Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 5, Number 2, April 2021
DOI : https://doi.org/10.33395/sinkron.v5i2.10883
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Application of the K-Means algorithm in grouping households by province and ownership status of owned houses

Kristin D R Sianipar[1)*], Septri Wanti Siahaan[2)], Indra Gunawan[3)]
[1)2)3)]STIKOM Tunas Bangsa Pematangsiantar, Indonesia
[1)]kristinsianipar7@gmail.com, [2)]septriwanti26@gmail.com, [3)]indra@amiktunasbangsa.ac.id

Submitted : Feb 19, 2021 | **Accepted** : Apr 6, 2021 | **Published** : Apr 8, 2021

Abstract: The house is a place or container to protect people from hot or cold air. A household is a group of people who gather in a place and share food, shelter, and others. People who live in a house, some live in a boarding house and some already have their own home ownership status. Research conducted by this author discusses the Application of Household Clustering by Province and Ownership Status of Owned Houses using K-Means Algorithm. The research data is sourced from the National Statistics Agency. The data used in doing this is the data on the percentage of households by province and the status of self-owned house ownership in 1999-2019. The data consists of 34 provinces. The variables of this study are based on the average number of percentages of households by province and ownership status of self-owned houses. The data clustering will be carried out by dividing it into 3 clusters, namely the high level of self-owned house ownership status, the moderate level of self-owned house ownership status and the low level of self-owned house ownership status. The results of this study are based on the Household index by Province and Ownership Status of Owned Houses, as many as 12 provinces with high self-owned house ownership, namely 86,14, 21 provinces with medium self-owned house ownership, namely 75,11, and 1 provinces with low self-owned house ownership, namely 51,61. The existence of this research can make people more open-minded about having their own house with ownership status.

Keywords: Household; Grouping; Community; K-Means Algorithm; Ownership

INTRODUCTION

The house is a place or container to protect people from hot or cold air. Housing is a group of houses or shelters that are useful as a residential environment or a place to live which is filled with infrastructure from the environment, namely the basic needs of the environment, such as the availability of drinking water, trash cans, electricity, roads, telephones that allow the residential environment to be useful as it should be. A household is a group of people who gather in a place and share food, shelter, and others. John FC Turner, 1972, said in his book entitled freedom to build, namely that, “The house is an integral part of a settlement, and is not a one-time physical result, but is a process that continues to develop and is related to the socio-economic mobility of its inhabitants over time, what matters and the house is the impact on the occupants, not their form or physical standards. Furthermore, it is said that the interaction between the house and the occupants is what the house gives to the occupants and what the residents do to the house”.

Law of the Republic of Indonesia number 1 of 2011 concerning housing and settlement areas in point d states that "Regional growth and development that does not pay attention to balance for the interests of low-income people results in difficulties for people to obtain decent and affordable housing". The Indonesian nation is a pluralistic nation. The census data counts 236,728,379 Indonesian citizens as well as 73,217 foreign nationals who are in Indonesia and 839,730 are unknown, with a total number of 34 provinces (Ong, 2013).

From the description above, we can see that Indonesia has a large population. Therefore, to make it easier for the government to calculate the population census and achieve equitable development, the authors conducted a study on grouping households by province and ownership status of their own houses. The data source that will be used in the research is from the Central Statistics Agency (BPS).

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 5, Number 2, April 2021
DOI : https://doi.org/10.33395/sinkron.v5i2.10883
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Table 1
Data on Percentage of Households by Province and Ownership Status, 1999-2019

