ICSCCD 2018 p-ISSN 2716-1919 e-ISSN 2716-2907
Volume 1, January 2020 doi:10.31327/icusn-adri.v1i0.1147

Implementation of Fuzzy C-Means for Grouping Socio-Economic Statuses

M N Sutoyo

Universitas Sembilanbelas November Kolaka, Sulawesi Tenggara, Indonesia

E-mail:mns.usn21@gmail.com

Abstrak. Economic status is the level of prestige held by a person in a society on the basis of work or other material property which may indicate individual status. Applying data mining can help analyze data obtained from economic status conditions. The data mining technique used is to use clustering techniques. One method of clustering is the Fuzzy C-Means algorithm. F-C-Means has a high degree of precision and a fast calculation time. Testing the validity of the results of clustering to classify socio-economic status using the calculation of the partition coefficient (PC) obtained at 0.791, this means that we can say that the results of clustering are classified in one or the other. 'other category. Based on the results of the calculations that the Fuzzy C-Means method can classify socio-economic status.

1. Introduction

Socioeconomic status indicates certain inequalities, where community members have jobs that vary in their achievements, and some people have better access to higher status jobs than others, different levels of education, better access to better education than others, different economic resources and the level of power to influence community institutions [1]. We can therefore conclude that socio-economic status is a condition for someone who has more status than others.

Computing (computer science) is generally defined as the study of calculation, hardware and software. Computing covers a variety of computer-related topics, ranging from abstract analysis of algorithms to more concrete topics such as programming languages, software, including hardware. As a discipline, computer science puts more emphasis on computer programming and software engineering, while computer engineering is more likely to be related to things such as computer hardware. However, these two terms are often misunderstood by many people.

One method of grouping is to use the fuzzy grouping method, namely the Fuzzy C-Means algorithm. Where F-C-Means has a high level of precision and a fast calculation time [2]. Some previous studies using F-C-Means, including [3] applying the Fuzzy C-Means method to classify basic math skills and learning outcomes. Fuzzy C-Means method for the early detection of mathematical reasoning skills. While in this study, the Fuzzy C-Means method is used to classify socio-economic status [4]. The aspects of socioeconomic status classification are the number of dependents and income. Knowledge of socio-economic status can be used as a determinant of further policy, for example: improving economic status, providing assistance, etc.

ICSCCD 2018 p-ISSN 2716-1919 e-ISSN 2716-2907
Volume 1, January 2020 doi:10.31327/icusn-adri.v1i0.1147

2. Materials and methods

2.1 Fuzzy C-Means

Fuzzy grouping is part of pattern recognition. Fuzzy clustering plays the most important role in the search for structure in data [4]. Fuzzy grouping is a technique for determining optimal clusters in a vector space based on the normal Euclidean form for the distances between the vectors. The fuzzy grouping is a technique allowing to determine optimal clusters in a vector space on the basis of the normal Euclidean form for the distances between the vectors. F-C-Means is a data aggregation technique in which the existence of each data point in a cluster is determined by the degree of membership [5].

The basic concept of fuzzy F-C-Means, the first is to determine the center of the cluster, which will mark the average location for each cluster. In the initial state, the center of the cluster is still inaccurate. Each data point has a degree of membership for each cluster [6]. The F-C-Means algorithm is as follows.

1. Input data to group X, in the form of a matrix size n x m (n = number of data samples, m = attribute of each data). X_{ij} = sample data i (i=1,2,...,n), attribute j (j=1,2,...,m).
2. specify it:
   - Number of clusters = c
   - Rank = w
   - Maximum iteration = Maxlter
   - Smallest expected error = ξ
   - Objective functions = $P_0 = 0$
   - Initial iteration = t = 1
3. Generate random numbers μ_{ik}, i=1,2,...,n; k=1,2,...,c; as elements of the initial partition matrix U. Count the number of each column:

   $$Q_i = \sum_{k=1}^{c} \mu_{ik}$$

   With j=1,2,...,n. Count:

   $$\mu_{ik} = \frac{\mu_{ik}}{Q_i}$$

4. Calculate the center of the k cluster: V_{kj}, with k=1,2,...,c; and j=1,2,...,m.

   $$V_{kj} = \frac{\sum_{i=1}^{n}\left((\mu_{ik})^{w} * X_{ij}\right)}{\sum_{i=1}^{n}(\mu_{ik})^{w}}$$

5. Calculate the objective function on the iteration t, P_t:

   $$P_t = \sum_{i=1}^{n}\sum_{k=1}^{c}\left(\left[\sum_{j=1}^{m}\left(X_{ij} - V_{kj}\right)^{2}\right](\mu_{ik})^{w}\right)$$

