Advances in Intelligent Systems Research, volume 172
Sriwijaya International Conference on Information Technology and Its Applications (SICONIAN 2019)

Clustering of Food Security Status in South Sumatera Using Fuzzy C-Means Algorithm

Dian Palupi RINI[1*], Endang Lestari RUSKAN[2]

[1]*dprini@unsri.ac.id, Informatic Engineering, Faculty of Computer Science, Universitas Sriwijaya, Indonesia*
[2]*endanglestari@unsri.ac.id, Information Systems, Faculty of Computer Science, Universitas Sriwijaya, Indonesia*
**Corresponding author: dprini@unsri.ac.id*

ABSTRACT
Food security is defined as the condition of food to satisfy every society which reflected in the availability of sufficient food, a good number and quality, secure, equitable, affordable, and based on the diversity of local resources. According to data from BPS South Sumatera, the percentage of poor population also half-poor family is still high. This shows existence of barriers to accessing food and there are still gaps in distribution of food as well as food consumption inequality according to socio-economic groups and regions. Clustering food security area or food insecurity area in South Sumatera is very necessary because the condition of food availability at each region varies every time, depending on the results of food production and the factors which affected it. In this research, the methods which used for data clustering is Hybrid fuzzy C-means (FCM). From the results of testing using Davies Bouldin Index (DBI), obtained data that food security in South Sumatera is suitable to be cluster into three clusters, which are food secure, enough food and food insecure, where each cluster contains: 1 (one), 8 (eight) and 5 (five) region in South Sumatera. The results of cluster are expected to be used as one of the information for the Government of South Sumatera to plan strategies which effective and efficient with conditions related to food security.
Keywords: *food security, FCM, DBI, cluster*

INTRODUCTION

Food security is defined as condition of sufficient food for every society which is reflected from availability of food, good quantity and quality, secure, equitable, affordable, and based on the diversity of local resources. Food security is the main issue in fulfilling of society welfare and effected on the stability of the economic, social and politic in a country (Rianti, 2016).

According to data from BPS in South Sumatera, there are still percentage of poor population at urban areas in March 2016, it is about 12.74 percent dropped to 12.45 percent in March 2017. While the percentage of poor population at rural areas dropped from 13.99 percent in March 2016 becomes 13.62 percent in March 2017. But if the data was included half-poor family then the numbers become higher. This high number shows the existence of barriers to accessing food and there are still gaps in distribution of food as well as food consumption inequality according to socio-economic group and area (Raharto, 2010). This reality shows that despite in country food security is not a problem, but at the regional level and household food insecurity are still often encountered. With the indication of food insecurity, it is a challenge for South Sumatera to reduce food insecurity, in order to realize the national food security. So, researches about food security or food insecurity in South Sumatera area are very necessary. This research aims to describe the characteristics and model of factors which indicate affecting food security in South Sumatera.

This cluster results are expected to be used as one of the information for the Government of South Sumatera to plan effective strategies and efficient related to food security condition and also public are able to evaluate the production of food on their region.

Research about model of factors which indicate affecting food security, as well as to compare the classification results of food security status in actual with prediction model results of which have been done by some researchers. (Permatasari & Ratnasari, 2016) classify the area of food security using Ordinal Probity Regression approach. (Syamsiah, 2011) classify the area of food security using Fuzzy Tahani approach. (Masitoh & Ratnasari, 2016) classify the area of food security using Binary Probity Regression Method approach. while for clustering problem (Wulandari & Fauzy, 2016) classify the area food security using Self Organizations Maps (SOM) method.

Some of factors which affect food security an area sometimes has ambiguous value, so it is needing to use a modeling which can tolerate a value which fuzzy. Fuzzy C-Means (FCM) is a method of clustering which existence of each data point in a cluster is determined by degree of fuzzy member (Havens, Bezdek, Leckie, Hall, & Palaniswami, 2012). So, the clustering of food security or food insecurity area in South Sumatera will be done by Fuzzy C-Means method.

