Jurnal Media Pratama
Volume 17 No. 1 Issue: Juni 2023, Page: 33 - 44
e-ISSN: 2721-9690

K-Means Clustering Analysis on the Distribution of Stunting Cases In Mojokerto Regency in June 2022

Analisis Klastering K-Means Pada Persebaran Kasus Stunting Di Kabupaten Mojokerto Pada Juni Tahun 2022

Haris Jamaludin[1)] Bima Yusuf Dharmahita[2)]
[1)2)] Sistem Informasi - Universitas STEKOM Semarang
STEKOM PUSAT Jl. Majapahit 605, Semarang, Jawa Tengah - Indonesia
[1)]harisjp88@gmail.com, [2)]bimayusufdharmadhita@gmail.com

ABSTRACT

The distribution of stunting cases in the Mojokerto regency is still not well mapped by the government, so the handling is not optimal. This stunting case needs attention from the government because it will also affect the development of an area or district. The method used by this study is the *K-Means Clustering* algorithm, where this study groups data into *clusters* based on the number of stunting cases that occur. This study clustered stunting case data in Mojokerto district to know the distribution of cases in each sub-district, sub-districts that have high levels of cases will get more handling or attention. Clustering can also show the likelihood of stunting in children varies significantly not only by individual child and household-level characteristics but also by provincial and sub-district level characteristics. The data used by this study is data on stunting cases in Mojokerto district in June 2022, which was obtained from the https://data.go.id website in the form of stunting case data in Mojokerto district in each sub-district based on gender. The results of clustering stunting cases using the *K-Means* algorithm are grouping regions based on the level of occurrence of cases into 3 *clusters*, namely *cluster C1 (high), cluster C2 (medium), and* cluster *C3 (* low). Clustering in stunting cases in Mojokerto district using the *K-Means algorithm* has been successfully carried out, where the clustering resulted in 3 clusters. The *K-Means Clustering* method in this study produced 7 iterations so that the final results were obtained, namely 3 *sub-districts entering cluster C1 (high), 14 sub-districts entering cluster* C2 (medium), and 10 sub-districts entering *cluster C3 (* low). It is hoped that with the results of this clustering, the district government can be more optimal in handling stunting cases that occur.

Keywords : stunting, clustering, k-means

(Sent : 18 June 2023, Revised 19 Juni 2023: Accepted : 25 Juni 2023)

ABSTRAK

Persebaran kasus stunting di kabupaten Mojokerto yang masih belum terpetakan dengan baik oleh pemerintah, sehinga penanganannya pun kurang maksimal. Kasus stunting ini perlu mendapat perhatian dari pemerintah, karena akan berpengaruh juga terhadap perkembangan suatu daerah atau kabupaten. Metode yang penelitian ini gunakan adalah algoritma *K-Means Clustering*, dimana penelitian ini mengelompokkan data ke dalam *cluster* berdasarkan jumlah kasus stunting yang terjadi. Penelitian ini melakukan klastering pada data kasus stunting di kabupaten Mojokerto dengan tujuan untuk mengetahui persebaran kasus di tiap kecamatan, kecamatan yang memiliki kasus dengan tingkat tinggi akan mendapatkan penanganan atau perhatian lebih. Klasterisasi juga dapat menunjukkan kemungkinan stunting pada anak bervariasi secara signifikan tidak hanya oleh karakteristik tingkat anak dan rumah tangga individu, tetapi juga oleh karakteristik tingkat provinsi dan kecamatan. Data yang penelitian ini gunakan adalah data kasus stunting di kabupaten Mojokerto pada bulan juni tahun 2022, yang di dapatkan dari situs https://data.go.id berupa data kasus stunting di kabupaten Mojokerto tiap kecamatan berdasarkan jenis kelamin. Hasil klastering kasus stunting menggunakan algoritma *K-Means* berupa pengelompokan daerah berdasarkan tingkatan terjadinya kasus ke dalam 3 *cluster* yaitu *cluster* C1 (tinggi), *cluster* C2 (sedang), dan *cluster* C3 (rendah). Klastering pada kasus stunting di kabupaten Mojokerto menggunakan algoritma *K-Means* telah berhasil dilakukan, dimana klastering tersebut menghasilkan 3 *cluster*. Metode *K-Means Clustering* pada penelitian ini menghasilkan 7 iterasi, sehingga diperoleh hasil akhir yaitu 3 kecamatan masuk *cluster* C1 (tinggi), 14 kecamatan masuk *cluster* C2 (sedang), dan 10 kecamatan masuk *cluster* C3 (rendah). Diharapkan dengan hasil klasterisasi ini, pemerintah Kabupaten bisa lebih maksimal dalam penanganan kasus stunting yang terjadi.
Kata Kunci: stunting, plastering, *k-means*

