Classification of provinces in Indonesia based on 2014 presidential election participation levels using fuzzy c-means

A Yuswatun¹ and A M Abadi²
Mathematics Study Program, Faculty of Mathematics and Natural Sciences, Yogyakarta State University, Yogyakarta, Indonesia

¹anna.yuswatun2015@student.uny.ac.id

Abstract. The presidential and vice presidential elections held every five years is a benchmark for democracy in Indonesia. However, the expected level of participation cannot be achieved. Election socialization as an effort to increase the level of participation in each province is different because the level of participation in each province is also varied. Therefore, there is a need for regional grouping based on the level of electoral participation in order to facilitate the electoral socialization process. This study uses fuzzy c-means with two attributes, namely the number of voters and the number of voting rights users. All provinces in Indonesia are classified into two clusters. If the provincial participation rate is below the national participation standard (69.58%) then the province is considered to have a low participation rate and vice versa. Therefore, the first cluster constitutes provinces with high electoral participation rate and the second cluster constitutes provinces with low electoral participation rate. The results of the classification are 9 provinces in first cluster and 24 provinces in the second cluster.

1. Introduction
Indonesia has many provinces and in terms of sovereignty, it “is in the hands of the people and is implemented according to the Constitution” [1]. This statement in the 1945 Constitution confirms that Indonesia is a democratic country, from the people by the people and to the people, including in the election of the president and vice president, in which citizens elect directly in the election. Article 6A of the 1945 Constitution states that the President and Vice President shall be elected as a pair by the people. General elections shall be conducted in a direct, general, free, secret, honest and fair every five years [1], so elections are usually called democratic parties because not every year happens.

Elections that are routinely carried out certainly produce data that can be processed mathematically. For example, it is predicting election results and predicting presidential candidates to win or lose. Research conducted by Jiao using fuzzy adaptive network shows that fuzzy can predict the results of the presidential election ideally [2]. Whereas Singh forms a model to calculate the percentage of win or lose from a candidate or party in the election using fuzzy [3]. In addition to the election results, the level of election participation also received attention. Voter participation is one indicator of the results of a general election with a strong degree of legitimacy. Participation rate is often a reference for election organizers whether the elections they hold have a strong appeal to citizens to be involved. Not surprisingly, General Election Commission (KPU) as the election organizer monitors the participation rate which will always be attempted to increase [4].
KPU commonly targets national participation rates. In 2014 KPU targeted 75% for election participation rates of the President and Vice President. However, voter participation in the 2014 presidential election missed the KPU's goal. Moreover, voter participation also declined compared to 2009 elections and April 2014 legislative elections. Voter participation in the 2014 presidential election based on data reported by KPU was only 69.58%. Generally the causes of decreasing participation differ in each region so that the treatment to increase the participation level of each region also varies. Therefore, grouping regions based on the level of electoral participation is deemed necessary and it is expected that the election socialization in the future could be carried out more easily. In this paper the Indonesian provinces will be divided into two clusters, with two attributes or variables, namely number of voters and number of voting right users. Provinces with a participation rate below the national participation standard of 69.58% are considered to have a low participation rate. Meanwhile, the provinces with a participation rate above the national participation level are considered to have a high level of participation. Fuzzy C-Means (FCM) is greater than SOM neural networks, K means and traditional hierarchical clustering algorithms in clustering. FCM shows stability even though there are outliers and overlapping data [5]. Clustering using FCM was used to classify areas in West Java based on economic potential. The study produced nine clusters based on the most potential economic in each region [6]. Similar to the study, this study will cluster provinces using FCM. This study uses FCM, which is commonly used for grouping data. Based on the background above, this grouping needs to be carried out because the level of voter attendance as a form of election participation becomes very important. It relates to the level of legitimacy and the future of democracy. The low level of voter attendance in the election can pose a risk for a country’s democracy [7].

2. Methodology
The data processing use rules in clustering with FCM. The data were collected from the results of the 2014 presidential and vice presidential elections on the General Election Commission web page (kpu.go.id). The variables used in this study are variables that represent the level of regional participation. These variables are:
V1: number of voters,
V2: number of voting right users.
After determining the variable, data were processed using FCM.