DATA PREPROCESSING OF FOOD SECURITY

The first stage is dataset making which is collecting data. Data collection begins with analyzing the variables which affecting food security. From some paper which being analyzed, obtained 9 variables which declared the most influential to calculate food security in an area (Masitoh & Ratnasari, 2016), with variable data output are 0 and 1 which indicate that area is food insecurity and food

security. In this research, the number of outputs will be clustered based on value of the index data. From that data index value will figure out, the number of food security cluster in South Sumatera.

Food security data processed from macro and micro data of BPS in 2011-2017 from 3238 villages in 13 districts and 4 cities in South Sumatera. After the data is collected then go into early process until data group are ready to be classified. There are some steps which will be conduct in initial processing, which are Handles missing data due to the presence of regional expansion yet in that year, which overcome by replacing them with data on the early years of the district. Solving Outlier is by limiting the input value with upper threshold and lower threshold average of values in each variable. Based on initial dataset which already formed, it created covariant matrix to see how big the relation between variables. Variables which have negative connection with all other variable considered unrelated which mean will not be used.

Based on covariant matrix in figure 1, all attributes have connection with other attributes, even though there are attributes which have negative relationship with several attributes, but there are some attributes that remained positive. As an example percentage of poor population, malnutrition toddler and toddler's life expectancy has many negative relation with other attribute, but there are still some relation that have positive values, such as percentage of poor population by ratio of serelia consumption and the percentage of female illiteracy, malnutrition toddler by ratio of serelia consumption, percentage of female illiteracy and distance of village < 5 km from clinics, and toddler's life expectancy with less adequate road conditions, Percentage of households without access to electricity, so all attributes become factors which affecting the food security in South Sumatera. So, from that 9 variables, the amount of data which obtained are 119 data, which is stored in DKP dataset (food security data).

Attribut...	Rasio K...	Persent...	Kondisi ...	Persent...	Persent...	Persent...	Jarak D...	Balita K...	Angka ...
Rasio K...	1	0.158	0.429	0.625	0.332	0.805	0.621	0.014	-0.232
Persenta...	0.158	1	-0.040	-0.064	-0.077	0.086	-0.008	-0.115	-0.404
Kondisi ...	0.429	-0.040	1	0.233	0.279	0.232	0.414	-0.335	0.085
Persenta...	0.625	-0.064	0.233	1	0.299	0.527	0.513	-0.053	0.003
Persenta...	0.332	-0.077	0.279	0.299	1	0.227	0.500	-0.005	-0.095
Persenta...	0.805	0.086	0.232	0.527	0.227	1	0.517	0.025	-0.290
Jarak De...	0.621	-0.008	0.414	0.513	0.500	0.517	1	0.197	-0.169
Balita Ku...	0.014	-0.115	-0.335	-0.053	-0.005	0.025	0.197	1	-0.116
Angka H...	-0.232	-0.404	0.085	0.003	-0.095	-0.290	-0.169	-0.116	1

Figure 1. Covariant Matrix

Fundamental theory

Fuzzy C-Means (FCM) firstly was developed by Dunn in 1973 than it developed again by Bezdek in 1981. The basic idea of this method is similar to K-Means method. FCM based on fuzzy logic, each data point was put into a cluster based on the value of its member in cluster (Gopal & Karnan, 2010). This following is step of the Fuzzy C-Means (Tsai & Lin, 2011):

1. The input data which will be clustered, X in form of matrix of size n x m (n = total data, m = attribute data) and specify the parameters which are involved: c = total clusters; w = rank or index; *MaxIter* = maximum iteration; ε = Error minimum expected; P0 = 0 = initial objective Function; t = 1 = early Iterations.
2. Rise random numbers as elements of the initial partition of matrix U with Xi is the data to-i. The number of column elements each value in a row is 1;

$$U = \begin{bmatrix} \mu_{11}(X_1) & \mu_{21}(X_1) \dots & \mu_{c1}(X_1) \\ \mu_{21}(X_2) & \mu_{22}(X_2) \dots & \mu_{c2}(X_2) \\ \vdots & \vdots & \vdots \\ \mu_{1i}(X_i) & \mu_{2i}(X_i) \dots & \mu_{ci}(X_i) \end{bmatrix} \quad (1)$$