Provinsi	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2011*	2012	2012*	2013	2013*	2014	2015	2016	2017	2018	2019
Aceh	83,56	-	-	52,53	86,87	82,98	78,33	79,04	77,89	77,95	77,46	76,59	78,80	79,02	78,43	78,71	79,63	79,86	80,33	82,36	81,66	80,42	81,21	80,32
Sumatera Utara	74,32	72,05	69,91	71,03	73,64	70,42	70,62	69,07	66,28	66,14	67,34	66,58	65,43	66,07	68,02	68,79	67,62	68,23	68,50	71,09	69,95	67,53	68,32	68,35
Sumatera Barat	79,12	78,33	72,56	73,32	74,03	74,98	72,91	69,84	64,91	68,28	66,55	68,53	69,42	69,51	70,04	70,46	70,09	70,51	69,27	74,13	72,89	70,58	71,21	70,75
Riau	76,37	74,71	70,29	74,08	73,66	72,49	74,37	71,59	66,89	67,04	67,25	66,56	67,11	68,42	67,51	68,80	69,15	70,17	68,09	71,56	71,53	68,15	69,90	70,55
Jambi	79,29	81,07	75,63	77,82	78,34	76,99	77,42	77,54	75,06	75,37	76,47	76,98	76,98	78,02	77,81	78,84	79,05	79,93	80,38	83,94	84,13	81,39	83,64	82,26
Sumatera Selatan	80,64	80,85	77,18	78,76	80,71	79,18	79,62	77,57	76,30	76,60	75,51	75,89	76,63	77,37	80,33	80,94	79,45	79,95	80,43	83,02	83,09	80,60	80,65	81,52
Bengkulu	77,17	81,11	80,78	81,52	81,06	78,60	79,76	77,10	75,78	75,39	77,36	78,05	78,35	79,15	79,55	80,56	80,62	81,25	80,64	83,52	84,27	80,83	83,39	82,72
Lampung	91,59	91,61	89,88	90,47	90,11	89,88	88,80	88,07	86,05	87,01	85,97	86,70	86,23	86,64	85,92	86,50	87,30	87,62	88,30	90,35	90,71	86,36	87,89	88,60
Kepulauan Bangka Belitung	-	-	79,20	84,97	86,27	82,29	82,72	84,23	80,88	81,54	80,91	80,75	80,89	81,71	84,25	85,10	83,63	84,26	84,31	87,85	87,59	83,75	85,91	85,51
Kepulauan Riau	-	-	-	-	-	-	67,21	67,33	63,83	70,95	71,86	67,09	62,72	64,33	64,54	66,78	62,31	64,73	67,91	67,67	68,99	65,70	69,44	66,62
DKI Jakarta	65,43	62,10	51,42	59,23	56,57	55,25	53,30	51,69	47,76	50,26	48,02	45,19	46,63	47,95	53,43	56,59	46,18	47,26	47,76	51,09	50,16	48,33	47,85	47,12
Jawa Barat	84,42	84,64	82,21	83,47	84,03	83,22	82,83	81,20	79,10	77,28	78,09	75,67	77,94	78,39	78,12	78,99	78,16	78,87	78,54	80,63	81,58	76,83	77,71	77,89
Jawa Tengah	92,33	91,74	88,47	89,43	90,19	89,06	89,12	89,18	88,35	88,31	88,45	87,88	87,64	87,98	88,56	88,95	87,76	88,14	88,06	90,93	90,77	88,01	88,17	89,04
DI Yogyakarta	77,90	80,84	73,19	77,25	75,13	72,61	73,31	72,41	74,09	75,36	78,63	74,50	76,51	75,70	76,62	76,83	76,45	76,88	74,97	76,99	77,40	75,26	76,54	73,29
Jawa Timur	89,87	89,92	86,25	87,69	89,54	88,08	88,20	88,27	86,66	87,63	88,09	87,05	86,62	86,85	88,13	88,56	87,28	87,66	87,14	90,46	90,76	87,55	87,46	87,58
Banten	-	-	74,45	81,29	81,26	82,57	80,82	82,16	76,56	75,41	74,20	72,33	75,96	76,20	76,98	78,05	76,70	77,48	77,96	80,94	81,52	81,38	81,33	80,36
Bali	84,27	83,63	78,74	79,24	83,65	79,46	79,03	77,01	73,86	77,11	76,84	71,28	70,25	70,30	75,00	76,17	71,47	72,36	69,80	77,31	74,05	70,54	71,75	72,88
Nusa Tenggara Barat	91,66	89,56	86,48	88,94	88,61	87,33	87,87	85,56	84,45	85,06	84,34	84,46	85,26	85,77	85,07	85,62	86,78	87,10	85,29	87,85	88,30	85,66	85,48	87,15
Nusa Tenggara Timur	90,62	91,78	88,30	87,92	88,87	88,39	88,00	87,14	85,11	85,13	85,36	83,74	86,78	86,56	86,01	86,30	87,31	87,43	86,33	88,52	88,53	86,50	85,91	86,88
Kalimantan Barat	89,48	87,33	85,28	85,91	88,87	87,17	86,51	87,60	84,65	85,08	85,82	84,00	84,85	85,30	87,60	87,87	88,58	88,81	87,53	90,07	89,06	87,99	88,84	88,52
Kalimantan Tengah	81,64	80,40	78,11	79,78	80,87	78,55	78,90	78,47	74,55	77,53	76,93	72,68	75,20	75,97	77,09	77,87	73,22	74,10	73,51	77,99	75,74	76,80	75,24	75,58
Kalimantan Selatan	81,63	79,81	76,14	77,81	79,85	77,93	78,52	77,80	75,18	75,78	74,79	73,78	73,74	74,67	75,23	76,08	75,69	76,41	74,39	79,22	79,84	77,30	76,77	76,33
Kalimantan Timur	70,68	68,41	66,71	66,30	67,91	68,81	69,64	69,10	65,65	64,92	66,05	63,88	64,93	66,18	69,93	71,35	66,96	68,10	68,91	72,69	71,12	68,87	70,31	69,13
Kalimantan Utara	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	74,77	70,02	69,72	67,23	72,06
Sulawesi Utara	79,95	78,96	71,13	73,89	78,18	74,51	76,58	75,95	74,65	72,11	72,75	73,75	73,21	73,47	74,97	75,40	75,76	76,12	74,60	80,44	79,77	76,79	77,65	77,27
Sulawesi Tengah	84,67	85,42	83,91	82,84	84,28	84,96	82,63	82,22	80,53	80,78	81,50	81,34	82,71	83,31	85,29	85,84	83,04	83,91	84,66	87,14	85,37	83,70	84,32	84,31
Sulawesi Selatan	88,08	88,14	84,24	85,20	86,21	85,98	84,50	83,52	80,49	82,26	81,11	82,40	82,59	82,73	83,10	83,53	83,49	83,84	83,82	86,85	85,72	82,75	83,61	83,69
Sulawesi Tenggara	87,71	87,22	83,20	82,65	85,89	85,25	83,44	82,02	82,83	82,30	83,94	82,32	84,56	84,72	84,50	85,10	85,03	85,42	86,32	86,47	86,37	84,78	84,43	85,28
Gorontalo	-	-	74,24	73,68	77,05	74,22	73,49	70,86	72,48	77,60	75,10	74,44	78,14	78,38	80,82	81,20	80,23	80,71	79,24	81,66	82,93	80,45	79,48	80,64
Sulawesi Barat	-	-	-	-	-	-	-	88,67	86,20	86,72	86,77	83,99	86,65	86,37	88,17	88,05	88,62	88,71	89,00	91,47	90,99	88,98	86,85	86,80
Maluku	85,73	-	75,69	62,74	81,69	79,64	81,04	78,79	78,59	79,02	78,96	74,57	79,18	79,39	81,35	81,81	79,36	79,53	79,84	81,51	80,48	80,15	80,09	79,48
Maluku Utara	-	-	83,53	64,80	82,67	84,41	83,70	84,32	81,41	82,13	81,74	82,27	83,84	84,77	85,24	85,50	85,41	85,88	86,13	87,84	87,65	86,83	83,87	82,73
Papua Barat	-	-	-	-	-	-	-	69,98	65,19	66,06	67,71	63,67	67,23	66,63	66,79	68,37	72,46	73,01	70,88	74,57	72,50	72,06	69,19	71,27
Papua	79,18	82,52	76,51	50,07	77,23	77,34	73,74	76,80	75,37	76,95	77,35	81,71	80,57	80,57	81,99	82,59	81,28	81,90	81,70	81,69	83,16	81,00	81,36	82,12