6. Calculate the partition matrix changes:

   $$\mu ik = \frac{\left[\sum_{j=1}^{m}\left(X_{ij} - V_{kj}\right)^{2}\right]^{\frac{-1}{w-1}}}{\sum_{k=1}^{c}\left[\sum_{j=1}^{m}\left(X_{ij} - V_{kj}\right)^{2}\right]^{\frac{-1}{w-1}}}$$

   with: i = 1,2,...,n; andd k = 1,2,...,c

ICSCCD 2018 p-ISSN 2716-1919 e-ISSN 2716-2907
Volume 1, January 2020 doi:10.31327/icusn-adri.v1i0.1147

7. Check the stop condition:
   - If ($|P_t - P_{t-1}| < \xi$) or (t > MaxIter) so stop.
   - Otherwise t = t + 1, repeat steps 4.

2.2 Validitas Clustering

Test the validity of clustering results with the fuzzy clustering approach, using the partition coefficient or partition coefficient (PC) calculation. Bezdek (1981), Partition Coeffecient (PC) as evaluation of the data in each cluster, where the value of Partition Coeffecient (PC) evaluates only the value of the degree of belonging and whatever the value of the data containing information. The formula used to calculate the validity of the partition coefficient (PC) is

$$PC = \sum_{k=1}^{n}\sum_{i=1}^{c}(u_{ik})^2 / n$$

where n is the number of data, c is the number of clusters and uik indicates the membership value of the k-th data in the i-th cluster. The value of the validity test results in the interval [0, 1], plus the value (close to 1) means that the quality of the clusters obtained improves.

2.3 Socio-economic status

Social status is a place where a person in general is associated with other people, relationships with others in the community, their prestige and their rights and obligations [7]. Position or status means the position or place of a person in a social group. The higher the position of a person, the easier it is to obtain the necessary and desired facilities [8].

The aspects of determining socioeconomic status are those of Table 1 below.

Table 1. Criteria for socio-economic status

Criteria	The description
Number of dependents	The number of dependents is the number of family members who are under the responsibility of the head of the family
Receipts	Income is the amount of money received by parents or guardians of a student

3. Results and discussion

The number of dependents and the income of 1414 people are known as in Table 2 below.

Table 2. Income and amount of dependents

No	Code	Receipts	Number of dependents
1	190001	965000	4
2	190002	1067000	3
3	190003	1912000	4
4	190004	1846000	4
5	190006	2620000	3
6	190007	2300000	5
7	190008	500000	5
8	190009	534000	5
9	190010	332000	6

ICSCCD 2018 p-ISSN 2716-1919 e-ISSN 2716-2907
Volume 1, January 2020 doi:10.31327/icusn-adri.v1i0.1147

10	190011	3971000	7
11	190012	650000	4
12	190013	442000	1
13	190014	408000	6
14	190015	3130000	3
15	190017	3344000	5
16	190018	500000	3
17	190019	1000000	4
18	190020	995000	4
19	190021	859000	5
20	190022	317000	4
...	...	....	....
1400	191490	431000	3
1401	191491	3449000	6
1402	191492	2756000	1
1403	191493	3993000	6
1404	191494	3686000	1
1405	191495	1993000	4
1406	191496	646000	4
1407	191497	760000	6
1408	191498	464000	1
1409	191499	271000	5
1410	191500	577000	4
1411	191501	3358000	5
1412	191502	891000	5
1413	191504	1067000	3
1414	191505	1345000	3

On the basis of the above data, it will be grouped into 3 groups, namely from the lowest group (group 1) to the highest group (group 5). The grouping process uses the grouping of fuzzy means C, with the following values.

- Number of clusters = 5
- Maximum iteration = 100
- Smallest expected error = 10^{-5}
- Objective functions = 0
- Initial iteration = 1

By generating the initial value, U is obtained randomly. When the cumulative value of belonging to all the clusters is equal to 1. An example of generation of a random matrix U is illustrated by the following illustration.

a. Value μ_1 = 0.3

b. the rest 1 - μ_1 = 0.7 on value μ_2 cannot be greater than the rest

c. value μ_2 = 0.4

d. the rest 1 - μ_1 - μ_2 = 0.3 on value μ_3 cannot be greater than the rest

e. value μ_3 = 0.1

ICSCCD 2018 p-ISSN 2716-1919 e-ISSN 2716-2907
Volume 1, January 2020 doi:10.31327/icusn-adri.v1i0.1147

f. the rest 1 - μ_1 - μ_2 - μ_3 = 0.2 on value μ_4 cannot be greater than the rest

g. value μ_4 = 0.2

So the result of generating U values is μ_1 = 0.3; μ_2 = 0.4; μ_3 = 0.1; and μ_4 = 0.2. Examples of accumulation of U matrix values are presented in Table 3.