1. INTRODUCTION

Stunting is a growth delay caused by a lack of nutritional fulfillment in toddlers, especially in the early 1,000 days of life. This situation can be seen from his physique which does not match his height with his age and also his low weight compared to his height, all of it can be caused by hunger and vitamin deficiency. (Unicef, 2019).

Stunting has important significance and requires serious handling because it has an impact on the quality of human resources. In addition to causing impaired physical growth and increasing the risk of disease, stunting can also inhibit cognitive development which has the potential to affect children's intelligence levels and productivity in the future. However, unfortunately, there are still many people who do not fully realize that stunting in children is a serious problem, because children who are stunted look like children in general in their environment.

In addition to these factors, the factor of limited data on regional clustering based on the high and low number of stunting cases that occur in each region is still lacking. Therefore, data on the distribution of stunting cases is needed in the form of clustering for each region based on the level of stunting cases.

Therefore, this study will cluster stunting case data in each sub-district in Mojokerto Regency. The clustering method can group areas with similar stunting characteristics and prevalence to provide a more comprehensive and in-depth picture of the determinants of stunting at the regional level (Mulyaningsih et al., 2021). The cluster aims to identify sub-districts that have a high rate of stunting cases in Mojokerto Regency so that the government can make the right decisions in providing attention and handling to reduce the rate of stunting cases in each sub-district. It is hoped that the results of this research can help the government in designing the right program or policy to overcome the problem of stunting in the Mojokerto Regency.

2. LITERATURE REVIEW

2.1 Previous Research

Previous research was an attempt by researchers to find comparisons and then find new inspiration to conduct research. In this section, the researcher lists some previous studies related to the research he wants to do and produces a summary. Below is previous research that is still relevant to the topic that the author is researching.

Clustering can indicate the likelihood of stunting in children varies significantly not only by individual child and household-level characteristics but also by provincial and sub-district level characteristics. Among the child-level covariates included in our model, dietary habits, neonatal weight, history of infection, and sex significantly influenced stunting risk. Parental wealth and education status were significant household-level covariates associated with a higher risk of stunting. Finally, the risk of stunting is higher for children living in communities without access to water, sanitation, and hygiene. Stunting is not only associated with child-level characteristics but also family- and community-level characteristics. Therefore, interventions to reduce stunting must also consider family and community characteristics to achieve effective outcomes. (Mulyaningsih et al., 2021).

In addition to the fields above, the k-means algorithm can also be used to group customers based on the transaction method used, to obtain groups that have not been known before. This information is useful for SMEs to be utilized based on their needs, for example in preparing methods and supporting infrastructure in implementing transactions. Data in this study were collected from customer attributes, transaction amounts, and payment methods. The results of the iteration can be found; first, based on the number of subscribers, groups can be classified into three namely C1(18%) is an automatic transfer payment, C2 (45%) is post-date payment, C3 (36%) is non-auto-transfer payment and combination. Second, based on the average number of transactions, postpaid payments are ranked first. From these results, it can be analyzed that this situation burdens SMEs because the more the number of transactions, the more investment that must be prepared. (Marisa et al., 2021).