In conducting this research, the authors use an algorithm found in Data Mining, namely the K-Means Algorithm. In this study, results will be obtained which states which provinces have the highest, medium, and low own home ownership status.

This study is based on the novelty of previous research researched by (Sadewo et al., 2017) which states that the results of the assessment are based on the chicken population index with 1 high population level province namely West Java, 6 moderate population level provinces namely North Sumatra, Central Java, East Java, Banten South Kalimantan and East Kalimantan, and 27 other provinces including low population levels, using data from 2009-2016 consisting of 34 provinces. The data will be processed by clustering in 3 clusters, namely high population level clusters, medium and low population level clusters. Centroid data for high population level clusters 4711403141, Centroid data for medium population level clusters 304240647, and Centroid data for low population level clusters 554200.

LITERATURE REVIEW

Data mining is a technique or method of processing data that is useful for finding hidden patterns from that data. The results of data processing with data mining methods can be used to make decisions in the future. Data mining is also known as pattern recognition (Sihombing, 2017). Data mining is a process that employs one or more computer learning techniques (machine learning) to analyze and extract knowledge automatically (Sadewo et al., 2017).Data mining is a process that uses statistical techniques, mathematics, artificial intelligence, and machine learning to extract and identify useful information and related knowledge from large databases(Rony, 2016).One of the approaches used in developing the clustering method is the K-Means method, where this method is a non-hierarchical data grouping method that seeks to partition data into two or more groups (clusters) with the same characteristics and put into the same group (Darmi & Setiawan, 2016).

Data mining can also be called Knowledge Discovery in Database (KDD) or pattern recognition. The term KDD or it can be said the discovery of data knowledge because the main purpose of data mining is to utilize data in the database by processing it so as to produce new useful information. While the term pattern recognition or called pattern recognition has the purpose of knowledge that will be extracted from the chunks of data that are being faced (Nur et al., 2017).

Based on Erene Gernaria Sihombing (2017) in her research conducted clustering of data on the Percentage of Households by Province and Ownership Status of Contract / Leased Houses from 1999-2016 consisting of 34 provinces using the K-Means method. The K-Means algorithm is defined as a method of clustering to separate data into different groups (Sihombing, 2017). The results of research conducted by Erene Gernaria Sihombing (2017) were that the percentage of households was assessed according to Province and Ownership Status of

