6. HASIL CEK_60960140

by 60960140 Te

Submission date: 08-Aug-2022 11:07AM (UTC+0700)

Submission ID: 1880089223

File name: 6. TE-60960140-Clustering Majors for New Students at Vocational High School Muhammadiyah 3 Yogyakarta Using Fuzzy C-Means.pdf (277.16K)

Word count: 4120

Character count: 18503

KINETIK, Vol. 3, No. 3, August 2018, Pp. 271-278
ISSN : 2503-2259
E-ISSN : 2503-2267

Clustering Majors for New Students at Vocational High School Muhammadiyah 3 Yogyakarta Using Fuzzy C-Means

Rifqi Rahmatika Az-Zahra[*1], Rusydi Umar[2], Abdul Fadlil[3]
[1,2,3]Universitas Ahmad Dahlan/Magister of Informatics Engineering
rifqi1707048012@webmail.uad.ac.id[*1], rusydi_umar@rocketmail.com[2], fadlil@mti.uad.ac.id[3]

Abstract

The development of the era demands quality human resources. For that need to be equipped with the knowledge and skills to be ready to compete in the world of work. Vocational High School has many skills programs. Examples of skills programs in Vocational High School are Computer Network Engineering, Audio and Video Engineering, Mechanical Engineering, Building, Drawing Techniques, Motorcycle Engineering, Installation and Power Engineering, systems that can assist in the decision-making process of prospective students. The grouping system will be created using the Fuzzy C-Means method. Grouping majors using fuzzy c-means algorithm is expected to help prospective students choose majors from the many majors that exist. This is done so that prospective students do not experience difficulties in learning activities, can develop optimally and is expected to work in accordance with the expertise that has been owned, so that ultimately can improve the quality of output and outcome of vocational education. The results obtained in this study in the form of data grouped based on 3 clusters, so that prospective students can choose the majors in accordance with the selected cluster.

Keywords: *Clustering, Fuzzy C-Means (FCM), Vocational High School*

1. Introduction

The development of the era demands quality human resources. For that need to be equipped with the knowledge and skills to be ready to compete in the world of work. In Law no. 20 Year 2003 of National Education System Article 18 explains that: "Vocational education is secondary education that prepares students primarily to work in a particular field". Of the above law, it is necessary to develop a form of vocational education at Vocational High School that can be implemented in accordance with the needs of the world of work. Djohar [1] vocational education is an educational program that prepares individual learners to become a professional workforce.

Vocational High School has many skills programs. The skills program implemented in Vocational High School adjusts to the needs of the world of work. Here are some examples of majors / areas of job skills: Computer Networking Techniques, Audio and Video Techniques, Engineering Machines, Buildings, Image Techniques, Motorcycle Engineering, Installation and Power Engineering, Light Vehicle Engineering, Pharmacy. With the many majors / areas of expertise in Vocational High School, many choices of majors face for prospective students who will choose the major, So that required decision support system that can assist in the decision-making process of prospective students based on predetermined criteria quickly.

Research in pattern recognition is developed so diverse. In example Machine vision, Object character recognition (OCR), Computer aided diagnosis, Speech recognition, Face recognition, Biometric, Image database retrieval, Data mining, and Bioinformatic. The application uses several algorithms: Artificial Neural Network (ANN), Backpropagation, Learning Vector Quantization LVQ), Reeman Chain Code, perception, and Fuzzy C-Means (FCM). In the study of Abdul Fadlil [2] In this research resulted from ANN with the best recognition result. Abdul Fadlil [3] In this research has developed a new approach of human skin irritation identification system automatically based on image pattern recognition. To evaluate the performance of the system is done by combining the methods of characterization and classifier extraction. The experimental results show that the best accuracy is 83.33% obtained in the system design implemented using GLH or GLCM characteristics through the LVQ-NN classifier.

