

# The Classification of Districts in East Java Based on Results of Monitoring Family Planning Using K-Modes Clustering Approach

*by* Fitria Rakhmawati

---

**Submission date:** 15-Feb-2018 02:23PM (UTC+0800)

**Submission ID:** 916339510

**File name:** J.\_Basic.\_Appl.\_Sci.\_Res.,\_6\_8\_1-6,\_2016.pdf (195.91K)

**Word count:** 2978

**Character count:** 15394



## The Classification of Districts in East Java Based on Results of Monitoring Family Planning Using K-Modes Clustering Approach

Fitria Rakhmawati<sup>1\*</sup>, Windu Purnomo<sup>1</sup>, Soenarnatalina<sup>1</sup>, Bambang Widjanarko Otok<sup>2</sup>

<sup>15</sup> <sup>26</sup>  
<sup>1</sup> Faculty of Public Health, Airlangga University in Surabaya (Indonesia)  
<sup>2</sup> Laboratory of Environmental and Health Statistic, Sepuluh Nopember Institute of Technology (ITS) Surabaya (Indonesia)

Received: April 13, 2016  
Accepted: July 26, 2016

### ABSTRACT

Mini survey BKKBN is a national scale survey conducted periodically every year since 2003. The classification of data will facilitate the monitoring of more focus on the view of the characteristics of a data. The implementation of the method k-modes clustering in the classification of data based on the results of the survey mini family planning (KB) monitoring as age, the number of children and education is kategorikal data. The results of the study showed that KB participation there is a relationship with the level of education, the number of children and age marry first. The classification of District based on the four variables become 2, 3, 4 5 cluster provides the value of the purity of each of 67.0 percent, 61.1 percent, 63.3 percent and 54.7 percent. The classification District/City based on the results of the survey mini with II-Cluster with the level of highest accuracy and have are different characteristics in Kediri District, Banyuwangi District, Bondowoso District, Sidoarjo District, Mojokerto District, Madiun District, Magetan District, Sampang District, Kediri City, Blitar City, Malang city, Mojokerto City, City Madiun, City Surabaya, City Batu. Group-1 characterized by education graduate Senior High, the number of 2 children and age marry first 20-25 years, whereas groups-2 characterized by education graduate Junior High, the number of children 1 children and age marry first 15 - 20 years.

**KEYWORDS:** clustering, K-modes, mini monitor survey, family planning

### 1. INTRODUCTION

The results of the monitoring KB mini survey in East Java still is an individual and has not done grouping. Mini survey data 2014 contain not only the data but also the continuous scale data categories namely membership KB, education level, number of children, so the cluster analysis will be done to make the grouping should lead to cluster analysis for categorical data [1]. According to [2] says there are several factors that affect the use of birth control include age, number of children, education, and knowledge. Classification parity according to [3] and [4] is nullipara, primiparity, multipara and grandemultipara.

Age plays a role in patterns of contraceptive services to the public with regard to a healthy reproductive period, in which the woman has several phases of which the phase space pregnancy, delay or prevent pregnancy and terminating the pregnancy [5]. The number of children a woman can affect whether or not a birth control method suitable medically. Mothers who already have two or more children should terminate fertility and it is advisable not to have any more children because of medical reasons and other reasons [4]. A person's education is crucial in the decisionmaking patterns and receive information than those with low education. Education is one of the factors that determine a person's knowledge and perception of the importance of a case, including the importance of participation in family planning [6][7].

Clustering techniques currently has many used to resolve issues related to the segmentation data. The implementation of clustering methods can be applied in various fields. The main purpose of the cluster method is the grouping of a number of data or objects into the cluster so that in each cluster will contain the data as closely [8]. The system on the method of clustering placing objects that are similar in one cluster and make the distance between the cluster which as far as possible [9]. This means that the objects in a cluster are very similar to each other and different from the objects in the cluster of the other clusters [10][11].