Table 3. Matrix values of U

No	u1	u2	u3	u4	u5
1	0,34	0,22	0,36	0,33	0,10
2	0,31	0,22	0,05	0,15	0,23
3	0,13	0,27	0,16	0,33	0,08
4	0,11	0,35	0,35	0,26	0,25
5	0,21	0,16	0,16	0,04	0,29
6	0,15	0,12	0,01	0,36	0,10
7	0,04	0,32	0,38	0,17	0,28
8	0,02	0,16	0,32	0,32	0,29
9	0,11	0,38	0,25	0,25	0,08
10	0,37	0,38	0,02	0,38	0,10
11	0,06	0,11	0,03	0,05	0,23
12	0,16	0,12	0,08	0,27	0,10
13	0,35	0,15	0,18	0,32	0,08
14	0,10	0,19	0,17	0,17	0,24
15	0,14	0,15	0,15	0,28	0,15
16	0,25	0,25	0,20	0,32	0,12
17	0,12	0,32	0,36	0,26	0,37
18	0,08	0,38	0,32	0,09	0,10
19	0,22	0,24	0,03	0,08	0,40
20	0,07	0,32	0,10	0,26	0,08
...	**...**	**...**	**...**	**...**	**...**
1400	0,07	0,05	0,20	0,03	0,20
1401	0,18	0,15	0,13	0,08	0,14
1402	0,06	0,08	0,36	0,25	0,21
1403	0,17	0,17	0,12	0,16	0,38
1404	0,29	0,04	0,04	0,01	0,23
1405	0,36	0,13	0,01	0,15	0,12
1406	0,23	0,07	0,18	0,32	0,06
1407	0,11	0,17	0,39	0,32	0,02
1408	0,20	0,35	0,3	0,12	0,01
1409	0,40	0,23	0,35	0,12	0,04
1410	0,17	0,07	0,03	0,09	0,13
1411	0,11	0,12	0,26	0,16	0,31
1412	0,35	0,39	0,31	0,33	0,13
1413	0,16	0,35	0,06	0,13	0,17
1414	0,37	0,33	0,23	0,30	0,22

Then calculate the center point of the cluster using the equation. Here is an example of calculating the center of the cluster, namely the center of the first cluster.

ICSCCD 2018 p-ISSN 2716-1919 e-ISSN 2716-2907
Volume 1, January 2020 doi:10.31327/icusn-adri.v1i0.1147

Table 4. Calculation of cluster centers

U1	u^2	u^2 * X1	u^2 * X2
1	0,12	111554,00	0,4624
2	0,10	102538,70	0,2883
3	0,02	32312,80	0,0676
4	0,01	22336,60	0,0484
5	0,04	115542,00	0,1323
6	0,02	51750,00	0,1125
7	0,00	800,00	0,008
8	0,00	213,60	0,002
9	0,01	4017,20	0,0726
10	0,14	543629,90	0,9583
11	0,00	2340,00	0,0144
12	0,03	11315,20	0,0256
13	0,12	49980,00	0,735
14	0,01	31300,00	0,03
15	0,02	65542,40	0,098
16	0,06	31250,00	0,1875
17	0,01	14400,00	0,0576
18	0,01	6368,00	0,0256
19	0,05	41575,60	0,242
20	0,00	1553,30	0,0196
...	**...**	**...**	**...**
1400	0,00	2111,90	0,0147
1401	0,03	111747,60	0,1944
1402	0,00	9921,60	0,0036
1403	0,03	115397,70	0,1734
1404	0,08	309992,60	0,0841
1405	0,13	258292,80	0,5184
1406	0,05	34173,40	0,2116
1407	0,01	9196,00	0,0726
1408	0,04	18560,00	0,04
1409	0,16	43360,00	0,8
1410	0,03	16675,30	0,1156
1411	0,01	40631,80	0,0605
1412	0,12	109147,50	0,6125
1413	0,03	27315,20	0,0768
1414	0,14	184130,50	0,4107
Jumlah	75,28	131261188,39	295,97
Pucat Cluster		1743753,09	3,93

And so on to calculate the other central points of the cluster. In order to obtain the value of the central point of each cluster (V) in the 1st iteration, proceed as follows.