Sapriani Gustina et al [4], the data in their study were used as a standard reference of 3 samples of mushrooms of 5 types of mushrooms. The test results for the fungal image

Az-Zahra, R., Umar, R., & Fadlil, A. (2018). Clustering Majors for New Students at Vocational High School Muhammadiyah 3 Yogyakarta Using Fuzzy C-Means. Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control, 3(3). doi:http://dx.doi.org/10.22219/kinetik.v3i4.642
Receive March 29. 2018: Revise July 06. 2018: Accepted July 06. 2018

identification system showed a high level of accuracy of 82% with the method of feature extraction method of Order Statistic 1 and distance classification method. Saifudin and Abdul Fadlil [5], in their research on wooden images can be analyzed in character to obtain patterns that indicate the type of wood. The system includes image acquisition, image processing, feature extraction, and feature extraction classification using the Gray Level Co-occurrence Matrix (GLCM) matrix while closer to using the Euclidean distance method. The results from resizing the original image to 30x30, 20x20 and 10x10 each produce 82.5%, 65.7% and 77.5% accuracy. Ahsan Anwar S, et al [6], in their study discussing the application of one method of mining Fuzzy C-Means data into mobile web-based system can simplify and speed up in clustering data.

Fuzzy logic was first developed by Lotfi A. Zadeh, an Iranian American scientist from the University of California at Barkele [7]. One application of fuzzy logic is in clusters or grouping. Clustering is a grouping method based on the size of proximity (similarity). The grouping is based on the proximity of an existing sample characteristic.

Fuzzy C-means Clustering (FCM), also known as Fuzzy ISODATA, is one of the clustering methods that is part of the Hard K-Means method. Fuzzy clustering is one technique for determining the optimal cluster in a vector space based on the normal Euclidean form of the spacing between vectors. Fuzzy C-Means (FCM) is a data grouping technique that the existence of each datum in a group is determined by a certain value or degree of membership and this technique was first introduced by Jim Bezdek in 1981 [8].

In this research using fuzzy C-Means method. With the clustering majors using fuzzy c-means algorithm is expected to help prospective students choose majors from the many majors that exist. This is done so that prospective learners do not experience difficulties that, when entering teaching and learning activities and can develop secure optimally. With a study period of about three or four years, Vocational High School graduates are expected to be able to work in accordance with the expertise that has been occupied. So that in the end can improve the quality of output and outcome of vocational education.

2. Research Method

2.1 Research Subject

Subjects in this research are how to classify the majors with the number phone existing majors. Grouping majors according to what the prospective student needs by considering several indicators such as skills, talents, interests. The results will be given in the form of clustering majors in accordance with the abilities and interests of the child.

2.2 Data Collection

The data used for the object of research in this study is the data related to the classification of majors. Data were obtained from questionnaires that had been scattered and several weighted questionnaires for the study. From the data that has been obtained made the supporting criteria for grouping majors.

2.3 Fuzzy C-Means Clustering

Fuzzy logic is used as a way to map problems from inputs to expected outputs. In fuzzy logic there is fuzzy clustering, which is one of the methods to determine the optimal cluster in a vector space based on the normal form of Euclidean for the distance between vectors.

Fuzzy C-Means is one of the data grouping techniques in which the existence of each data point in a cluster is determined by the degree of membership. The Fuzzy C-Means method includes a supervised clustering method in which the number of cluster centers is determined in the clustering process. The purpose of the Fuzzy C-Means is to obtain a cluster center, which will be used to know the data that goes into a cluster [9]. According to Kusumadewi and Purnomo [10], the fuzzy c-means algorithm is as follows:

1. Input data to be clustered X, a matrix of size n x m (n = number of data samples, m = attribute of each datum). Xij the I-the sample data (I = 1,2,... n), the j-attribute (j = 1,2, .. m).
2. Define the cluster
   Number of clusters = c;
   Rank = w;
   Maximum iteration = MaxIter;
   The smallest expected error = ξ

The initial objecting function = PO = 0;
Initial iteration = t = 1;

3. Generate random numbers μ_{ik} i=1,2,...,n; k=1,2...,c; elements of the initial partition matrix U.
$Qi = \sum_{k=1}^{c} \mu_{ik}$
$j = 1,2,\ldots,n.$
Count on Equation 1.