K-means of cluster method used to analyze data is continuously but for data categorical k-means method of cannot be used. According to [12] analysis k-means of cluster for data categorical by introducing numeric encoding scheme on the categorical data called dummy variable. The election code that vain will be fatal in the calculation of the distance between categorical data to expand the algorithm k-means cluster, especially to expand the analysis of the group on the categorical data [13]. The difficulty will be faced if faced with a large number of categorical data is possible errors on metric euclidean space where data is positioned in the coordinates of the measured. This problem causes the use of euclidean distance yang usually used for continuous data on the analysis of the group does not applicative again [14][15].

32  
**Corresponding Author:** Fitria Rakhmawati, Faculty of Public Health, Airlangga University in Surabaya (Indonesia).  
E-mail: fna\_2007@yahoo.com

According to [16] asking algorithm which is updated, namely k-modes algorithm. This algorithm is used as a replacement for the mode from the mean to represent the data center. This algorithm is iterative, namely with how to renew the membership from each mode and data from the group on each iteration [8][12].

The algorithm K-modes is clustering algorithm used for type the categorical data. The main modification of k-means using the function of the distance of the cluster center and the process of iteration [16]. Mini survey data 2014 relating to the use of family planning among the membership KB, age, number of children and education as the basis for the grouping of districts/city in East Java with K - Modes Clustering approach. So the result of this grouping is expected to bring new information that can be used as a basis for policy making, especially regarding family planning program.

## 2. METHODOLOGY

The type of research used in this research is non reactive studies [17]. The data source is the secondary data BKKBN East Java Province mini results monitoring survey KB years 2014 as much as 2729 which is distributed in the district / city [1]. The variables used in this research are participation family planning (KB), level of education, number of children and first age marry. Data analysis techniques using the K-modes Clustering, namely grouping mini data monitoring survey KB based four of these variables. The framework of the concept as follows [1][2].

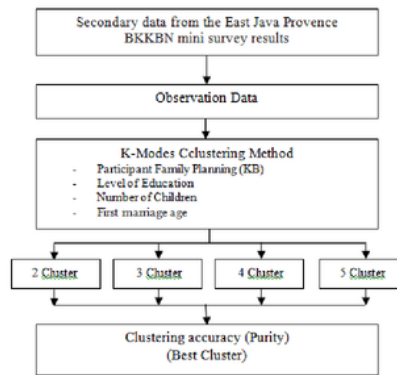


Figure 1. Clustering analysis frame with K-Modes

K - mode in Figure 1 can be explained by the following steps as follows [8]:

1. Initialize the cluster center as much as k cluster
2. Generate the initial centroid is randomly from the data available where many according to the cluster in the form (one point of the center in the form for each cluster) 31
3. Count the distance each data against all the focal point of the cluster. Allocate every object to the nearest cluster using the size of the distance dissimilarity.
4. Update cluster center using the frequency method.
5. Repeat steps 2 and 3 until no objects that switch cluster (convergence) for one time iteration.

Cluster pure said all objects with the same class in the same clusters. To measure the accuracy of clustering (purity) or 'r', measured using common [16]

$$r = \frac{1}{n} \sum_{i=1}^k a_i$$

with:

r : the level of accuracy of clustering

k : the number of cluster

$a_i$  : object that appears in the cluster  $C_i$  and on the label class the appropriate

## 3. RESULTS AND DISCUSSION

Research results include descriptive variables, dependency analysis and K-Modes. The description and analysis of tendency in detail is presented in the following table.

Table 1. The Description & Independency Test of the Participation Family Planning (KB) with independent variables