3. THEORETICAL BASIS

Clustering is an unattended method of data mining. It doesn't require a training dataset. The two most popular types of grouping methods are partition grouping and hierarchical grouping. In contrast to linear regression which is a predictive method, clustering is a classification method that categorizes subjects (data points) into different groups (clusters), each with several characteristic measurements (Zhou, 2020). Clustering divides and groups data into clusters based on the similarity of data types. Among several algorithms that can be used to perform clustering, k-means is the algorithm most often used for data mining.

Among the many grouping methods, the k-means developed by MacQueen is the most widely used. The simplicity of k-means makes this algorithm popular and widely applied in various fields. This algorithm can group large amounts of data quickly and efficiently, including data that is out of the ordinary (outliers). K-means remain the basic framework in the development of numerical or conceptual clustering with various distance options and prototypes. (Nepal, Yamaha, Sahashi, & Yokoe, 2019).

The K-Means algorithm has been applied in problem-solving in various fields. In agriculture, the K-Means algorithm is used to group vegetables based on their production level (Harahap, Fuadi, Rosnita, Darnila, &; Meiyanti, 2022). In economics, the K-Means algorithm is used to group customers by payment method (Marisa et al., 2021).

4. METODE

4.1 Stunting

Stunting is a growth delay caused by a lack of nutritional fulfillment in toddlers, especially in the early 1,000 days of life. This situation can be seen from his physique which does not match his height with his age and also his low weight compared to his height, all of it can be caused by hunger and vitamin deficiency (Unicef, 2019).

4.1 K-Means Clustering

K-means clustering is a type of partition clustering, where each cluster is defined by the centroid (mean) of the data points in the cluster (Zhou, 2020). In this study, the variable used was the number of stunting cases in girls (P) and boys (L). Data on the number of stunting cases by sex will be processed using clusters and divided into 3 clusters, namely C1 (High), C2 (Medium), and C3 (Low).

4.3 Euclidean Distance

Euclidean distance is a method used to measure the distance between two points in Euclidean space, we will store the Euclidean distance from each data point the b:

$$\boldsymbol{D(x_i, \mu_j) = \sqrt{(x_1 - \mu_j)^2}} \quad \textbf{(1)}$$

Where D = Document Point
χ_i = Criteria Data
μ_j = Centroid on cluster

4.4 Centroid

K-means clustering defines each cluster by centroid (mean). The midpoint (centroid) is a randomly determined point (Zhou, 2020), and random determination is carried out when determining a new (initial) midpoint. In centroid determination for the second iteration and so on, use the following formula: $\boldsymbol{y_i = \frac{\sum_{i=0}^{n} x_i}{n}}$ **(2)**

Remarks :
y_i = Centroid on the cluster
x_i = Object of observation to i
n = Number of objects that are members of the cluster

4.5 Stages of the K-Means

K-Means Cluster Analysis has the following working algorithm:

1. Determine the number of clusters (C)
2. Determine the initial centroid, by taking a random cluster value for the first time from the data.
3. Calculates the Euclidean distance value using the formula in equation (1).
4. Group data, based on the minimum value among calculated values.
5. Determine the new centroid using equation (2) formula.
6. If the position of the new centroid is not the same as the old centroid, then go back to step 3.

The convergence check is stopped if the termination criteria are met, but if it has not been met (not yet convergent) then iterations need to be done until the termination criteria (convergence) have been met.