$$\mu_{ik} = \frac{\mu_{ik}}{Qi} \quad (1)$$

4. Calculate the center of the k-cluster on Equation 2: V_{kj}, with k=1,2,…,c; and j=1,2,…,m,[11].

$$\boldsymbol{V_{jk}} = \frac{\sum_{i=1}^{n} ((\boldsymbol{\mu_{ik}})^{w} * \boldsymbol{X_{ij}})}{\sum_{i=1}^{n} (\boldsymbol{\mu_{ik}})^{w}} \quad (2)$$

5. Using Equation 3 to calculate the objective function on the iteration to -t, P_t [11].

$$P_t = \sum_{i=1}^{n} \sum_{k=1}^{c} \left(\left[\sum_{j=1}^{m} (Xij - Vkj)2\right](\mu_{ik})\mathrm{w}\right) \quad (3)$$

6. Calculate the partition matrix change on Equation 4 [11].

$$\mu ik = \frac{[\sum_{j=1}^{m} (Xij - Vkj)2]^{\frac{-1}{w-1}}}{\sum_{k=1}^{c} \sum_{j=1}^{m} (Xij - Vij)2]^{\frac{-1}{w-1}}} \quad (4)$$

7. Check stop condition:
If: (| PT - PT - 1 | Ɛ) or (t> Maxtler) then stop.
If not: t = t + 1, reset step 4th.

Table 1. Table Cluster

Cluster	Information
TAV/ C1	Audio Video Engineering
TKJ/C2	Computer and Network Engineering
TSM/C3	Engineering and Business of Motorcycles

Table 1, shows the groups contained in this study. namely: TAV / C1 who majored in Audio Video Engineering, TKJ / C2, who majored in Computer and Network Engineering, TSM / C3 who majored in Engineering and Business Motorcycles

Table 2. Table Data X (data questionnaire)

	Parameter		
Name	Skill	Talent	Interest
A	35	31	35
B	37	33	33
C	38	29	32
D	35	27	32
E	33	32	30
F	33	28	24
G	33	31	28
H	40	38	39
I	33	29	26
J	41	39	38

Table 2 is the X data, which will be used for grouping data, namely: Consisting of 10 data which is the result of the questionnaire. based on the data obtained, then grouped the object into 3 clusters with attributes TAV, TKJ, and TSM.

Table 3. Table Value of Initial Parameters

Cluster	3
w (Pembobot)	2
e (error)	0.0001
P0 (func Objectif)	0
Max Iterasi	10
First Iterasi	1

In the Table 3, the researcher has determined the initial parameter value of the group number (c) of 3, run (w) or weighed by weighting 2. The smallest expected error of 0.001, Function Objective (P0) is 0, and the maximum iteration of up to 10 iterations. The initialization of the above determination shall be made before producing a random number. This research has determined the initial parameter value of the group number (c) of 3, run (w) or weighed by weighting 2. The smallest expected error of 0.01, Function Objective (P0) is 0, and maximum iteration of up to 10 iterations. The initialization of the above determination should be made before generating random numbers.

3. Results and Discussion

Basically, grouping majors can be grouped after data. Grouping is done to help simplify the determination/grouping of data prospective students in accordance with skills, interests, and talents so that the results of the grouping will facilitate prospective students in teaching and learning activities in schools to be addressed. Results for groups with Fuzzy C-Means as shown in the T.able 4.

Table 4. Random Number Generation Table (μik)

	Cluster Membership (*random*)		
	c1	c2	c3
1	0.349	0.430	0.220
2	0.533	0.333	0.135
3	0.703	0.261	0.037
4	0.415	0.123	0.463
5	0.613	0.281	0.106
6	0.121	0.491	0.389
7	0.419	0.308	0.272
8	0.746	0.185	0.069
9	0.710	0.217	0.074
10	0.098	0.517	0.385

Table 4, it is a table of kernel values (μik) which can be randomly derived from fuzzy, c-means calculation example, according to random number generator η initial partition U as an element of initial partition matrix U. Begin by forming random membership with an assumption of random value with the total of each row of data is always worth 1.