ICSCCD 2018 p-ISSN 2716-1919 e-ISSN 2716-2907
Volume 1, January 2020 doi:10.31327/icusn-adri.v1i0.1147

$$V = \begin{pmatrix} 1743753,09 & 3,93 \\ 1692125,42 & 4,02 \\ 1722230,47 & 3,98 \\ 1688093,09 & 3,95 \\ 1732897,28 & 4,01 \end{pmatrix}$$

Then calculate the value of the objective function. Where the results of the calculation of the objective value obtained by the calculation of the difference in objective function (P) in iteration-1 of 65083.45. Since the value of the objective function difference (P)> 10-5, repeat the calculation.

Then correct the value of the matrix U using the equation. The new degrees of adhesion as shown in Table 5.

Table 5. New levels of membership

$\mu_{ik} = \frac{\left[\sum_{j=1}^{2}(x_{ij}-v_{1j})^2\right]^{\frac{-1}{w-1}}}{\sum_{k=1}^{3}\left[\sum_{j=1}^{2}(x_{ij}-v_{1j})^2\right]^{\frac{-1}{w-1}}}$	$\mu_{ik} = \frac{\left[\sum_{j=1}^{2}(x_{ij}-v_{1j})^2\right]^{\frac{-1}{w-1}}}{\sum_{k=1}^{3}\left[\sum_{j=1}^{2}(x_{ij}-v_{1j})^2\right]^{\frac{-1}{w-1}}}$	$\mu_{ik} = \frac{\left[\sum_{j=1}^{2}(x_{ij}-v_{1j})^2\right]^{\frac{-1}{w-1}}}{\sum_{k=1}^{3}\left[\sum_{j=1}^{2}(x_{ij}-v_{1j})^2\right]^{\frac{-1}{w-1}}}$	$\mu_{ik} = \frac{\left[\sum_{j=1}^{2}(x_{ij}-v_{1j})^2\right]^{\frac{-1}{w-1}}}{\sum_{k=1}^{3}\left[\sum_{j=1}^{2}(x_{ij}-v_{1j})^2\right]^{\frac{-1}{w-1}}}$	$\mu_{ik} = \frac{\left[\sum_{j=1}^{2}(x_{ij}-v_{1j})^2\right]^{\frac{-1}{w-1}}}{\sum_{k=1}^{3}\left[\sum_{j=1}^{2}(x_{ij}-v_{1j})^2\right]^{\frac{-1}{w-1}}}$
0,185	0,213	0,196	0,215	0,191
0,183	0,215	0,195	0,217	0,189
0,262	0,153	0,206	0,148	0,231
0,298	0,131	0,203	0,125	0,243
0,213	0,190	0,203	0,188	0,207
0,220	0,184	0,204	0,182	0,211
0,191	0,208	0,198	0,209	0,194
0,191	0,208	0,198	0,210	0,194
0,192	0,207	0,198	0,208	0,195
0,205	0,196	0,201	0,195	0,203
0,190	0,209	0,197	0,211	0,193
0,191	0,207	0,198	0,209	0,195
0,192	0,207	0,198	0,209	0,195
0,208	0,193	0,202	0,192	0,205
0,207	0,194	0,201	0,193	0,204
0,191	0,208	0,198	0,209	0,194
0,185	0,213	0,196	0,216	0,190
0,185	0,213	0,196	0,216	0,190
0,187	0,211	0,197	0,213	0,192
0,192	0,207	0,198	0,208	0,195
...	...	...	...	...
0,191	0,207	0,198	0,209	0,195
0,207	0,195	0,201	0,194	0,204
0,211	0,191	0,202	0,189	0,206
0,205	0,196	0,201	0,195	0,203
0,206	0,195	0,201	0,194	0,203
0,243	0,167	0,206	0,162	0,223
0,190	0,209	0,197	0,211	0,194
0,188	0,210	0,197	0,212	0,193
0,191	0,208	0,198	0,209	0,194
0,192	0,207	0,198	0,208	0,195
0,190	0,208	0,198	0,210	0,194

ICSCCD 2018 p-ISSN 2716-1919 e-ISSN 2716-2907
Volume 1, January 2020 doi:10.31327/icusn-adri.v1i0.1147

0,207	0,194	0,201	0,193	0,204
0,187	0,212	0,197	0,214	0,192
0,183	0,215	0,195	0,217	0,189
0,171	0,226	0,191	0,231	0,181

After the 32 iteration, the objective function difference (P) ≤ 10-5 is obtained, then the iteration stops. The central point of each cluster (V) at the end of the iteration is as follows.

$$V = \begin{pmatrix} 442710,55 & 4,22 \\ 985207,95 & 4,19 \\ 1649850,30 & 3,92 \\ 3315082,22 & 3,82 \\ 3871214,22 & 3,75 \end{pmatrix}$$

From the new diploma membership matrix, one can obtain information on the tendency of a given item to enter the same group. The greater degree of belonging shows the tendency of a given to become a member of a group. Table 6 shows the trend of the data in some clusters.