4.6 K-Means Algorithm Scheme

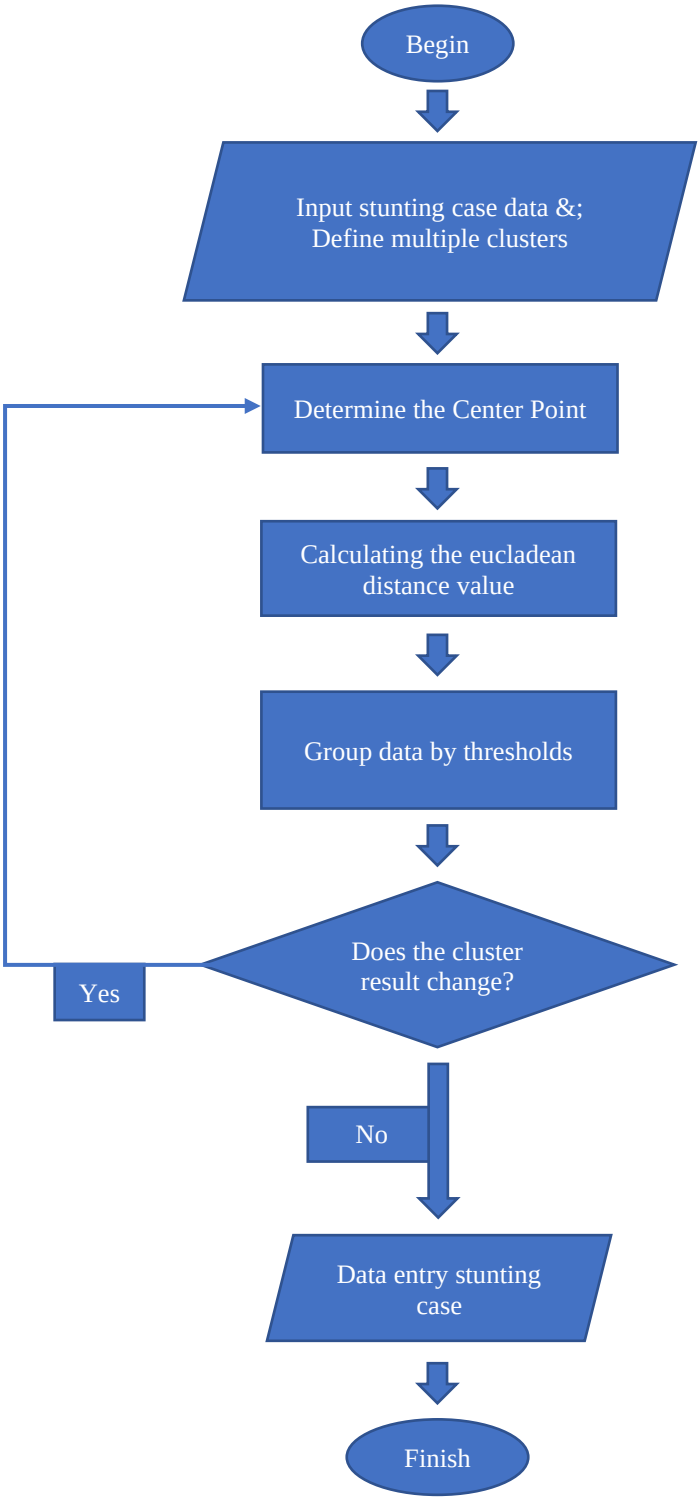

Chart_1 Scheme K-Means

Chart 1 _ The K-Means scheme is the system or process design scheme in this study, from the initial stage to the end, with the following flow :

1. Starting the k-means scheme.

2. Entering stunting case data based on regional puskesmas and also gender, namely women and men.
3. Making the initial centroid by randomly selecting from the data obtained, this study used data to 2, 4, and 6, as the initial centroid. Then determine the 3 groups or clusters that will result from this clustering process, with cluster details as follows :
    1) C1 (High)
    2) C2 (Medium)
    3) C3 (Low)
4. After determining the starting point and number of clusters, calculate the Euclidean distance using the formula in equation (1).
5. Group using the results of the Euclidean distance calculation that has been done.
6. If the results of the new grouping or clustering have results that differ from the results of the previous clustering, then regrouping must be done by iterating until the results of the new grouping or clustering are the same as the previous results.
7. The clustering results will appear when the calculation has been completed so that it can determine which cluster the data is entering (high, medium, or low cluster).
8. The calculation process is complete.