2.1. Fuzzy C Means
Fuzzy Clustering is one technique for determining optimal clusters in a vector space based on the Euclidian normal form for distances between vectors. There are several data clustering algorithms; one of which is FCM. FCM is a data clustering technique where the existence of each data point of a cluster is determined by the degree of membership. This technique was first introduced by Jim Bezdek in 1981. The first basic concept of FCM is to determine the center of the clusters, which marks the average location for each cluster. In the initial conditions, the cluster center is still not accurate. Each data point has a degree of membership for each cluster. The second basic concept of FCM comprises fixing the cluster center and the degree of membership of each data point repeatedly. It can be seen that the center of the cluster will move towards the right location. This looping is based on the minimization of the objective function which describes the distance from the data points to the cluster center weighted by the degree of membership of the data points [8]. Membership degrees between zero and one are used in fuzzy clustering instead of crisp assignments of the data to clusters. The resulting data partition improves data understanding and reveals its internal structure. Partition clustering algorithms divide a data set into clusters or classes, where similar data objects are assigned to the same cluster, whereas dissimilar data objects should belong to different clusters[9].
The FCM algorithm is as follows [8]:
1) Input the data in cluster X, in the form of a matrix measuring $n \times m$ ($n =$ number of data samples, $m =$attribute of each data). $X_{ij}$ = sample data to-i (i=1,2,3,...,n), attributed to j(j=1,2,...,m)

2) Specify
- Number of cluster (c)
- Degree (w)
- Maximum iteration
- The smallest error expected ($\xi$)
- First objective function ($P_0 = 0$)
- First Iteration (t=1)

3) Generate the random numbers $\mu_{ik}, i = 1,2,...,n; k = 1,2,...,c$; as elements of the initial partition matrix $U$.

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

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

5) Calculate objective function in iteration to-t, $P_t$:

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

6) Calculate the partition matrix changes

$$\mu_{ik} = \frac{\sum_{j=1}^{m}(X_{ij} - V_{kj})^2}{\sum_{k=1}^{c}(\mu_{ik})^w}$$

7) Check the stop condition
- If ($|P_t - P_{t-1}| < \xi$) or (t>MaxIter) then stop;
- If no, t=t+1, repeat from step 4.

3. Results and Discussion
Since its post-reform election history, Indonesia has held elections four times, including the last 2014 Election. If compared with the 1999 election when the president is still elected by the MPR, voter participation in the legislative elections in 2014 is very high. Voter participation in the 1999 election reached 97.7%. This means that only 2.3% of voters did not vote. A number of voters registered in the 1999 election were 117,815,953 people. However, after the 1999 elections, the tendency of voter participation rates has declined [2]. The election participation rate is calculated by the following formula:

$$\frac{V_0}{N_r} \times 100\%$$

$N_r$ : number of voters
$V_0$ : number of voters who give voting rights

Based on this formula, the level of community participation in the 2014 presidential and vice presidential election was 69.58%. The number of voters is the sum of the Final Voters List (DPT), Additional Voters List (DPTb), Voters registered in the Special Voters List (DPK) and Additional Special Voters List (DPKTb). Table 1 demonstrates results of the total votes from each province in the 2014 presidential and vice presidential elections. The data were collected from www.kpu.go.id.
Table 1. Data resulting from vote counting per province

| Province       | Number of voters | Number of voters who give voting rights | Percentage of voting rights |
|----------------|------------------|----------------------------------------|-----------------------------|
| Aceh           | 3,357,159        | 2,061,084                              | 61%                         |
| Sumatra Utara  | 10,129,891       | 6,356,025                              | 63%                         |
| Sumatra Barat  | 3,693,822        | 2,354,327                              | 64%                         |
| Riau           | 4,319,920        | 2,709,778                              | 63%                         |
| Jambi          | 2,525,649        | 1,782,570                              | 71%                         |
| Sumatra Selatan| 5,941,085        | 4,190,631                              | 71%                         |
| Bengkulu       | 1,396,279        | 963,992                                | 69%                         |
| Lampung        | 6,070,978        | 4,360,992                              | 72%                         |
| Bangka Belitung| 943,944          | 619,174                                | 66%                         |
| Kepulauan Riau | 1,396,550        | 830,025                                | 59%                         |
| DKI Jakarta    | 7,523,101        | 5,441,705                              | 72%                         |
| Jawa Barat     | 33,821,378       | 23,990,089                             | 71%                         |
| Jawa Tengah    | 27,606,063       | 19,668,404                             | 71%                         |
| DIY            | 2,812,144        | 2,245,164                              | 80%                         |
| Jawa Timur     | 30,933,642       | 22,184,407                             | 72%                         |
| Banten         | 8,230,615        | 5,651,467                              | 69%                         |
| Bali           | 2,992,122        | 2,167,221                              | 72%                         |
| NTB            | 3,579,559        | 2,569,997                              | 72%                         |
| NTT            | 3,237,432        | 2,274,079                              | 70%                         |
| Kalimantan Barat| 3,560,852    | 2,621,933                              | 74%                         |
| Kalimantan Tengah| 1,880,910   | 1,173,833                              | 62%                         |
| Kalimantan Selatan| 2,888,127  | 1,919,794                              | 66%                         |
| Kalimantan Timur| 3,023,405   | 1,893,441                              | 63%                         |
| Sulawesi Utara | 1,934,354        | 1,349,868                              | 70%                         |
| Sulawesi Tengah| 1,985,135        | 1,407,759                              | 71%                         |
| Sulawesi Selatan| 6,426,837   | 4,274,615                              | 67%                         |
| Sulawesi Tenggara| 1,827,083  | 1,139,678                              | 62%                         |
| Gorontalo      | 803,465          | 603,448                                | 75%                         |
| Sulawesi Barat | 902,061          | 624,943                                | 69%                         |
| Maluku         | 1,238,067        | 881,448                                | 71%                         |
| Maluku Utara   | 859,717          | 565,970                                | 66%                         |
| Papua          | 3,270,840        | 2,833,245                              | 87%                         |
| Papua Barat    | 730,426          | 536,270                                | 73%                         |

