0.787
25	0.488	0.512	0.512	0.488	0.023
26	0.989	0.011	0.011	0.989	0.978

Table 5 introduced the truth membership, indeterminacy membership and falsity membership for the clusters in this study. The results of the table showed that the highest value of truth membership was for cluster one, while cluster two had the highest value of falsity. Finally, indeterminacy membership which represented the difference between cluster one and cluster two had the highest percent with 97.8%.

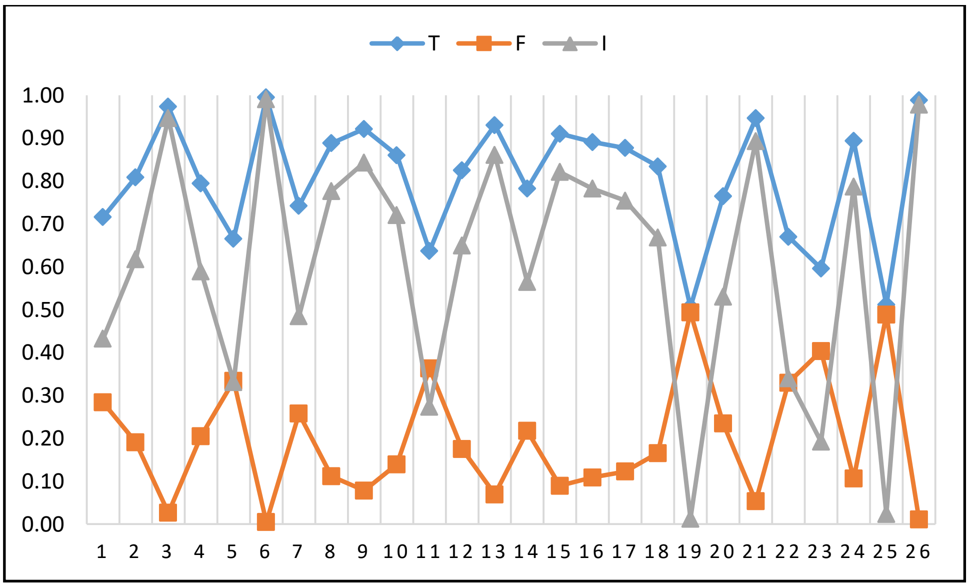

Figure3: I,T& F membership for neutrosophic logic

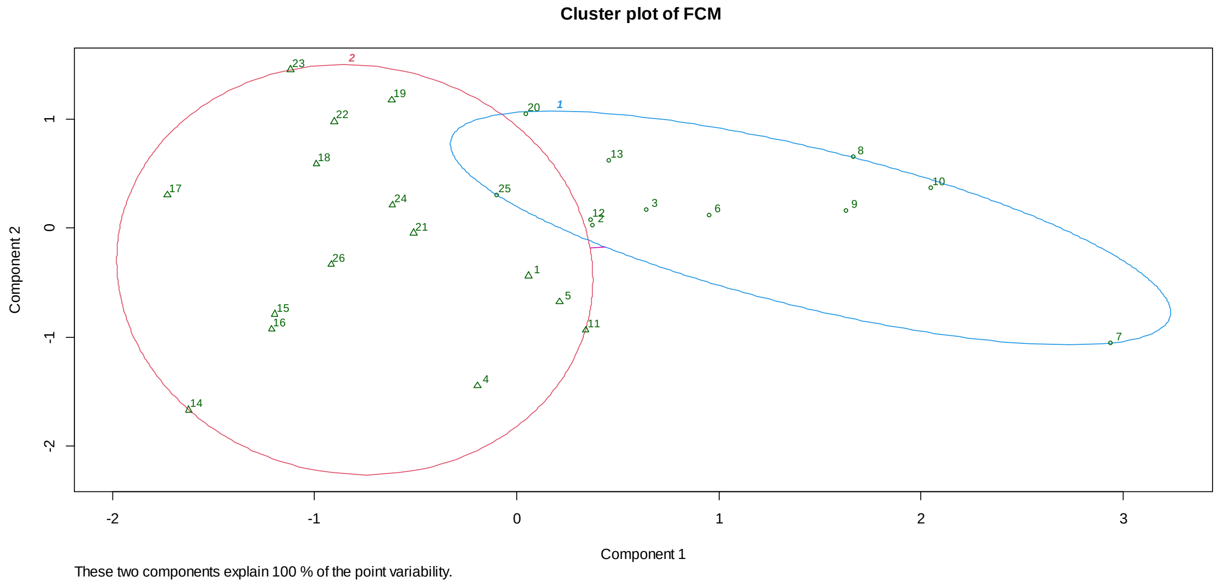

Figure 3 displayed the memberships for neutrosophic logic for each cluster. The figure showed that these memberships varied from increasing to decreasing through observations.

Figure4: Cluster plot of FCM

.