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 5, Number 2, April 2021
DOI : https://doi.org/10.33395/sinkron.v5i2.10883
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Contract / Leased Houses with 2 high-level cluster provinces, namely the Riuan Islands and DKI Jakarta, 12 medium-level cluster provinces namely North Sumatra, West Sumatra, Riau, Bengkulu, Central Java, Banten, Bali, South Kalimantan, East Kalimantan, Central Kalimantan, West Papua and Papua and 20 low-level cluster provinces namely Aceh, Jambi, South Sumatra, Lampung, Bangka Belitung, West Java, Central Java, East Java, NTB, NTT, West Kalimantan, South Kalimantan, North Sulawesi, Central Sulawesi, South Sulawesi, Southeast Sulawesi, Gorontalo and West Sulawesi. Based on research conducted by Nurul et.al on clustering of scholarship applicants producing research results with a purity value on a partially codified data dataset for the cluster algorithm k-means results of 61.11%. In the codification dataset, the purity value of the cluster algorithm result is 80.56%. And for the original data dataset, the purity value of the k-means cluster algorithm is 75%(nurul rohmawati, sofi defiyanti, 2015).

Erni and Siti conducted a research on clustering to determine the nutritional value of toddlers using the k-means algorithm. The results showed that there were 30% obese children under five and 11% malnutrition, therefore it is necessary to have the care and attention from cadres of Posyandu and related Puskesmas to parents of toddlers(Irfiani & Rani, 2018). Endang conducted research on lecturer data clustering related to activities and performance in accordance with the implementation and responsibility using the K-Means algorithm (Nurzahputra et al., 2017).

The K – Means algorithm is an algorithm that plays an important role in the field of data mining and is simple to implement and run. In addition, there is a development of a variant of the K-Means method, the K-Medoids Algorithm, which appears to overcome the weakness of the K-Means Algorithm which is sensitive to outliers because an object with a large value may deviate substantially from the data distribution. Research conducted by (Kamila et al., 2019) in making comparisons between the K-Means Algorithm and the K-Medoids Algorithm by comparing based on a dataset. From the experiments that have been carried out, the results of K-Means processing only takes an average of 1 second, while data processing on K-Medoids takes an average of 1 minute 38 seconds on the Rapid Miner.

METHOD

The K-Means algorithm is a non-hierarchical data clustering method or technique that attempts to partition existing data into two or more groups. This method partitions data into groups so that data with the same characteristics are put into the same group and data with different characteristics are grouped into other groups. The purpose of grouping this data is to minimize the objective function arranged in the grouping process, which generally seeks to minimize variations within a group and maximize variation between groups (Sari et al., 2018).The simplicity of the k-means algorithm makes the K-means algorithm applicable in various fields, such as a study conducted by Sulastri et.al on the application of data mining in the Classification of Thalassemia Patients (Sulastri & Gufroni, 2017). The similarity measure used in the cluster is a function of distance. So that maximization of data similarity is obtained based on the shortest distance between the data and the centroid point (Prayoga et al., 2019).

Basically, the use of the K-Means algorithm in conducting the clustering process depends on the existing data and the conclusions to be achieved (Agustin, 2015). An algorithm used in data clustering by using separators to divide it into different blocks is an understanding of the K-Means Algorithm (Wardhani, 2016). The steps in completing calculations on the K-Means Algorithm are: (R Sianipar et al., 2020), the first, specify k as the number of clusters to be created. Define the cluster center. The second, calculate the distance of each data set to the center of the cluster using the Euclidean equation.

$$d_{ik} = \sqrt{\sum_{j}^{m} (Cij - Ckj)^2} \quad (1)$$

The third, group data into clusters with the shortest distance using equations.

$$\text{Min} \sum_{k=1}^{k} d_{ik} = \sqrt{\sum_{j}^{m} (Cij - Ckj)^2} \quad (2)$$

The fourth, calculate the center of the new cluster using the equation

$$C_{kj} = \frac{\sum_{i=1}^{p} x_{ij}}{p} \quad (3)$$

Where:
Xij E k - th cluster
P = number of members of the kcluster

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 5, Number 2, April 2021
DOI : https://doi.org/10.33395/sinkron.v5i2.10883
e-ISSN : 2541-2019
p-ISSN : 2541-044X

The fifth, repeat steps 2 through 4 until no more data has moved to another cluster.

RESULT

Research conducted by the author refers to the data that has been obtained from BPS (Central Bureau of Statistics).

Number of Clusters : 3
Amount of Data : 34
Number of Attributes : 1

Iteration-1

The first, determine the starting centroid. The data used for the centroid are:

{Lampung, Bali, DKI Jakarta}
M1 = (88.27), M2 = (75.67), M3 = (51.61)

The second, calculate the distance of each data to the center of the centroid cluster to produce the closest centroid.

$$DM_1 = \sqrt{(78,82-88,27)^2}$$
$$DM_1 = \sqrt{(-9,45)^2}$$
$$DM_1 = 9,45$$

$$DM_2 = \sqrt{(78,82-75,67)^2}$$
$$DM_2 = \sqrt{(3,15)^2}$$
$$DM_2 = 3,15$$

$$DM_3 = \sqrt{(78,82-51,61)^2}$$
$$DM_3 = \sqrt{(27,21)^2}$$
$$DM_3 = 27,21$$

The third, calculating the centroid until the nth data. Then, the calculation results will be obtained in iteration-1 as in Table 1.