Table 5 is a table that calculates the value of congestion (μik) squared from the calculation of fuzzy c-means. The above stage is the first stage of the cluster calculation with the vector value μ ik2 for the calculation of the cluster center becomes easy. By means of the existing random number data are: c1, c2, c3 are doubled 2.

Table 6 is the table of calculation of the cluster center on the first iteration. Calculation of the cluster center by using equation according to formula (2.3.2). The above calculation results in the value of centroid in Table 7.

Table 7 is Cluster V Center with centroid obtained for 3 clusters in the first iteration. In Table 8. This is a table the calculation of the objective function of the first iteration. With the objective function value reaching 209,415. Difference of objective function = 209,415, Because the change

in the value of the objective function is still above the threshold, the process is continued to the next iteration.

Table 5. Miu Squares

(μ i1) ^2	(μ i2) ^2	(μ i3) ^2
0.122	0.185	0.049
0.284	0.111	0.018
0.494	0.068	0.001
0.172	0.015	0.214
0.375	0.079	0.011
0.015	0.241	0.151
0.176	0.095	0.074
0.556	0.034	0.005
0.504	0.047	0.005
0.010	0.268	0.148
0.122	0.185	0.049

Table 6. Calculation Table Cluster Center 1-3 Iterations 1

Xij (Data is *clustered*)			Degree of Membership in Cluster -1			(μ ik)² * xij			(μik)²*xij			(μ ik)² * xij		
Skill (X_{11})	Talent (X_{12})	Interest (X_{13})	(μ i1) ^2	(μ i2)^2	(μi3)^2									
35	31	35	0.122	0.185	0.049	4.270	3.782	4.270	6.486	5.744	6.486	1.698	1.504	1.698
37	33	33	0.284	0.111	0.018	10.498	9.363	9.363	4.091	3.649	3.649	0.672	0.600	0.600
38	29	32	0.494	0.068	0.001	18.761	14.317	15.798	2.582	1.970	2.174	0.051	0.039	0.043
35	27	32	0.172	0.015	0.214	6.014	4.640	5.499	0.527	0.406	0.482	7.497	5.783	6.854
33	32	30	0.375	0.079	0.011	12.389	12.013	11.263	2.603	2.524	2.367	0.374	0.362	0.340
33	28	24	0.015	0.241	0.151	0.479	0.407	0.349	7.950	6.745	5.782	4.985	4.229	3.625
33	31	28	0.176	0.095	0.074	5.806	5.454	4.926	3.140	2.949	2.664	2.443	2.295	2.073
40	38	39	0.556	0.034	0.005	22.250	21.137	21.693	1.364	1.296	1.330	0.193	0.184	0.188
33	29	26	0.504	0.047	0.005	16.621	14.606	13.095	1.551	1.363	1.222	0.179	0.157	0.141
41	39	38	0.010	0.268	0.148	0.394	0.374	0.365	10.969	10.434	10.166	6.070	5.774	5.626
			2.707	1.142	0.677	97.482	86.094	86.622	41.262	37.082	36.321	24.161	20.927	21.187
$\frac{\sum_{i=1}^{n} ((\mu_{ik})^2 * X_{ij})}{\sum_{i=1}^{n} (\mu_{ik})^2}$														

Table 7. Is the Cluster Iterations 1

$$V = \begin{pmatrix} 36.015 & 31.808 & 32.003 \\ 36.118 & 32.458 & 31.792 \\ 35.693 & 30.915 & 31.300 \end{pmatrix}$$