The frequency (Adjusted Residual) Pearson Chi-Square [18]		Participation Family Planning		
		Yes	Not	
Level of Education	Graduate Colloge	117 (-2.1)	66 (2.1)	
	Graduate Diploma	32 (0.8)	10 (-0.8)	
	Graduate Senior High	511 (-1.6)	234 (1.6)	
	Graduate Junior High	474 (3.6)	144 (-3.6)	
	Elementary School	562 (2.3)	196 (-2.3)	
	Not Elementary School	197 (-1.6)	97 (1.6)	
	Not Schools	42 (-5.0)	47 (5.0)	
	Pearson Chi-Square = 47.148 Asymptotic Sig. (2-sided) = 0.000			
	Number of children	Do not have children	13 (-17.3)	138 (17.3)
1 children		650 (-1.3)	288 (1.3)	
2 children		893 (8.7)	224 (-8.7)	
3 children		281 (1.4)	99 (-1.4)	
4 children		65 (-1.3)	35 (1.3)	
5 children		25 (0.9)	7 (-0.9)	
6 children		5 (-0.5)	3 (0.5)	
7 children		3 (1.1)	0 (-1.1)	
Pearson Chi-Square = 335.199 Asymptotic Sig. (2-sided) = 0.000				
First Marriage Age	< 15 year	82 (-0.1)	34 (0.1)	
	15- <20 year	820 (3.1)	285 (-3.1)	
	20- <25 year	793 (0.4)	319 (-0.4)	
	25- <30 year	200 (-3.8)	123 (3.8)	
	>= 30 year	40 (-3.1)	33 (3.1)	
	Pearson Chi-Square = 27.762 Asymptotic Sig. (2-sided) = 0.030			

Table 1 shows that KB participation there is a relationship with the level of education, the number of children and age marry first. It can be seen from the Asymptotic Sig. (2-sided) smaller than alpha 5 percent. Meanwhile, if viewed from the adjusted residual value is out at interval (-1.96; 1.96) then the cause of the relationship is participation family planning (KB) has a tendency on the level of education Graduate Colloge, Graduate Junior High, Elementary School and Not Schools and tend to have 2 children. This also happened in KB participation with age marry first 15 - < 20 years, 25 - < 30 years and more than 30 years.

The classification of District based on the results of the monitoring Mini Survey KB years 2014 using cluster 2, 3, 4, and 5 using the *software R* obtained the following result.



29 Table 2. Clustering results with the method K-Modes

The Number of observation The variables	II - Cluster	III - Cluster	IV - Cluster	V - Cluster
<b>Group - 1</b>	1755 Eligibility KB (yes) Education (Graduate Senior High) The number of children 2 Age marry first 20-25 year	1473 Eligibility KB (yes) Education (Elementary School) The number of children 1 Age marry first 15-20 year	1351 Eligibility KB (yes) Education (Graduate Senior High) The number of children 2 The first marriage age < 14 year	886 Eligibility KB (not) Education (Graduate Senior High) The number of children 2 Age marry first 20-25 year
<b>Group - 2</b>	974 Eligibility KB (yes) Education (Graduate Junior High) The number of children 1 Age marry first 15-20 year	752 Eligibility KB (not) Education (Graduate Senior High) The number of children 2 Age marry first 20-25 year	523 Eligibility KB (not) Education (Elementary School) The number of children 1 The age of the first marriage >=30 year	762 Eligibility KB (yes) Education (Graduate Junior High) The number of children 1 Age marry first 15-20 year
<b>Group - 3</b>		504 Eligibility KB (yes) Education (Graduate Junior High) The number of children 2 Age marry first 20-25 year	2087 Eligibility KB (yes) Education (Graduate Junior High) The number of children 1 The first marriage age < 14 year	616 Eligibility KB (yes) Education (Elementary School) The number of children 2 Age marry first 15-20 year
<b>Group - 4</b>			145 Eligibility KB (not) Education (Graduate Senior High) The number of children (not yet have) Age marry first 20-25 year	222 Eligibility KB (yes) Education (Graduate Junior High) The number of children 2 Age marry first 20-25 year
<b>Group - 5</b>				138 Eligibility KB (not) Education (Not Elementary School) The number of children (not yet have) Age marry first 15-20 year
<b>Purity (r)</b>	67.0 %	61.1 %	63.3 %	54.7 %