Table 2
Results of Distance Calculation in Iteration-1

Provinsi	Rata-rata	C1	C2	C3	Jarak Terdekat	*Cluster*
Aceh	78,82	9,456864838	3,149648812	27,20872046	3,149648812	C2
Sumatera Utara	68,97	19,30163316	6,695119507	17,36395214	6,695119507	C2
Sumatera Barat	71,34	16,93057757	4,324063918	19,73500773	4,324063918	C2
Riau	70,26	18,01314886	5,406635205	18,65243645	5,406635205	C2
Jambi	78,93	9,34188049	3,26463316	27,32370481	3,26463316	C2
Sumatera Selatan	79,28	8,989977682	3,616535969	27,67560762	3,616535969	C2
Bengkulu	80,02	8,25095454	4,35555911	28,41463076	4,35555911	C2
Lampung	88,27	0	12,60651365	36,6655853	0	C1
Kepulauan	83,57	4,704096197	7,902417453	31,9614891	4,704096197	C1
Bangka Belitung Kepulauan Riau	66,67	21,60628035	8,9997667	15,05930495	8,9997667	C2
DKI Jakarta	51,61	36,6655853	24,05907165	0	0	C3
Jawa Barat	79,99	8,280889085	4,325624566	28,38469622	4,325624566	C2
Jawa Tengah	89,02	0,749231071	13,35574472	37,41481637	0,749231071	C1

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 5, Number 2, April 2021
DOI : https://doi.org/10.33395/sinkron.v5i2.10883
e-ISSN : 2541-2019
p-ISSN : 2541-044X

DI Yogyakarta	75,78	12,49528993	0,11122372	24,17029537	0,11122372	C2
Jawa Timur	88,05	0,218539234	12,38797442	36,44704607	0,218539234	C1
Banten	78,45	9,823044785	2,783468865	26,84254052	2,783468865	C2
Bali	75,67	12,60651365	0	24,05907165	0	C2
Nusa Tenggara Barat	86,65	1,620477871	10,98603578	35,04510743	1,620477871	C1
Nusa Tenggara Timur	87,23	1,046956884	11,55955677	35,61862842	1,046956884	C1
Kalimantan Barat	87,20	1,076301018	11,53021263	35,58928428	1,076301018	C1
Kalimantan Tengah	76,95	11,32669519	1,279818462	25,33889011	1,279818462	C2
Kalimantan Selatan	76,86	11,41125144	1,195262206	25,25433386	1,195262206	C2
Kalimantan Timur	68,19	20,08388817	7,477374523	16,58169713	7,477374523	C2
Kalimantan Utara	70,76	17,51235935	4,905845699	19,15322595	4,905845699	C2
Sulawesi Utara	75,74	12,52830776	0,078205887	24,13727754	0,078205887	C2
Sulawesi Tengah	83,70	4,57748695	8,029026701	32,08809835	4,57748695	C1
Sulawesi Selatan	84,08	4,195395334	8,411118317	32,47018997	4,195395334	C1
Sulawesi Tenggara	84,66	3,617023629	8,989490021	33,04856167	3,617023629	C1
Gorontalo	77,59	10,68021951	1,926294138	25,98536579	1,926294138	C2
Sulawesi Barat	87,82	0,44877396	12,15773969	36,21681134	0,44877396	C1
Maluku	79,11	9,158618221	3,447895429	27,50696708	3,447895429	C2
Maluku Utara	83,48	4,788057845	7,818455805	31,87752746	4,788057845	C1
Papua Barat	69,27	19,00409057	6,397576918	17,66149473	6,397576918	C2
Papua	78,61	9,660852333	2,945661318	27,00473297	2,945661318	C2

The fourth, creating clustering based on the clusters that have been obtained. Based on Table 1, the data clustering in iteration-1 has been obtained.

Table 3
Data Clustering in Iteration-1

Provinsi	*Cluster*
Lampung	C1
Kepulauan Bangka Belitung	C1
Jawa Tengah	C1
Jawa Timur	C1
Nusa Tenggara Barat	C1
Nusa Tenggara Timur	C1
Kalimantan Barat	C1
Sulawesi Tengah	C1
Sulawesi Selatan	C1
Sulawesi Tenggara	C1
Sulawesi Barat	C1
Maluku Utara	C1
Aceh	C2
Sumatera Utara	C2
Sumatera Barat	C2
Riau	C2
Jambi	C2
Sumatera Selatan	C2

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 5, Number 2, April 2021
DOI : https://doi.org/10.33395/sinkron.v5i2.10883
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Bengkulu	C2
Kepulauan Riau	C2
Jawa Barat	C2
DI Yogyakarta	C2
Banten	C2
Bali	C2
Kalimantan Tengah	C2
Kalimantan Selatan	C2
Kalimantan Timur	C2
Kalimantan Utara	C2
Sulawesi Utara	C2
Gorontalo	C2
Maluku	C2
Papua Barat	C2
Papua	C2
DKI Jakarta	C3

The fifth, create a new centroid cluster. In obtaining the new centroid, then, must calculate the average value of each cluster.