5. RESULTS AND DISCUSSION

5.1 Calculation Analysis of the K-Means Algorithm

The data that has been obtained will be analyzed using the k-means algorithm. The analysis was conducted using two attributes, namely the L attribute for the male gender and the P attribute for the female gender in stunting case data in Mojokerto district in June 2022. Using the attributes in Table 4.1 _ 1 Attributes.

Table 5.1 _ 1 Attributes

Attributes	Initial
Man	L
Woman	P

Next determine the number of clusters to be used, to perform calculations using the k-means algorithm. Clusters as in Table 4.1 _ 2 Initial Centroid Determination.

Table 5.1 _ 2 Penentuan Sentroid Awal

Cluster	Sentroid Awal
C1	Using Data Number 2 in Table 4.1 _ 3
C2	Using Data Number 4 in Table 4.1 _ 3
C3	Using Data Number 6 in Table 4.1 _ 3

Data on stunting cases in June in Mojokerto regency, contained in table 4.1 _ 3 as follows :

Table 5.1 _ 3 Data On Stunting cases Juni 2022

No.	Indication	L	P
1	Puskesmas Sooko	113	77
2	Puskesmas Trowulan	34	28
3	Puskesmas Tawangsari	40	42
4	Puskesmas Puri	13	15
5	Puskesmas Gayaman	34	32
6	Puskesmas Bangsal	7	9
7	Puskesmas Gedeg	45	39
8	Puskesmas Lespadangan	29	20
9	Puskesmas Kemlagi	40	22
10	Puskesmas Kedungsari	15	17
…	……	…	…
…	……	…	…
24	Puskesmas Pandan	54	37
25	Puskesmas Trawas	49	35
26	Puskesmas Gondang	68	60
27	Puskesmas Jatirejo	55	49

The steps taken to complete the calculation using the k-means algorithm are as follows :

1) Determine the number of clusters
   The cluster that this study uses is 3 clusters, with the following details: Cluster for high case rate (C1), cluster for medium case rate (C2), and cluster for low case rate (C3). The clustering in this study was based on data on the number of stunting cases in male and female toddlers in Mojokerto Regency as many as 27 regional health centers.

2) Creating a Starting Centroid
   Determining the initial centroid by taking data in the range of stunting case data randomly, this study used data 2 for cluster 1 (C1), data 4 for cluster 2 (C2), and data 6 for cluster 3 (C3). The data that this study used is contained in Table 4.1 _ 3 Stunting Case Data June 2022, initial centroids can be seen in Table 4.1 _ 4 Initial centroids.

Table 5.1 _ 4 Sentroid Awal

Data	Centroid	L	P
2	1	34	28
4	2	13	15
6	3	7	9

3) Calculating the Euclidean distance value
   $(x^1,y^1) = \sqrt{(113-34)^2 + (77-28)^2} = 92{,}96$ (C1)

$(x^2,y^2) = \sqrt{(113-13)^2 + (77-15)^2} = 117{,}66$ (C2)
$y^3) = \sqrt{(113-7)^2 + (77-9)^2} = 125{,}94$ (C3)
…….
Perform on all rows up to the last data, calculate Euclidean on variables L and P using a predefined initial centroid.