3. Calculate the center of each cluster (Vkj), for that partition matrix with k = 1, 2, ..., c; and j = 1, 2, ..., m;

$$V_{kj} = \frac{\sum_{i=1}^{n}((\mu_{ik})^w \,.\, X_{ij})}{\sum_{i=1}^{n}((\mu_{ik})^2} \quad (2)$$

Where V = Vector from center cluster; X = Data; n = total data; μ = degree of membership

$$P_t = \sum_{i=k}^{n}\sum_{k=1}^{c}([\sum_{j=1}^{m}\left(\left(x_{ij\_}v_{kj}\right)^2\right](\mu_{ik})^w) \quad (3)$$

4. Calculate objective function on iteration -t;
5. Calculate degree of membership;

$$\mu_{ik} = \frac{[\sum_{j=1}^{m}(x_{ij} - v_{kj})^2]^{-\frac{1}{w}-1}}{\sum_{k=1}^{c}[\sum_{j=1}^{m}x_{ij} - v_{kj})^2]^{-\frac{1}{w}-1}} \quad (4)$$

6. Check stop condition

- If ($P_t - P_{t-1} < \varepsilon$) or (t > *MaxIter*) the stop
- Else : t = t + 1, go to step-4

On data of food security case, there are 119 data and 9 variables then matrix X in form of matrix 119x9. Example parameter which included: c = 2; w = 2; *Maxlter* = 100; Next step is raising random number as elements of initial partition U matrix: ε = 0.05; P0 = 0; t=1. Then raising random number as elements of initial partition U matrix is initial below.

$$U = \begin{bmatrix} \mu_{11}(X_1) & \mu_{21}(X_1) \\ \mu_{21}(X_2) & \mu_{22}(X_2) \\ \vdots & \vdots \\ \mu_{119}(X_{119}) & \mu_{119}(X_{119}) \end{bmatrix} \quad (5)$$

After obtained the value, calculate each cluster Center (V_{kj}), for that partition matrix with k = 1.2; and j = 1 .. 9; so it can be calculate each cluster centers of its variable

$$V_{11} = \frac{(\mu_{11})^2 \,.\, X_{11} + (\mu_{21})^2 \,.\, X_{21} + \dots + (\mu_1}{(\mu_{11})^2 + (\mu_{21})^2 + \dots + (\mu_{119}}$$

$$V_{21} = \frac{(\mu_{12})^2 . X_{11} + (\mu_{22})^2 . X_{21} + \cdots + (\mu_{\mathbf{1}}}{(\mu_{11})^2 + (\mu_{21})^2 + \cdots + (\mu_{119}} \quad (6)$$

$$\vdots$$

$$\vdots$$

$$V_{19} = \frac{(\mu_{11})^2 . X_{19} + (\mu_{21})^2 . X_{29} + \cdots + (\mu_{\mathbf{1}}}{(\mu_{11})^2 + (\mu_{21})^2 + \cdots + (\mu_{119}$$

$$V_{29} = \frac{(\mu_{12})^2 . X_{19} + (\mu_{22})^2 . X_{29} + \cdots + (\mu_{\mathbf{1}}}{(\mu_{11})^2 + (\mu_{21})^2 + \cdots + (\mu_{119}$$

The next step is to calculate objective function at iteration 1, where its objective function is the total square of the difference in overall data with center cluster multiplied by the square of fuzzy weight from membership function each data.

$$P_1 = \sum_{i=1}^{119} \sum_{k=1}^{2} ([\sum_{j=1}^{9} \left(\left(x_{ij} - v_{kj} \right)^2 \right] (\mu_{ik})^2) \quad (7)$$

After the objective function is obtained, new membership function values can be calculated. Then checked back if it meets stop conditions, then process will stop, if not, then it will repeat the process of calculating center cluster. Process would be repeated until stop condition is met and cluster was formed.