In the 1st cluster there are 12 data, so:

$$C1_{Mean} = \frac{88,27+83,57+89,02+88,05+86,65+87,23+87,20+83,70+84,08+84,66+87,82+83,48}{12}$$

$$C1_{Mean} = 86,14$$

In the 2st cluster there are 21 data, so:

$$C2_{Mean} = \frac{\begin{array}{l}78,82+68,97+71,34+70,26+78,93+79,28+80,02+\\66,67+79,99+75,78+78,45+75,67+76,95+76,86+\\68,19+70,76+75,74+77,59+79,11+69,27+78,61\end{array}}{21}$$

$$C2_{Mean} = 75,11$$

In the 3st cluster there are 1 data, so:

$$C3_{Mean} = \frac{51,61}{1}$$

$$C3_{Mean} = 51,61$$

Thus, a new centroid is obtained, namely :
M_1 = {86,14} ; M_2 = {75,11} ; M_3 = {51,61}
After obtaining the new centroid, it is forwarded to the 2nd iteration.

After obtaining the new centroid, the process is returned to step 2 in iteration-1 and the calculation is carried out again. So, the results of the distance calculation in iteration-2 are obtained, namely:

Table 4
Distance Calculation Results in Iteration-2

Provinsi	Rata-rata	C1	C2	C3	Jarak Terdekat	Cluster
Aceh	78,82	7,323916265	3,706083735	27,2060837	3,706083735	C2
Sumatera Utara	68,97	17,16868458	6,138684583	17,3613154	6,138684583	C2
Sumatera Barat	71,34	14,797629	3,767628995	19,732371	3,767628995	C2
Riau	70,26	15,88020028	4,850200282	18,6497997	4,850200282	C2
Jambi	78,93	7,208931916	3,821068084	27,3210681	3,821068084	C2
Sumatera Selatan	79,28	6,857029108	4,172970892	27,6729709	4,172970892	C2

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 5, Number 2, April 2021
DOI : https://doi.org/10.33395/sinkron.v5i2.10883
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Bengkulu	80,02	6,118005966	4,911994034	28,411994	4,911994034	C2
Lampung	88,27	2,132948573	13,16294857	36,6629486	2,132948573	C1
Kepulauan Bangka Belitung	83,57	2,571147623	8,458852377	31,9588524	2,571147623	C1
Kepulauan Riau	66,67	19,47333178	8,443331777	15,0566682	8,443331777	C2
DKI Jakarta	51,61	34,53263673	23,50263673	0,00263673	0,00263673	C3
Jawa Barat	79,99	6,147940511	4,882059489	28,3820595	4,882059489	C2
Jawa Tengah	89,02	2,882179644	13,91217964	37,4121796	2,882179644	C1
DI Yogyakarta	75,78	10,36234136	0,667658643	24,1676586	0,667658643	C2
Jawa Timur	88,05	1,914409339	12,94440934	36,4444093	1,914409339	C1
Banten	78,45	7,690096212	3,339903788	26,8399038	3,339903788	C2
Bali	75,67	10,47356508	0,556434923	24,0564349	0,556434923	C2
Nusa Tenggara Barat	86,65	0,512470702	11,5424707	35,0424707	0,512470702	C1
Nusa Tenggara Timur	87,23	1,085991689	12,11599169	35,6159917	1,085991689	C1
Kalimantan Barat	87,20	1,056647556	12,08664756	35,5866476	1,056647556	C1
Kalimantan Tengah	76,95	9,193746615	1,836253385	25,3362534	1,836253385	C2
Kalimantan Selatan	76,86	9,27830287	1,75169713	25,2516971	1,75169713	C2
Kalimantan Timur	68,19	17,9509396	6,9209396	16,5790604	6,9209396	C2
Kalimantan Utara	70,76	15,37941078	4,349410776	19,1505892	4,349410776	C2
Sulawesi Utara	75,74	10,39535919	0,63464081	24,1346408	0,63464081	C2
Sulawesi Tengah	83,70	2,444538376	8,585461624	32,0854616	2,444538376	C1
Sulawesi Selatan	84,08	2,06244676	8,96755324	32,4675532	2,06244676	C1
Sulawesi Tenggara	84,66	1,484075056	9,545924944	33,0459249	1,484075056	C1
Gorontalo	77,59	8,547270939	2,482729061	25,9827291	2,482729061	C2
Sulawesi Barat	87,82	1,684174614	12,71417461	36,2141746	1,684174614	C1
Maluku	79,11	7,025669648	4,004330352	27,5043304	4,004330352	C2
Maluku Utara	83,48	2,655109271	8,374890729	31,8748907	2,655109271	C1
Papua Barat	69,27	16,871142	5,841141995	17,658858	5,841141995	C2
Papua	78,61	7,527903759	3,502096241	27,0020962	3,502096241	C2