4) Determine the minimum value of the calculated Euclidean distance value as in table 5.1 _ 5 Euclidean values :

Table 4.1 _ 5 Euclidean values

No.	Indikator	L	P	C1	C2	C3	Minimum	Cluster
1	Puskesmas Sooko	113	77	92,96236	117,6605	125,9365	92,96236	C1
2	Puskesmas Trowulan	34	28	0	24,69818	33,01515	0	C1
3	Puskesmas Tawangsari	40	42	15,23155	38,18377	46,66905	15,23155	C1
4	Puskesmas Puri	13	15	24,69818	0	8,485281	0	C2
5	Puskesmas Gayaman	34	32	4	27,01851	35,4683	4	C1
6	Puskesmas Bangsal	7	9	33,01515	8,485281	0	0	C3
7	Puskesmas Gedeg	45	39	15,55635	40	48,41487	15,55635	C1
8	Puskesmas Lespadangan	29	20	9,433981	16,76305	24,59675	9,433981	C1
9	Puskesmas Kemlagi	40	22	8,485281	27,89265	35,4683	8,485281	C1
10	Puskesmas Kedungsari	15	17	21,9545	2,828427	11,31371	2,828427	C2
….	…….	….	….	……	……	……	……	…..
….	…….	….	….	……	……	……	……	…..
24	Puskesmas Pandan	54	37	21,93171	46,52956	54,70832	21,93171	C1
25	Puskesmas Trawas	49	35	16,55295	41,18252	49,39636	16,55295	C1
26	Puskesmas Gondang	68	60	46,69047	71,06335	79,51101	46,69047	C1
27	Puskesmas Jatirejo	55	49	29,69848	54,03702	62,482	29,69848	C1

5) Create a new centroid
Once the results of the previous iteration are known, then determine the new centroid for each cluster using the formula in equation (2).

The new centroids for cluster 1 or C1 are calculated based on the sum of all data on the variables that go into cluster 1.
L = (113+34+40+…+…+68+55) / 17 = 47
P = (77+28+42+…+…+60+49) / 17 = 38

The new centroid for cluster 2 or C2 is calculated based on the sum of all data on the variables that go into cluster 2.
L = (13+15+…+…+…+26+12) / 6 = 16
P = (15+17+…+…+…+14+12) / 6 = 14

The new centroid for cluster 3 or C3 is calculated based on the sum of all data on the variables that go into cluster 3.
L = (7+10+9+6) / 4 = 8

P = (9+4+7+8) / 4 = 7

The calculation for the determination of the new centroid gets the result as in Table 4.1 _ 6 New Centroid for iteration 1 below :

Table 5.1 _ 6 New Centroid for Literation 1

Centroid	L	P
1	47	38
2	16	14
3	8	7

Then calculate the Euclidean distance value again, based on the new centroid point by repeating the steps above. If the calculation produces the same cluster as the result of the previous calculation, then the calculation or iteration is stopped. However, if the calculation has a different result from the previous calculation, then the calculation or iteration is continued until it produces convergence 0. In this study, the iterations that this study needs to achieve convergent results 0 require 7 iterations. The results of the last iteration on clustering that this study conducted are in Table 4.1 _ 7 Final Results of Clustering Stunting Cases in Banyumas Regency as follows:

Table 5.1 _ 7 Final Results of Iteration of Stunting Case Clustering in Banyumas Regency

No.	Indicator	L	P	C1	C2	C3	Minimum	Cluster
1	Puskesmas Sooko	113	77	32	87	120	32,36	C1
2	Puskesmas Trowulan	34	28	62	6	27	5,69	C2
3	Puskesmas Tawangsari	40	42	49	11	41	10,55	C2
4	Puskesmas Puri	13	15	86	30	4	4,11	C3
5	Puskesmas Gayaman	34	32	59	4	30	4,45	C2
6	Puskesmas Bangsal	7	9	95	39	6	6,01	C3
7	Puskesmas Gedeg	45	39	46	10	43	9,92	C2
8	Puskesmas Lespadangan	29	20	71	15	19	14,93	C2
9	Puskesmas Kemlagi	40	22	62	10	29	9,70	C2
10	Puskesmas Kedungsari	15	17	84	28	7	6,52	C3
...	.....	...	...	...	...	...	.....	....
....	.....	....	....	...	....	....	.....	....
23	Puskesmas Pacet	27	37	63	13	30	12,65	C2
24	Puskesmas Pandan	54	37	41	16	49	16,49	C2
25	Puskesmas Trawas	49	35	46	11	44	11,11	C2
26	Puskesmas Gondang	68	60	16	41	74	15,81	C1
27	Puskesmas Jatirejo	55	49	33	24	57	24,05	C2

5.2 The final result of clustering stunting cases

The following is a visualization of the clustering results using the pie chart contained in Chart 5.2._.1 Pie Chart Case Percentage.