The next step for clustering is determining number of cluster, degree, maximum iteration, the smallest error expected, first objective function, and first iteration. So, there are two clusters, thus the degree used is two and the results of the cluster center calculation, first objective function, and matrix partition changes are not too large. The smallest expected error is $10^{-5}$ and the desired maximum iteration is 5.

- Number of cluster (c) = 2
- Degree (w) = 2
- Maximum iterations = 5
- The smallest error expected ($\xi$) = $10^{-5}$
- First objective function ($P_0$) = 0
First iteration (t) = 1
Based on the above data, it was obtained a U random matrix formed from random numbers $\mu_{ik}$.

\[
U = \begin{bmatrix}
0.058 & 0.091 & 0.116 & 0.091 & 0.282 & 0.281 & 0.245 & 0.313 & 0.161 & 0.011 & 0.325 & 0.291 & 0.299 & 0.508 & 0.310 \\
0.942 & 0.909 & 0.884 & 0.909 & 0.718 & 0.719 & 0.755 & 0.687 & 0.839 & 0.989 & 0.675 & 0.709 & 0.701 & 0.492 & 0.690 \\
0.236 & 0.328 & 0.312 & 0.274 & 0.357 & 0.083 & 0.182 & 0.088 & 0.263 & 0.291 & 0.183 & 0.082 & 0.393 & 0.251 & 0.297 \\
0.764 & 0.672 & 0.688 & 0.726 & 0.643 & 0.917 & 0.818 & 0.912 & 0.737 & 0.709 & 0.817 & 0.918 & 0.607 & 0.749 & 0.703 \\
\end{bmatrix}
\]

0.167 0.674 0.352
0.833 0.326 0.648

Afterwards, the center of cluster and objective function was observed using U Matrix random.

**Table 2.** The results of the center of cluster calculations and objective functions in the 1st and 2nd iterations

| Iteration to- | Cluster center ($V_{kj}$) | Objective Function ($P_t$) |
|----------------|---------------------------|---------------------------|
| 1 | 5,999,763.11 4,377,335.71 | $P_1 = 2,010,749,480,634,760$ |
| 2 | 6,985,077.97 4,890,074.67 | $P_2 = 1,666,544,569,392,200$ |

In the first iteration, the condition cannot stop due to the following calculation.

\[|P_t - P_{t-1}| = |2,010,749,480,634,760 - 0| = 2,010,749,480,634,760 > \xi\]

Therefore, it goes on to the next iteration. In the second iteration of the first cluster center the first attribute is 6,985,077.97 and the second attribute is 4,890,074.67. Whereas in the second cluster, the first attribute cluster center is 4,726,856.46 and the second attribute is 3,304,538.84. Hence, the new U matrix is formed,

\[
U = \begin{bmatrix}
0.14 & 0.76 & 0.10 & 0.04 & 0.20 & 0.59 & 0.26 & 0.72 & 0.28 & 0.27 & 0.95 & 0.54 & 0.55 & 0.16 & 0.54 \\
0.86 & 0.24 & 0.90 & 0.96 & 0.80 & 0.41 & 0.74 & 0.72 & 0.73 & 0.05 & 0.46 & 0.45 & 0.84 & 0.46 \\
0.89 & 0.16 & 0.10 & 0.14 & 0.10 & 0.24 & 0.17 & 0.17 & 0.23 & 0.23 & 0.85 & 0.24 & 0.29 & 0.28 & 0.27 & 0.29 & 0.11 & 0.29 & 0.29 \\
0.11 & 0.84 & 0.90 & 0.86 & 0.90 & 0.76 & 0.83 & 0.83 & 0.77 & 0.77 & 0.15 & 0.76 & 0.71 & 0.72 & 0.73 & 0.71 & 0.09 & 0.71 & 0.71 \\
\end{bmatrix}
\]