From Table 3, the cluster is obtained as in iteration-1 and data clustering is carried out as in Table 6, which is as follows:

Table 5
Data Clustering in Iteration-2

Provinsi	*Cluster*
Lampung	C1
Kepulauan Bangka Belitung	C1
Jawa Tengah	C1
Jawa Timur	C1
Nusa Tenggara Barat	C1
Nusa Tenggara Timur	C1

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 5, Number 2, April 2021
DOI : https://doi.org/10.33395/sinkron.v5i2.10883
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Kalimantan Barat	C1
Sulawesi Tengah	C1
Sulawesi Selatan	C1
Sulawesi Tenggara	C1
Sulawesi Barat	C1
Maluku Utara	C1
Aceh	C2
Sumatera Utara	C2
Sumatera Barat	C2
Riau	C2
Jambi	C2
Sumatera Selatan	C2
Bengkulu	C2
Kepulauan Riau	C2
Jawa Barat	C2
DI Yogyakarta	C2
Banten	C2
Bali	C2
Kalimantan Tengah	C2
Kalimantan Selatan	C2
Kalimantan Timur	C2
Kalimantan Utara	C2
Sulawesi Utara	C2
Gorontalo	C2
Maluku	C2
Papua Barat	C2
Papua	C2
DKI Jakarta	C3

From the cluster results in iteration-1 and the cluster results in iteration-2 there is no change in the position of the cluster, so the iteration or the process stops and 3 clusters have been obtained.

DISCUSSION

In the research results, the authors used data sourced from the central statistics agency and the authors completed the results using manual counting and validated using the Rapidminer tool to get more accurate results.From this research, the results of the first cluster (86.14) with 12 clusters, the results of the second cluster (75.11) with 21 clusters, and the third result (51.61) with 1 cluster. So, it can be concluded that the K-Means Algorithm can be implemented in the Clustering of Households according to Province and Ownership Status of Owned Houses. Explanation of the results in this research are : In the first cluster (86.14) with 12 clusters, the level of households according to province and own house ownership is classified as “high”. Provinces classified as "high" are Lampung, Bangka Belitung Islands, Central Java, East Java, West Nusa Tenggara, East Nusa Tenggara, West Kalimantan, Central Sulawesi, South Sulawesi, Southeast Sulawesi, West Sulawesi and North Maluku, in the second cluster (75.11) with 21 clusters, the level of the Household by Province and Ownership Status is classified as “medium”. Provinces classified as "medium" are Aceh, North Sumatra, West Sumatra, Riau, Jambi, South Sumatra, Bengkulu, Riau Islands, West Java, DI Yogyakarta, Banten, Bali, Central Kalimantan, South Kalimantan, East Kalimantan, North Kalimantan, Sulawesi. North, Gorontalo, Maluku, West Papua, Papua, In the third cluster (51.61) with 1 cluster, the level of the Household by Province and Ownership Status is classified as “low”. The province that is classified as "low" is DKI Jakarta.

Comparison of results with other studies related to this research are : based on the research conducted by (Sihombing, 2017) related to data clustering, the Percentage of Households by Province and Ownership Status of Contract / Leased Homes from 1999-2016 consisting of 34 provinces using the K-Means method resulted in 2 high-level provinces, 12 medium-level provinces, and 20 provinces with low levels, while in this study resulted 12 clusters with high-level, 21 clusters with medium-level, and 1 cluster with low-level.

The study limitation of this research is that in this research does not have a level of accuracy, only obtaining clusters. So that, this research it can be developed by further researchers.

CONCLUSION

In the research that has been conducted by the author, the authors conclude that the K-Means Algorithm can cluster data on Households by Province and Ownership Status based on the average percentage of available data. From the results of calculations carried out, there are 3 clusters based on the average level, namely. In the first cluster (86.14) with 12 clusters, the level of households according to province and own house ownership is

e-ISSN : 2541-2019
p-ISSN : 2541-044X

classified as “high”. Provinces classified as "high" are Lampung, Bangka Belitung Islands, Central Java, East Java, West Nusa Tenggara, East Nusa Tenggara, West Kalimantan, Central Sulawesi, South Sulawesi, Southeast Sulawesi, West Sulawesi, and North Maluku.In the second cluster (75.11) with 21 clusters, the level of the Household by Province and Ownership Status is classified as “medium”. Provinces classified as "moderate" are Aceh, North Sumatra, West Sumatra, Riau, Jambi, South Sumatra, Bengkulu, Riau Islands, West Java, DI Yogyakarta, Banten, Bali, Central Kalimantan, South Kalimantan, East Kalimantan, North Kalimantan, Sulawesi. North, Gorontalo, Maluku, West Papua, Papua.In the third cluster (51.61) with 1 cluster, the level of the Household by Province and Ownership Status is classified as “low”. The province that is classified as "low" is DKI Jakarta.