Table 8. Table Function Calculation Objective Iteration 1

Cluster 1			Cluster 2			Cluster 3			L1+L2+L3
(Xij-Vkj)²	μi1^2	L1	(Xij-Vkj)²	μi2^2	L2	(Xij-Vkj)²	μi3^2	L3	
10.666	0.122	1.301	13.665	0.185	2.532	14.178	0.049	0.688	4.521
3.384	0.284	0.960	2.530	0.111	0.280	8.945	0.018	0.163	1.402
11.825	0.494	5.838	15.548	0.068	1.056	9.479	0.001	0.013	6.907
24.150	0.172	4.150	31.087	0.015	0.468	16.298	0.214	3.491	8.109
13.142	0.375	4.934	13.142	0.079	1.037	10.120	0.011	0.115	6.085
87.644	0.015	1.273	90.319	0.241	21.759	69.039	0.151	10.428	33.461
25.770	0.176	4.534	26.229	0.095	2.495	18.150	0.074	1.344	8.373
103.172	0.556	57.388	97.731	0.034	3.334	128.038	0.005	0.618	61.340
53.016	0.504	26.702	55.232	0.047	2.596	39.010	0.005	0.211	209.509
112.531	0.010	1.080	105.164	0.268	28.134	138.421	0.148	20.493	49.708
					Objective function				209.415

Table 9. Table Function Calculation Objective Iteration 2

Cluster 1			Cluster 2			Cluster 3			L1+L2+L3
$(Xij-Vkj)^2$	$\mu i1^2$	L1	$(Xij-Vkj)^2$	$\mu i2^2$	L2	$(Xij-Vkj)^2$	$\mu i3^2$	L3	
15.981	0.077	1.226	15.837	0.126	1.994	10.335	0.136	1.401	4.621
8.840	0.052	0.458	11.446	0.029	0.332	0.125	0.362	0.045	0.836
13.240	0.103	1.363	10.915	0.178	1.943	18.479	0.066	1.223	4.529
20.433	0.114	2.329	15.240	0.189	2.878	39.183	0.052	2.034	7.241
7.382	0.130	0.962	8.262	0.130	1.077	23.524	0.077	1.818	3.857
67.109	0.126	8.450	62.935	0.134	8.415	119.653	0.078	9.347	26.212
15.360	0.135	2.073	14.975	0.140	2.094	42.603	0.067	2.852	7.018
127.715	0.098	12.565	136.981	0.088	12.091	72.185	0.152	10.937	35.593
37.131	0.130	4.813	34.222	0.141	4.814	78.574	0.070	5.514	15.141
137.145	0.100	13.694	147.785	0.087	12.888	80.081	0.151	12.099	38.682
Objective function									143.728

In Table 9, this is a table the calculation of the objective function of the first iteration. With the objective function value reaching 143,728. Because the change in the value of the objective function is still above the threshold, the process is continued to the next iteration.

Table 10. Table Function Calculation Objective Iteration 3

Cluster 1			Cluster 2			Cluster 3			L1+L2+L3
$(Xij-Vkj)^2$	$\mu i1^2$	L1	$(Xij-Vkj)^2$	$\mu i2^2$	L2	$(Xij-Vkj)^2$	$\mu i3^2$	L3	
9.158	0.144	1.316	11.427	0.141	1.613	32.969	0.060	1.982	4.911
0.197	0.188	0.037	0.528	0.314	0.166	29.001	0.000	0.001	0.204
16.865	0.096	1.626	22.389	0.066	1.467	22.002	0.188	4.134	7.227
37.537	0.075	2.797	46.949	0.041	1.946	17.016	0.274	4.662	9.405
25.704	0.036	0.913	31.378	0.044	1.396	5.617	0.361	2.026	4.335
122.744	0.072	8.866	137.773	0.064	8.753	34.924	0.230	8.019	25.638
45.241	0.044	2.006	53.622	0.042	2.260	4.320	0.341	1.474	5.741
70.915	0.144	10.192	59.235	0.165	9.794	187.877	0.046	8.625	28.611
81.207	0.061	4.981	93.529	0.052	4.873	14.221	0.275	3.906	13.760
79.435	0.141	11.214	66.785	0.164	10.948	198.894	0.048	9.573	31.735
Objective function									131.567

In Table 10, this is a table the calculation of the objective function of the first iteration. With the objective function value reaching 131,568. The change in the value of the objective function has already fallen below the specified threshold, then the process is stopped.

4. Conclusion

Based on the results of the discussion, it can be concluded as follows: From the research using the data above. Grouping Department Using Fuzzy C-Means have been done. Fuzzy C-Means is able to solve existing problems, and the data used consists of 10 data. The results above can be summarized in the table below.