In this research cluster will be grouped by Davies Bouldin Index values (Kovács, Legány, & Babos, 2005)et al. 2005). Davies Bouldin Index (DBI) is a method of cluster validation for quantitative evaluation from results of clustering (Xiao, Lu et al. 2017). This measurement aims to maximize inter-cluster distance between one cluster with other clusters. DBI used for outlier detecting on each cluster which formed.

$$DBI = \frac{1}{n} x \sum_{c_i \in C} \max_{c_i \neq c_j} \left(\frac{\sigma_i + \sigma_j}{d(\bar{c_i}, \bar{c_j})} \right) \quad (8)$$

Where n is the number of clusters, c_i, c_j is the cluster to *i* and *j*, *d* is the average distance from cluster, σ_i , σ_j is variance of cluster to *i* and *j*. In the earlier research, food security cluster of districts into 2 which are food security and food insecurity. In this research, cluster which will be tested on cluster 2, 3, 4 and 5 then it will be seen the value it's DBI. The value of the smallest DBI indicates that those data are very close between groups. So, it could be conclude that the cluster is the most suitable data with food security in South Sumatera.

RESULTS AND ANALYSIS

This research was conducted on 2 experimental testing scenario which use a weight value (m) which different to get most optimal DBI value. Stopping criteria which used is maximum iterations value by 500 times and average value of error (Error Means Square) at least was 0.05.

The value of DBI show the proximity data Center cluster. The smaller of DBI value that means the closer distance that data into the center of cluster. This means that data value of the smallest DBI is considered the best cluster for that data. From the table above obtained that, smallest DBI value on clusters = 3.

From result of the experiment were obtained that on m is 3 for total cluster in 5 processes stop on 22nd – iteration because it has reached the value of error = 0.0361, on total cluster = 4 process stop on 37th – iteration with the stopping error is 0.0408, on total clusters = 3 process stop on 40th – iteration cause it has reached the value of error on 0.0382, and on total cluster = 2 process stops at 40th – iteration because it has reached the value of error = 0.0367. From the results of experiment above, also summed up the large number of repetitions of the process is only slightly affected at speeds reaching stop condition (stopping criteria) and not too affected the DBI value.

Tabel 1. The Comparison of DBI Result

m	Cluster	R	DBI
2	5	1,117.396	223.479
	4	1,129.015	282.254
	3	413.951	137.984
	2	2,834.745	1,417.372
3	5	30.163528	6.0327056
	4	478.8437	119.71092
	3	17.975439	5.9918131
	2	27005.462	13502.731

From the table above was obtained, smallest DBI value of the clusters exist at cluster = 3. So it can be summarize that the food security data clustering in South Sumatera is more appropriate to be cluster into 3 groups, which are: Food security Area, food enough area, food insecurity area. Figure 2 shows that the area which include in food security group is Palembang and Pali only. Palembang city entered this category from 2011 until 2017. For the latest data

which on 2017, area which entered cluster of food security only 2 area, clusters of food are 10 regions and clusters of food insecurity are 5 area.