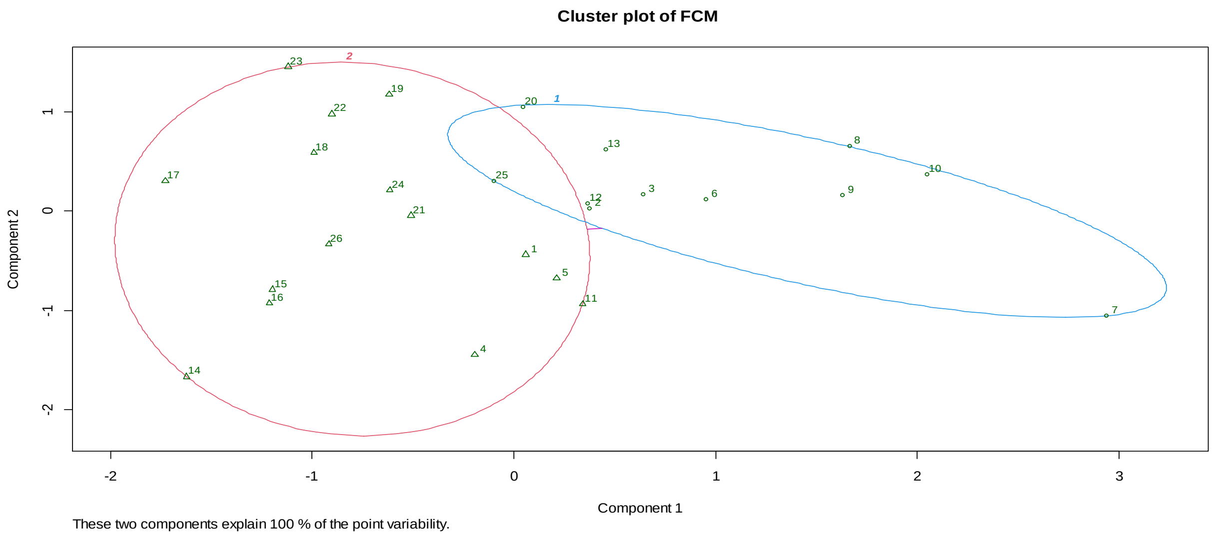

Figure5: Cluster plot of neutrosophic C-M.

Figure4 &5 displayed the plot of clusters under FCM and NCM. These figures had the first two principal components, which meant that 100% of variability of the data is captured by this plot of components 1 and 2.

Conclusion:-

This paper displayed the accuracy of applying Fuzzy C means and Neutrosophic C – means on two insurance companies with one dependent variable represented retention limit and one independent variable was the reinsurance commissions rate. The determined data divided into two clusters, each cluster has thirteen observation with the total number of observations contained 26 item to enhance the relationship between the two mentioned variables. The main concern of this study was to handle the disadvantages of FCM by incorporating FCM with neutrosophic sets extracting neutrosophic C-Means. The extracted approach enhancing the accuracy of the results under this type of data. The accuracy results increasing from65% under FCM to 77% under neutrosophic C-Means.

References:-

Abdelhafeez, A., & Mohamed, H. K. (2023). Skin cancer detection using neutrosophic c-means and fuzzy c-means clustering algorithms. *Journal of intelligent systems and internet of things*, *8*(1), 33-42.

Abdelhafeez, A., Mohamed, H. K., Maher, A., & Abdelmonem, A. (2023). A Neutrosophic based c-means approach for improving breast cancer clus tering performance. *Neutrosophic Sets and Systems*, *53*, 317-330

Abdullah, A. A., Ahmed, A. M., Rashid, T., Veisi, H., Rassul, Y. H., Hassan, B., ... & Shamsaldin, A. S. (2024). Advanced Clustering Techniques for Speech Signal Enhancement: A Review and Metanalysis of Fuzzy C-Means, K-Means, and Kernel Fuzzy C-Means Methods. arXiv preprint arXiv:2409.19448.

Al-Janabee, O., & Al-Sarray, B. (2022). Fuzzy C means Based Evaluation Algorithms For Cancer Gene Expression Data Clustering. Iraqi Journal For Computer Science and Mathematics, 3(2), 27-41

Guo, Y., & Sengur, A. (2015). NCM: Neutrosophic c-means clustering algorithm. *Pattern Recognition*, *48*(8), 2710-2724

Ikotun, A. M., Ezugwu, A. E., Abualigah, L., Abuhaija, B., & Heming, J. (2023). K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data. *Information Sciences*, *622*, 178-210

Lu, Z., Qiu, Y., & Zhan, T. (2019). Neutrosophic C-means clustering with local information and noise distance-based kernel metric image segmentation. *Journal of Visual Communication and Image Representation*, *58*, 269-276

Mahasneh, N. (2024). Claims Handling as an Illustration of Legal Rules Specific to Reinsurance.

Ménard, M., Demko, C., & Loonis, P. (2000). The fuzzy c+ 2-means: solving the ambiguity rejection in clustering. *Pattern recognition*, *33*(7), 1219-1237.

Qiu, H., Liu, Z., & Letchmunan, S. (2024). Incm: Neutrosophic c-means clustering algorithm for interval-valued data. *Granular Computing*, *9*(2), 34.

Smarandache, F. (1998). Neutrosophy: neutrosophic probability, set, and logic: analytic synthesis & synthetic analysis.

Trávez-Castellano, A. M., Arias-Palma, G. B., & Romero-Corral, R. A. (2024). The Rheological Analysis of Ancestral Beverages, Neutrosophic C-Means (NCM) Clustering Algorithm of Data. *Neutrosophic Sets and Systems*, *69*(1), 14.