Table 6. Clusters in trend data

Data	Kecenderungan Cluster				
	C1	C2	C3	C4	C5
1			*		
2			*		
3			*		
4			*		
5				*	
6				*	
7			*		
8			*		
9			*		
10					*
11			*		
12			*		
13			*		
14				*	
15				*	
16	*				
17			*		
18			*		
19			*		
20			*		
...	...	...	...	...	...

ICSCCD 2018 p-ISSN 2716-1919 e-ISSN 2716-2907
Volume 1, January 2020 doi:10.31327/icusn-adri.v1i0.1147

1400		*		
1401			*	
1402			*	
1403				*
1404				*
1405		*		
1406		*		
1407		*		
1408		*		
1409		*		
1410		*		
1411			*	
1412	*			
1413		*		
1414		*		

The information that can be obtained from these five cluster centers is:

a. The first group (1st cluster), contains an average income of around Rp. 442,710.55 and have an average number of dependents 4 people.
b. The second group (second cluster), contains an average income of around Rp. 985,207.95 and has an average number of dependents 4 people.
c. vs. The third group (3rd cluster), has an average income of around Rp. 1,649,850.30 and has an average number of dependents of 3 people.
d. The fourth group (4th cluster), contains an average income of around Rp. 3,315,082.22 and have an average number of dependents of 3 people.
e. The fifth group (5th cluster), contains an average income of around Rp. 3,871,214.22 and have an average number of dependents of 3 people.

At the end of the iteration, test the validity of the cluster results using the equation. Where in the validity test using the partition coefficient formula (PC), namely by calculating the square of each data membership value in each cluster and then adding it. The results of the partition coefficient (PC) calculation were obtained at 0.791. The validity test is in the right category because the value obtained is close to 1.

3. Conclusions

Based on the results of calculations that the Fuzzy C-Means method can determine socio-economic status, namely: The first group (1st cluster), containing an average income of around Rp. 442,710.55 and have a number medium dependents 4 people. The second group (second cluster), contains an average income of around Rp. 985,207.95 and has an average number of dependents 4 people. The third group (3rd cluster), has an average income of around Rp. 1,649,850.30 and has an average number of dependents of 3 people. The fourth group (4th cluster), contains an average income of around Rp. 3,315,082.22 and have an average number of dependents of 3 people. The fifth group (5th cluster), contains an average income of around Rp. 3,871,214.22 and have an average number of dependents of 3 people.

The validity test is in the right category because the value obtained is close to the number 1. Where the results of the calculation of the partition coefficient (PC) obtained 0.791.

ICSCCD 2018 p-ISSN 2716-1919 e-ISSN 2716-2907
Volume 1, January 2020 doi:10.31327/icusn-adri.v1i0.1147

References

[1] Santrock, JW., 2007, *Perkembangan Anak Edisi 11.* Jakarta: Erlangga
[2] Hammouda, K., Karaay, F. 2000. *A Comparative Study of Data Clustering Techniques.* University of Waterloo, Ontario, Canada.
[3] Sutoyo, MN., Rhita, N., Hayaty, N., 2014, Penerapan Fuzzy C-Means Untuk Mengelompokkan Kemampuan Dasar Berhitung dan Hasil Belajar, SEMINASIK Universitas Gadjah Mada.
[4]Sutoyo, MN dan Sumpala, AT., 2016, Penerapan Fuzzy C-Means untuk Deteksi Dini Kemampuan Penalaran Matematis, cientific Journal of Informatics Vol. 2 No. 2 pp 129-135.
[5] Kusumadewi, S., Hartati, S., Harjoko, A., Wardoyo, R., 2006. *Fuzzy Multi-Attribute Decision Making (Fuzzy MADM)*. Graha Ilmu. Yogyakarta.
[6] Zimmermann., 2001, *Fuzzy Set Theory and Its Applications*, Fourth Edition. Kluwer Academic Publishers.
[7] Abdulsyani. 2007. *Sosiologi, Skematika, Teori, dan Terapan*. Jakarta: PT.Bumi. Aksara.
[8] Wijianto dan Ulfa, IK., 2016, Pengaruh Status Sosial dan Kondisi Ekonomi Keluarga terhadap Motivasi Bekerja bagi Remaja Awal (Usia 12-16 Tahun) di Kabupaten Ponorogo, Al Tijarah: Vol. 2, No. 2 pp 190 – 210.