19

Table 2 shows that the clustering results based on the results of the survey mini Participatively monitoring KB to II cluster with the level of accuracy of 67.0 percent, there are 1755 people in group 1 with participation characteristics KB (yes), education (Graduate Senior High), the number of children 2 people, age marry first 20-25 years, and there are 974 people on the group 2 with participation characteristics KB (yes), education (Graduate Junior High), the number of children 2 people, age marry first 15-20 years. Cluster III with the level of accuracy of 61.1 percent, there are 1473 people in group 1 with participation characteristics KB (yes), education (Elementary School), the number of children 1 people, age marry first 15-20 years, there are 752 people on the group 2 with participation characteristics KB (not), education (Graduate Senior High), the number of children 2 people, age marry first 20-25 years, and there are 504 people in Group 3 with participation characteristics KB (yes), education (Graduate Junior High), the number of children 2 people, age marry first 20-25 years. Division IV cluster with the level of accuracy of 63.3 percent, there are 1341 people in group 1 with participation characteristics KB (yes), education (Graduate Senior High), the number of children 2 people, age marry first less than 14 years, there are 523 people on the group 2 with participation characteristics KB (not), education (Elementary School), the number of children 1 people, age marry first more than 30 years, there are 713 people in group 3 with participation characteristics KB (yes), education (Graduate Junior High), the number of children 1 people, age marry first less than 14 years, and there are 713 people in group 3 with participation characteristics KB (not), education (Graduate Senior High), have not had a child, age marry first 20 - 25 years. Cluster V with the level of accuracy of 54.7 percent.

Table 2 shows that the highest purity on the two cluster, seen based on the region of each District/City in East Java, in detail is presented in the following table:

Table 3. Many members of cluster in each District by the method of K-mode

District/City	Cluster		District/City	Cluster	
	I	II		I	II
Pacitan	32	18	Magetan	39	11
Ponorogo	41	26	Ngawi	45	30
Trenggalek	31	19	Bojonegoro	41	34
Tulungagung	51	24	Tuban	46	29
Blitar	27	23	Lamongan	38	37
Kediri	71	29	Gresik	47	28
Malang	66	34	Bangkalan	40	35
Lumajang	66	34	Sampang	52	23
Jember	51	49	Pamekasan	45	30
Banyuwangi	67	33	Sumenep	43	32
Bondowoso	33	42	Kediri City	60	15
Situbondo	43	32	Blitar City	44	5
Probolinggo	39	36	Malang City	62	7
Pasuruan	51	49	Probolinggo City	33	17
Sidoarjo	74	25	Pasuruan City	29	20
Mojokerto	34	41	Mojokerto City	38	8
Jombang	27	23	Madiun City	42	8
Nganjuk	49	26	Surabaya City	60	15
Madiun	46	4	Batu City	52	23

Table 3 shows that the characteristics which differentiate between groups occurred in some districts, namely in Kediri District, Banyuwangi District, Bondowoso District, Sidoarjo District, Mojokerto District, Madiun district, Magetan district, Sampang District, Kediri City, Blitar City, Malang city, Mojokerto City, Madiun city, Surabaya City and Batu City. It is shown from the difference between cluster 1 and cluster 2 is more than 50 percent, and specifically on cluster 2 is greater than the number of members of the cluster 1.

#### 4. CONCLUSION

The results of the study showed with the approach of K-Modes clustering that clustering data Mini survey result is II-Cluster with the level of accuracy of 67.0 percent. District/City which has different characteristics in Kediri District, Banyuwangi District, Bondowoso District, Sidoarjo District, Mojokerto District, Madiun district, Magetan district, Sampang District, Kediri City, Blitar City, Malang city, Mojokerto City, Madiun City, Surabaya City and Batu City. Groups-1 characterized by education completed Senior High School, the number of 2 people and age marry first 20 - 25 years, whereas groups-2 characterized by education ended Junior High School, the number of children 1 people and age marry first 15 - 20 years. Based on the results of research, need to examine the other variables which exist on the data the results of the survey mini 2014 which can affect the difference between the characteristics in each District in East Java.