Table 11 above is the level of student membership in the last iteration. Obtained information about the tendency of prospective students to enter into clusters where. The largest membership rate indicates that the highest tendency of prospective students to enter cluster members. With distance matrix: The first cluster member with membership title is 0.1676, Member of the second cluster with membership degree 0.4625, Member of the third cluster with membership level 0,5004 From Table 11 above can be obtained:

1. Cluster 1 consists of 1 prospective study
2. Cluster 2 consists of 3 prospective students
3. Cluster 3 consists of 6 prospective students

The data are grouped by 3 groups, making it easier for prospective students to choose the majors in accordance with the selected cluster.

Table 11. Membership Degrees of Each Student Data for Each Cluster in The Last Iteration.

Data	Degree of Membership of data on *cluster* On Last Iteration			Data entered into the cluster to-		
	1	2	3	TAV/ C1	TKJ/ C2	TSM/ C3
A	0.171	0.213	0.616		√	
B	0.007	0.018	0.976	√		
C	0.275	0.366	0.359		√	
D	0.370	0.463	0.168			√
E	0.410	0.500	0.090		√	
F	0.415	0.466	0.118			√
G	0.438	0.520	0.042			√
H	0.223	0.186	0.591			√
I	0.430	0.495	0.075			√
J	0.230	0.194	0.576			√

References

[1] Djohar, A. *"Technology and Vocational Education. In Science and Applications."* Bandung: Pedagogiana Press. Pp. 1285-1300.2007.

[2] Fadlil, Abdul. *"Comparison of Classification of Distance Function and Artificial Neural Networks on Facial Recognition System,"* SNATI (National Seminar on Information Technology Applications), ISSN: 1907-5022, 2007.

[3] Fadlil, Abdul. *"An Automatic Identification System of Human Skin Irritation."* TELKOMNIKA (Telecommunication Computing, Electronics and Control), Vol.8, No.3, 255-264., ISSN: 1693-6930, 2010.

[4] Gustina, S., Fadlil, A., & Umar, R. (2017). *"Mushroom Identification System Using the Statistical Order Extraction Method 1 and Classification of Distance."* Techno. Come, Vol.16, No. 4, 378-38678, 2017.

[5] Saifudin, S., & Fadlil, A. (2015) "*Site Identification of Wood Based Image Using Gray Level Co-occurrence Matrix (Glcm) With Euclidean Distance Classification."* SINERGI: Technical Journal of Mercu Buana, Vol. 19, No. 3,181-186, ISSN: 1410-233, 2015.

[6] Ahsan Anwar S, Abdul Fadlil, Rusydi Umar, *"Fuzzy C-Means Implementation for Clustering Poor Data With Realtime Web-Based Mobile, Presentation,"* SNTT (National Seminar on Applied Technology), ISSN: 2339-028X, 2017.

[7] Kusumadewi, Sri. *"Artificial Intelligence: Technique and Its Application."* Yogyakarta: Graha Ilmu. 2003.

[8] Kusumadewi, Sri. *"Fuzzy System Design Analysis Using Matlab Tool Box."* Graha Ilmu, Yogyakarta. 2002.

[9] Rizal, A. S., & Judge, R. F. "*Method of K-Means Clustering and Fuzzy C-Means Clustering."* In Proceeding National Seminar on Mathematics and Mathematics Education UMS, ISBN: 978.602.361.002.0, 2015.

[10] Kusumadewi, Sri; Purnomo, Hari. *"Fuzzy Logic Application for Decision Support."* Publisher Graha Ilmu. 2010.

[11] Yan Jun, Michael dan James Power. *"Using Fuzzy Logic (Toward Intelligent System)."* New York: Prentice-Hall. 1994.

6. HASIL CEK_60960140

ORIGINALITY REPORT

10%	10%	0%	0%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	dev.kinetik.umm.ac.id Internet Source	5%
2	garuda.ristekdikti.go.id Internet Source	5%

Exclude quotes	On	Exclude matches	< 5%
Exclude bibliography	On		