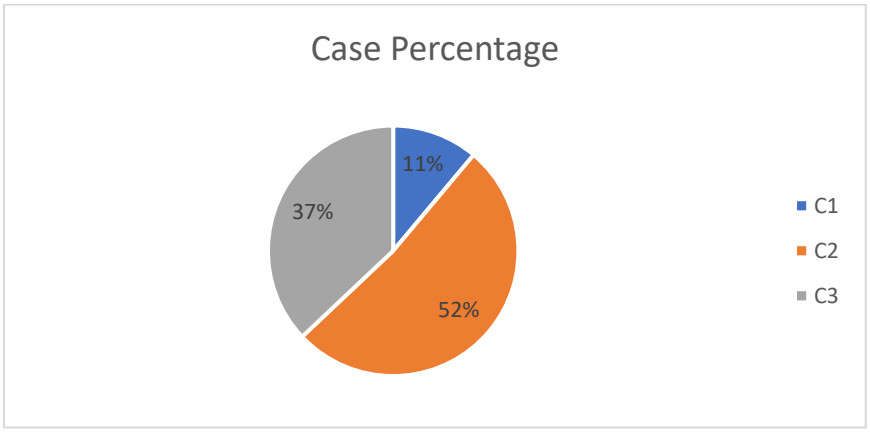

Chart 4.2 _ 1 Pie Chart Case Percentage

The iteration stopped at iteration 7, with the results of cluster 1 (C1) as many as 3 regional puskesmas, cluster 2 (C2) as many as 14 regional puskesmas, and cluster 3 (C3) as many as 10 regional puskesmas. With the mapping of stunting case rates visualized in Chart 5.2 _ 2 Distribution of Stunting Cases