#### REFERENCES

- [1] BKKBN. 2014. *Pemantauan Pasangan Usia Subur Melalui Mini Survei Indonesia*, Jakarta
- [2] Hartanto, H. 2004. *KB dan Kontrasepsi*. Jakarta : Pustaka Sinar Harapan
- [3] Manuaba. 2008. *Ilmu Kebidanan, Kandungan dan KB*. Jakarta: EGC
- [4] Rindang. 2008. Faktor Karakteristik Keluarga, Tingkat Fertilitas, dan Pemakaian Kontrasepsi. *Jurnal Kependudukan Padjajaran*, vol 10, 135-151

- [5] Gan, H. M., Buckley, L., Szegedi, E., Hudson, A., Sayka, M. 2009. *Identification of an rsh gene from a Novosphingobium sp. Necessary for quorum sensing signal accumulation*, USA
- [6] Royston E, Armstrong S. 1994. *Pencegahan Kematian Ibu Hamil*. Jakarta: Binarupa Aksara
- [7] Kumar, Srivastava Dhiraj, et al. 2011. A study to assess the unmet needs of family planning in Gwalior district and to study the factors that helps in determining it. *NATIONAL JOURNAL* 2.1: 28.
- [8] Khan and Ahmad. 2003. Computing Initial points using Density Based Multiscale Data Condensation for Clustering Categorical data, *2nd International Conference on Applied Artificial Intelligence, ICAAI'03*, Kolhapur, India
- [9] Barakbah, A.R, and Kiyoki, Y. 2009. *A pillar algorithm for k-means optimization by distance maximization for initial centroid designation* : pp. 61-68
- [10] Zhexue, H. 1998. *Extensions to the K-Means Algorithm for Clustering Large Data Sets with Categorical Values*, *Data Mining and Knowledge Discovery* 2, 283–304.
- [11] Zhexue, H. 2009. *Clustering Categorical Data with K-Modes*, 246–247.
- [12] Kant, Rao, T.L., Sundaram, P.N., 1994. An Automatic and Stable Clustering Algorithm, *Pattern Recognition Letters*, vol. 15, Issue 6, 543-549
- [13] Khan and Kant. 2007. *Computation of Initial Modes for K-modes Clustering Algorithm using Evidence Accumulation*, India
- [14] Hair, J.F. JR., Anderson, R.E, Tatham, R.L. & Black, W.C. 2006. *Multivariate Data Analysis*. Six Edition. New Jersey: Pearson Educational, Inc.
- [15] Johnson RA & Wichern DW. 2007. *Applied Multivariate Statistical Analysis*. Prentice Hall, Englewood Cliffs, New Jersey
- [16] Huang. 1997. *A Fast Clustering Algorithm to Cluster very Large Categorical Data Sets in Data Mining: DMKD*
- [17] Notoatmodjo, S. 2002. *Metodologi Penelitian Kesehatan*. Jakarta: Rineka Cipta.
- [18] Agresti, A. 2002. *Categorical Data Analysis* (Second ed.). New York: John Wiley & Sons

# The Classification of Districts in East Java Based on Results of Monitoring Family Planning Using K-Modes Clustering Approach

## ORIGINALITY REPORT

17%

SIMILARITY INDEX

15%

INTERNET SOURCES

10%

PUBLICATIONS

8%

STUDENT PAPERS

## PRIMARY SOURCES

1	<a href="http://digilib.its.ac.id">digilib.its.ac.id</a> Internet Source	1%
2	<a href="http://es.scribd.com">es.scribd.com</a> Internet Source	1%
3	Submitted to Stefan cel Mare University of Suceava Student Paper	1%
4	<a href="http://docplayer.net">docplayer.net</a> Internet Source	1%
5	<a href="http://www.butlercountycsea.org">www.butlercountycsea.org</a> Internet Source	1%
6	<a href="http://jb.asm.org">jb.asm.org</a> Internet Source	1%
7	<a href="http://ecda2014.eu">ecda2014.eu</a> Internet Source	1%
8	Rajput, Dharmveer Singh, Pramod Kumar Singh, and Mahua Bhattacharya. "IQRAM: a	1%



high dimensional data clustering technique",  
International Journal of Knowledge Engineering  
and Data Mining, 2012.