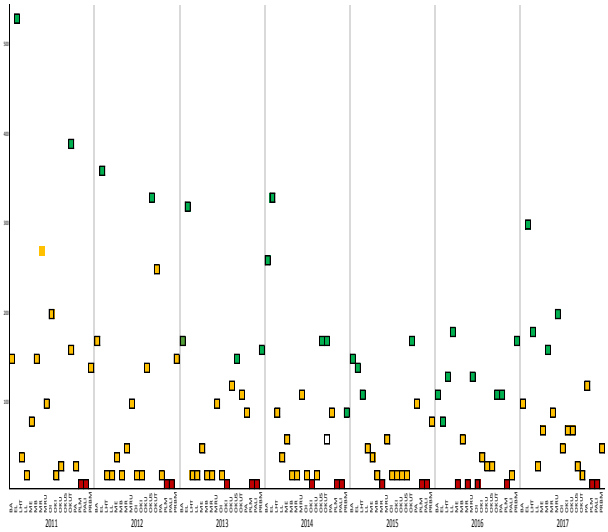

Figure 2. Data Cluster Food Security in South Sumatera

SUMMARY

Based on the research which conducted, then the contribution of this research are as follows: Criteria of food security in South Sumatera is consist of 9 variables which taken from the whole area which consists of 13 districts and 4 cities which spread in 3238 village. From the results of the data processing, preprocess data and calculation of covariant matrix, obtained food security data are 119 data which consisting of 9 variable which describe the condition of 17 areas in South Sumatera from 2011 to 2017.

Clustering which are processed using fuzzy C-Means which testing conducted by differentiating the weight exponent value which control the fuzziness of clusters which produced, obtained that data of food security in South Sumatera according to cluster into three clusters which based on food security characteristics of those three clusters are food security cluster, enough food and food insecurity.

To improve the performance of the process of clustering from Fuzzy C-Means, then there must be mechanism to overcome the initial values which are random. Swarm intelligence such as ant colony or particle swarm, bee colony may be as a solution. This method is optimization method of stochastic population-based, which enables mobile population in search space to determine the best value globally, so that it is possible to search values which are random.

REFERENCES

[1] Gopal, N. N., & Karnan, M. (2010). Diagnose brain tumor through MRI using image processing clustering algorithms such as Fuzzy C Means along with intelligent optimization techniques. Paper presented at the Computational Intelligence and Computing Research (ICCIC), 2010 IEEE International Conference on.

[2] Havens, T. C., Bezdek, J. C., Leckie, C., Hall, L. O., & Palaniswami, M. (2012). Fuzzy c-means algorithms for very large data. IEEE Transactions on Fuzzy Systems, 20(6), 1130-1146.

[3] Kovács, F., Legány, C., & Babos, A. (2005). Cluster validity measurement techniques. Paper presented at the 6th International symposium of hungarian researchers on computational intelligence.

[4] Masitoh, F., & Ratnasari, V. (2016). Pemodelan Status Ketahanan Pangan di Provinsi Jawa Timur dengan Pendekatan Metode Regresi Probit Biner. Jurnal Sains dan Seni ITS, 5(2).

[5] Permatasari, D. L., & Ratnasari, V. (2016). Pemodelan Ketahanan Pangan di Indonesia dengan Pendekatan Regresi Probit Ordinal. Jurnal Sains dan Seni ITS, 5(2).

[6] Raharto, S. (2010). Pemetaan Ketahanan Pangan Regional di Jawa Timur. JSEP (Journal of Social and Agricultural Economics), 4(3), 1-11.

[7] Rianti, T. S. M. (2016). Analisis Ketahanan Pangan Tingkat Desa di Kecamatan Purwoasri, Kecamatan Plemahan dan Kecamatan Mojo Kab. Kediri, Jawa Timur. Paper presented at the Seminar Nasional Pembangunan Pertanian Malang.

[8] Syamsiah, N. O. (2011). Sistem Klasifikasi Indikator Daerah Rawan Pangan Menggunakan Database Fuzzy Tahani. PARADIGMA, 13(2).

[9] Tsai, D.-M., & Lin, C.-C. (2011). Fuzzy C-means based clustering for linearly and nonlinearly separable data. Pattern Recognition, 44(8), 1750-1760.

[10] Wulandari, K., & Fauzy, A. (2016). Implementasi Self Organizing Maps untuk Clustering Ketahanan dan Kerentanan Pangan Desa di Kabupaten Magetan 2014. Paper presented at the Seminar Nasional Pendidikan Matematika 2016, Universitas Kanjuruhan Malang.