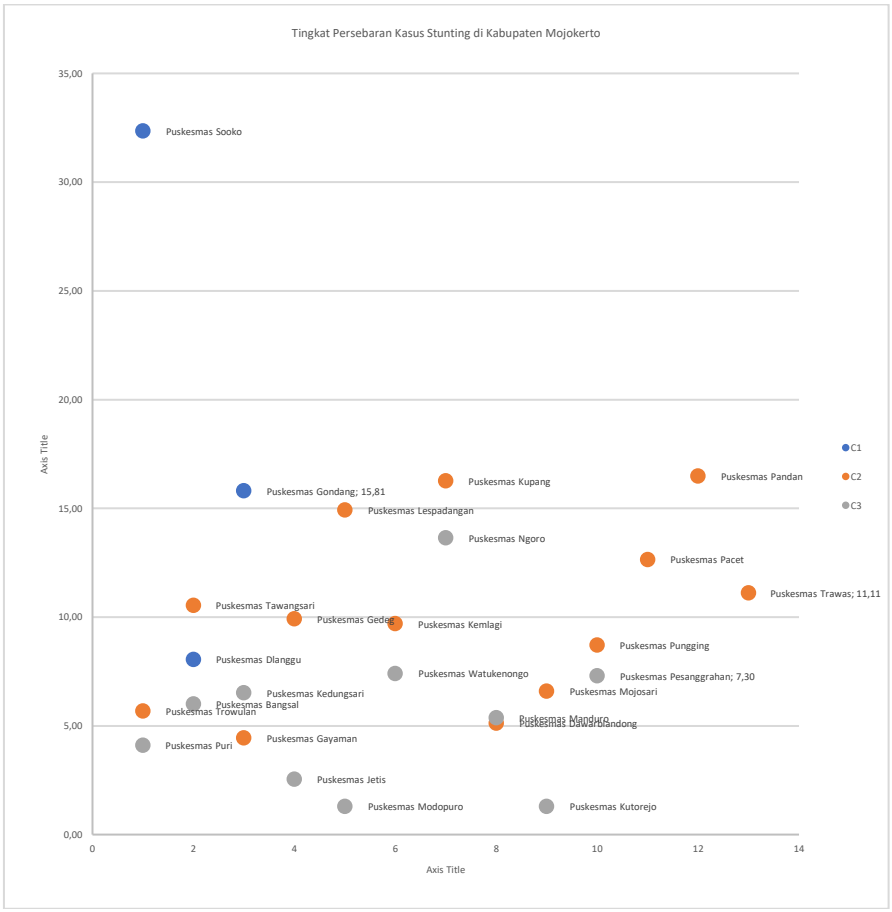

Chart 5.2 _ 2 Persebaran Kasus Stunting

6. CONCLUSION

The results of clustering stunting cases using the *K-Means* algorithm are grouping regions based on the level of occurrence of cases into 3 *clusters*, namely *cluster C1 (high), cluster C2 (medium), and* cluster *C3 (*low). Clustering in stunting cases in Mojokerto district using the *K-Means algorithm* has been successfully carried out, where the clustering resulted in 3 clusters. The *K-Means Clustering* method in this study produced 7 iterations so that the final results were obtained, namely 3 *sub-districts entering cluster C1 (high), 14 sub-districts entering cluster* C2 (medium), and 10 sub-districts entering *cluster C3 (*low). It is hoped that with the results of this clustering, the district government can be more optimal in handling stunting cases that occur. In future research, research can be integrated with additional data, such as demographic data, environmental data, or health data. This will help in gaining a deeper understanding of the factors influencing mapping patterns and can provide new insights into the analysis.

BIBLIOGRAPHY

Harahap, L. M., Fuadi, W., Rosnita, L., Darnila, E., &; Meiyanti, R. (2022). Clustering featured vegetables using the K-means algorithm. *Journal of Informatics Engineering and Information Systems*, 8(3). Retrieved from https://doi.org/10.28932/jutisi.v8i3.5277

Marisa, F., Sakinah Syed Ahmad, S., Izzah Mohd Yusof, Z., Mohammad Akhriza, T., Purnomowati, W., & Kumar Pandey, R. (2021). The Analyze of the Relationship between Revenue and Customer Payment Methods in Small and Medium Enterprise Based on Clustering K-Means. In *Journal of Physics: Conference Series* (Vol. 1908). IOP Publishing Ltd. Retrieved from https://doi.org/10.1088/1742-6596/1908/1/012021

Mulyaningsih, T., Mohanty, I., Widyaningsih, V., Gebremedhin, T. A., Miranti, R., & Wiyono, V. H. (2021). Beyond Personal Factors: Multilevel Determinants of Childhood Stunting in Indonesia. *PLoS ONE*, 16(11 November). Retrieved from https://doi.org/10.1371/journal.pone.0260265

Nepal, B., Yamaha, M., Sahashi, H., & Yokoe, A. (2019). Analysis of Building Electricity Use Pattern using K-means Clustering Algorithm by Determination of Better Initial Centroids and Number of Clusters. *Energies*, 12(12). Retrieved from https://doi.org/10.3390/en12122451

Nishom, M., &; Fathoni, M. Y. (2018). Implementation of the Rule-Of-Thumb Approach for K-Means Clustering Algorithm Optimization, 03(02).

Unicef. (2019). The State of The World's Children 2019. Children, Food, and Nutrition Growing Well in A Changing World.

Zhou, H. (2020). Learn Data Mining Through Excel. Learn Data Mining Through Excel. Apress. Retrieved from https://doi.org/10.1007/978-1-4842-5982-5