Publication

- 
- |   |   |    |
|---|---|----|
| 9 | <a href="http://eprints.ums.ac.id">eprints.ums.ac.id</a><br>Internet Source | 1% |
|---|---|----|
- 
- |    |  |    |
|----|--|----|
| 10 | Fu, Jun-Song, Yun Liu, and Han-Chieh Chao.<br>"ICA: An Incremental Clustering Algorithm<br>Based on OPTICS", Wireless Personal<br>Communications, 2015.<br>Publication | 1% |
|----|--|----|
- 
- |    |   |    |
|----|---|----|
| 11 | <a href="http://www.textroad.com">www.textroad.com</a><br>Internet Source | 1% |
|----|---|----|
- 
- |    |   |    |
|----|---|----|
| 12 | Submitted to Mansoura University<br>Student Paper | 1% |
|----|---|----|
- 
- |    |   |     |
|----|---|-----|
| 13 | <a href="http://www.dailydiary.com">www.dailydiary.com</a><br>Internet Source | <1% |
|----|---|-----|
- 
- |    |  |     |
|----|--|-----|
| 14 | Sharma, Abha, and R. S.. "A Variant of Genetic<br>Algorithm Based Categorical Data Clustering<br>for Compact Clusters and an Experimental<br>Study on Soybean Data for Local and Global<br>Optimal Solutions", International Journal of<br>Advanced Computer Science and Applications,<br>2016.<br>Publication | <1% |
|----|--|-----|
- 
- |    |   |  |
|----|---|--|
| 15 | Susilawaty, Andy; Otok, Bambang Widjanarko; |  |
|----|---|--|

Tuwo, Ambo; Daud, Anwar and Nurdin, Armyn. "DETERMINING SANITATION RISK INDEX FOR MAKASSAR AND SINJAI ISLANDS USING SECOND CONFIRMATORY FACTOR ANALYSIS (2CFA)", International Journal of Academic Research, 2015.

Publication

<1%

16

[www.psychology.ilstu.edu](http://www.psychology.ilstu.edu)

Internet Source

<1%

17

Bhattacharyya, R.. "Isolating top-k dense regions with filtration of sparse background", Pattern Recognition Letters, 20111001

Publication

<1%

18

[dblp.dagstuhl.de](http://dblp.dagstuhl.de)

Internet Source

<1%

19

"20th International Congress of Nutrition: Granada, Spain, September 15 20, 2013", Annals of Nutrition and Metabolism, 2013

Publication

<1%

20

[repository.unand.ac.id](http://repository.unand.ac.id)

Internet Source

<1%

21

[ejurnal.kependudukan.lipi.go.id](http://ejurnal.kependudukan.lipi.go.id)

Internet Source

<1%

22

[www.zju.edu.cn](http://www.zju.edu.cn)

Internet Source

<1%

23	<a href="http://water.usgs.gov">water.usgs.gov</a> Internet Source	<1%
24	<a href="http://repository.upi.edu">repository.upi.edu</a> Internet Source	<1%
25	<a href="http://ejournal.litbang.depkes.go.id">ejournal.litbang.depkes.go.id</a> Internet Source	<1%
26	<a href="http://www.sciedu.ca">www.sciedu.ca</a> Internet Source	<1%
27	<a href="http://www.regionalinvestment.com">www.regionalinvestment.com</a> Internet Source	<1%
28	<a href="http://epub.wu.ac.at">epub.wu.ac.at</a> Internet Source	<1%
29	<a href="http://scholar.sun.ac.za">scholar.sun.ac.za</a> Internet Source	<1%
30	<a href="http://www.cse.yorku.ca">www.cse.yorku.ca</a> Internet Source	<1%
31	<a href="http://www.aclweb.org">www.aclweb.org</a> Internet Source	<1%
32	<a href="http://akper-luwuk.ac.id">akper-luwuk.ac.id</a> Internet Source	<1%
33	<a href="http://textroad.com">textroad.com</a> Internet Source	<1%

---

Exclude quotes      On

Exclude matches      < 5 words

Exclude bibliography      Off

# The Classification of Districts in East Java Based on Results of Monitoring Family Planning Using K-Modes Clustering Approach

---

GRADEMARK REPORT

---

FINAL GRADE

**/100**

GENERAL COMMENTS

**Instructor**

---

PAGE 1

---

PAGE 2

---

PAGE 3

---

PAGE 4

---

PAGE 5

---

PAGE 6

---