REFERENCES

Agustin, F. E. M. (2015). Implementasi Algoritma K-Means Untuk Menentukan Kelompok Pengayaan Materi Mata Pelajaran Ujian Nasional (Studi Kasus: Smp Negeri 101 Jakarta). *Jurnal Teknik Informatika*, *8*(1), 73–78. https://doi.org/10.15408/jti.v8i1.1938

Darmi, Y., & Setiawan, A. (2016). Penerapan metode clustering k-means dalam pengelompokan penjualan produk. *Jurnal Media Infotama Universitas Muhammadiyah Bengkulu*, *12*(2), 148–157.

Irfiani, E., & Rani, S. S. (2018). Algoritma K-Means Clustering untuk Menentukan Nilai Gizi Balita. *Jurnal Sistem Dan Teknologi Informasi (JUSTIN)*, *6*(4), 161. https://doi.org/10.26418/justin.v6i4.29024

Kamila, I., Khairunnisa, U., & Mustakim, M. (2019). Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Data Transaksi Bongkar Muat di Provinsi Riau. *Jurnal Ilmiah Rekayasa Dan Manajemen Sistem Informasi*, *5*(1), 119. https://doi.org/10.24014/rmsi.v5i1.7381

Nur, F., Zarlis, M., & Nasution, B. B. (2017). Penerapan Algoritma K-Means Pada Siswa Baru Sekolahmenengah Kejuruan Untuk Clustering Jurusan. *InfoTekJar (Jurnal Nasional Informatika Dan Teknologi Jaringan)*, *1*(2), 100–105. https://doi.org/10.30743/infotekjar.v1i2.70

nurul rohmawati, sofi defiyanti, mohamad jajuli. (2015). Implementasi Algoritma K-Means Dalam Pengklasteran Mahasiswa Pelamar Beasiswa. *Jitter 2015*, *I*(2), 62–68.

Nurzahputra, A., Muslim, M. A., & Khusniati, M. (2017). Penerapan Algoritma K-Means Untuk Clustering Penilaian Dosen Berdasarkan Indeks Kepuasan Mahasiswa. *Techno.Com*, *16*(1), 17–24. https://doi.org/10.33633/tc.v16i1.1284

Ong, J. O. (2013). Implementasi Algotritma K-means clustering untuk menentukan strategi marketing president university. *Jurnal Ilmiah Teknik Industri*, *vol.12*, *no*(juni), 10–20.

Prayoga, Y., Tambunan, H. S., & Parlina, I. (2019). Penerapan Clustering Pada Laju Inflasi Kota Di Indonesia Dengan Algoritma K-Means. *BRAHMANA: Jurnal Penerapan Kecerdasan Buatan*, *1*(1), 24–30. https://doi.org/10.30645/brahmana.v1i1.4

R Sianipar, K. D., Wanti Siahaan, S., Siregar, M., & Fikrul Ilmi Zer, P. R. (2020). Penerapan Algoritma K-Means Dalam Menentukan Tingkat Kepuasan Pembelajaran Online Pada Masa Pandemi Covid-19. *Jurnal Teknologi Informasi*, *4*(1), 101–105.

Rony, S. (2016). Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Promosi Mahasiswa Baru (Studi Kasus : Politeknik Lp3i Jakarta). *Jurnal Lentera Ict*, *3*(1), 76–92.

Sadewo, M. G., Windarto, A. P., & Hartama, D. (2017). Penerapan Datamining Pada Populasi Daging Ayam Ras Pedaging Di Indonesia Berdasarkan Provinsi Menggunakan K-Means Clustering. *InfoTekJar (Jurnal Nasional Informatika Dan Teknologi Jaringan)*, *2*(1), 60–67. https://doi.org/10.30743/infotekjar.v2i1.164

Sari, R. W., Wanto, A., & Windarto, A. P. (2018). *IMPLEMENTASI RAPIDMINER DENGAN METODE K-MEANS (STUDY KASUS : IMUNISASI CAMPAK PADA BALITA BERDASARKAN PROVINSI)*. *2*, 224–230.

Sihombing, E. G. (2017). Klasifikasi Data Mining pada Rumah Tangga Menurut Provinsi dan Status Kepemilikan Rumah Kontrak/Sewa Menggunakan K-Means Clustering Method. *Journal of Computer Engineering System and Science (CESS)*, *2*(2), 74–82.

Sulastri, H., & Gufroni, A. I. (2017). Penerapan Data Mining Dalam Pengelompokan Penderita Thalassaemia. *Jurnal Nasional Teknologi Dan Sistem Informasi*, *3*(2), 299–305. https://doi.org/10.25077/teknosi.v3i2.2017.299-305

Wardhani, A. K. (2016). K-Means Algorithm Implementation for Clustering of Patients Disease in Kajen Clinic of Pekalongan. *Jurnal Transformatika*, *14*(1), 30. https://doi.org/10.26623/transformatika.v14i1.387