Checking the stop condition $t = 2$

\[|P_t - P_{t-1}| = |1,666,544,569,392,200 - 2,010,749,480,634,760| < \xi\]

\[|P_t - P_{t-1}| = |1,666,544,569,392,200 - 2,010,749,480,634,760| < 10^{-5}\]

Since it meets the requirements of $|P_2 - P_1| < \xi$ the iteration stops. The clustering results table is presented according to the new U matrix. Cluster determination is based on the largest value of the U matrix. For example, The Province of Aceh has membership values on matrix U 0.14 and 0.86. The value of 0.14 is the value of Aceh’s membership in the first cluster and 0.86 is membership value in the second cluster. As the membership value in second cluster is bigger, so The Province of Aceh goes into second cluster.

**Table 3.** Clustering Results

| Province     | $\mu_{i1}$ | $\mu_{i2}$ | 1st Cluster | 2nd Cluster |
|--------------|------------|------------|-------------|-------------|
| Aceh         | 0.14       | 0.86       | *           |             |
| Sumatra Utara| 0.76       | 0.24       | *           |             |
| Sumatra Barat| 0.10       | 0.90       | *           |             |
From the calculation results according to the algorithm, the first cluster (provinces of high electoral participation rate) consists of the provinces of North Sumatra, South Sumatra, Lampung, DKI Jakarta, West Java, Central Java, East Java, Banten and South Sulawesi. The second cluster (provinces of low electoral participation rate) consists of 24 other provinces, namely Aceh Province, West Sumatra, Riau, Jambi, Bengkulu, Bangka Belitung, Riau Islands, DIY, Bali, NTB, NTT, West Kalimantan, Southeast Kalimantan, South Kalimantan, East Kalimantan, North Sulawesi, Southeast Sulawesi, Central Sulawesi, Gorontalo, West Sulawesi, Maluku, North Maluku, Papua and West Papua.

4. Conclusion
Clustering with the Fuzzy C-Means method with two iterations has fulfilled the stop condition according to the algorithm, namely \((|P_2 - P_1| < \xi < 10^{-5})\). Provinces are classified into 2 clusters. The first cluster consists of the provinces with high electoral participation rate and the second cluster consists of the provinces with low electoral participation rate. This clustering method results in 9
provinces classified into the first cluster, and 24 provinces categorised into the second cluster. The first cluster comprises the Province of North Sumatra, South Sumatra, Lampung, DKI Jakarta, West Java, Central Java, East Java, Banten and South Sulawesi. The second cluster comprises 24 other provinces, namely Aceh Province, West Sumatra, Riau, Jambi, Bengkulu, Bangka Belitung, Riau Islands, DIY, Bali, NTB, NTT, West Kalimantan, Southeast Kalimantan, South Kalimantan, East Kalimantan, North Sulawesi, Southeast Sulawesi, Central Sulawesi, Gorontalo, West Kalimantan, Maluku, North Maluku, Papua and West Papua.

The center of the first attribute (number of voters) within the first cluster is 6,985,077.97 and the second attribute (number of voting right users) is 4,890,074.67. Meanwhile, in the second cluster, the first attribute is 4,726,856.46 and the second attribute is 3,304,538.84.

References
[1] 1945 Constitution of Republic Indonesia Amendment of 2002.
[2] Jiao Y, Syau Y and Lee E S 2006 Fuzzy adaptive network in presidential elections Mathematical and Computer Modelling 43 244-253.
[3] Singh H, Singh G and Bhatia N 2012 Election results prediction system based on fuzzy logic International journal of computer application 53 30-39.
[4] Ramadhanil F, Junaidi V and Ibromhim 2015 Community Partitioning Design in Election Monitoring (Jakarta: Kemitraan bagi Pembaruan Tata Pemerintahan di Indonesia and Perludem).
[5] Nurshasim M 2014 Voter Participation in 2014 Election: Scoping Study (Jakarta: Indonesian Institute of Science).
[6] Mingoti S A and Lima J O 2006 Comparing SOM Neural Network with Fuzzy C-Means, K-Means, and Traditional Hierarchical Clustering Algorithms European Journal of Operational Research 174 1742-1759.
[7] Rahayu S P, Febryani S, Kusuma H B, Suhartono and Prasetyo D D 2018 The approach of fuzzy C-means cluster and factor analysis on economic potential mapping of regency/city in East Java Province International Conference on Science and Applied Science (ICSAS): AIP Conference Proceedings 2014 020082-1-020082-8.
[8] Kusumadewi S and Purnomo H 2004 Fuzzy Logic Application for Decision Making (Yogyakarta: Graha Ilmu).
[9] Sowmya B and Sheelarani B 2011 Land cover classification using reformed fuzzy c-means Sadhana 36 153-165.