International Journal Of Artificial Intelegence Research ISSN: 2579-7298
Vol 5, No 1, June 2021, pp. 102-110

Performance Analysis Of Elbow Algorithms Against The Results Of Clustering K-means Algorithms In The Grouping Of Rice Production Areas In Indonesia

Paska Marto Hasugian[1], Bosker Sinaga[1], Jonson Manurung[1] Safa Ayoub Al Hashim[2]

[1]*Software Engineering, STMIK Pelita Nusantara, Medan, Indonesia*
[2]*College of Information Technology, University of Babylon, Iraq*
*Paskamarto86@gmail.com **

ARTICLE INFO

Article history:
Received
Revised
Accepted

ABSTRACT

Indonesia is the third largest country in the world with rice production reaching 83,037,000 million tons and became the highest production in southeast Asia spread in several regions in Indonesia The problem found that with such production has not been able to cover the needs of indonesian people with a very high population so in the study conducted information excavation to generate potential to the pile of data that has been described and analyzed by BP S (Central Bureau of Statistics) with the topic of clustering. The purpose of this research is to group rice production in Indonesia based on available data to help and provide solutions to the ministry of agriculture in determining land development priorities and can minimize the shortage of rice production nationally. Grouping process by involving K-means algorithm to group rice production followed by the utilization of elbow method provides recommendations of the most optimal clusters of several clusters produced based on the calculation of each iteration of the K-Means Algorithm, supporting attributes that are used as part of cluster determination is the area of harvest, productivity and production. Research methods with data cleaning activities, data integration, data transformation and the application of K-means with elbow combination and pattern evaluation. The results achieved based on the work description with a combination of K-Means and elbow provide cluster recommendations that are the best choice or the most optimal is iteration 2 which is the lowest rice production group with the number of 22 regions, rice production with a medium category of 9 and production with the highest category with 3 regions.

Keywords:
Cluster
K-Means
Elbow
production
rice

I. Introduction

Indonesia is an agrarian country where most of its people's livelihoods are farmers[1] . Until now, the agricultural sector still plays a role in the national economy. Rice is the main food commodity in Indonesia with a consumption rate of 132.98 kg/capita/year. So rice production becomes the top priority to overcome the supply shortage[2]. The increase in rice production occurs from several factors, namely the extent of harvest in Indonesia. Based on BPS data that Indonesia is the third-largest country in the world with rice production reached 83,037,000 million tons and became the highest production in southeast Asia spread in several regions in Indonesia[3]. The problem found that such products have not been able to cover the needs of Indonesian people with a very large population and still import from neighboring countries.

Based on agus Perdana Windarto's research explained that K-Means algorithm can be used to do the grouping process well and can be developed with additional rules to produce clusters that are more optimal and quality[4]. The use of a k-means algorithm in dharmarajan and T. Velmurugan publications compares with two methods, namely k-Means and Fuzzy C-Means (FCM) by providing information on decision making in detecting cancer-affected areas[5].Riski Annisa et al use K-Means Algorithm to predict software defect module errors by proposing a point center algorithm to determine the initial centroid value against k-means algorithm optimization have developed an algorithm in performing point center [6] . Dewi Pramudi

Ismi et al's research shows that K-means clustering-based feature selection can be done to produce a subset of features with 561 features and 1080 data with 857 features capable of delivering 80% accuracy results to classification dimensions [7]. S. Santha Subbulaxmi, G. Arumugam with the development of algorithms to improve the performance of algorithms by adapting ensemble techniques based on classification problems with unbalanced data[8].

Referring to the excerpts from some researches provides information and knowledge in accordance with the object studied and found some activities that have not been done by researchers. Some of these activities are in accordance with the analysis and the results of some researchers state that the K-Means algorithm has limitations in a grouping so that K-Means can be refined by using new algorithms or methods either in terms of determining the central point or determining the most optimal cluster that deserves to be used as Knowledge. in the summary of research on the comparison of K-Means Algorithm with C-Means has not been able to give an idea of the effectiveness of the K-Means algorithm and still put forward the last iteration to be the more recommended. To complete the obstacles that have been poured then in this study involved one method that has the task to provide opportunities to all groups formed have the same opportunities to be recommended, namely the Elbow method.

The purpose of this research is to conduct testing by utilizing the K-Means Algorithm and Elbow Method in Supporting the excavation of information on rice production areas in Indonesia as inputs to the ministry of agriculture in the development and expansion of the agricultural sector specifically on rice production so that certain areas become the top priority in development or know the region dominates the highest production that deserves to be the largest supply. The grouping process will be displayed in the study with three categories namely high, medium and low production. To support the grouping is utilized data mining studies with the main task to extract information and provide exposure to the information found by utilizing the principle of clustering [9]. The algorithm applied is K-Means Algorithm with activities to cluster based on centroid closest to data[10]. Data mining results with the kmeans algorithm can be improved with the elbow algorithm so that the result of data collection in determining clusters becomes better.

K-Means is data grouping by maximizing data similarity in a single cluster and minimizing data similarity between clusters [11]. The similarity size used in Cluster is the distance function[12]. So that the maximization of data similarity is obtained based on the shortest distance between the data against centroid point [13]. By the rules of the K-means algorithm in the process of determining the center point of the cluster is done randomly so that the cluster formed is not by the expected pattern, it is one of the activities that make the performance of the k-means algorithm does not run optimally.[14]. Identification of the number of k clusters is the most important and primary way of clustering using the K-Means algorithm where the cluster results will depend on the number of initial clusters. So if the number of clusters specified is not good then the cluster results will also not be as expected i.e. will not produce the information needed by the user [15]. So in the research involved the elbow method with the principle of the best cluster value to be taken from the sum of square error (SSE) which experienced a significant decrease and shaped elbow [16]. So that the quality of the grouping developed is more qualified and able to maximize the more dominant group[17]. Different percentage results of each cluster value can be shown by using a graph as the source of the information[18]. If the value of the first cluster with the value of the second cluster provides an angle in the graph or the value decreases the most then the value of the cluster is best[19].

II. METHOD AND MATERIALS

Research work steps to explore the potential of rice production data stacks based on the following flow :

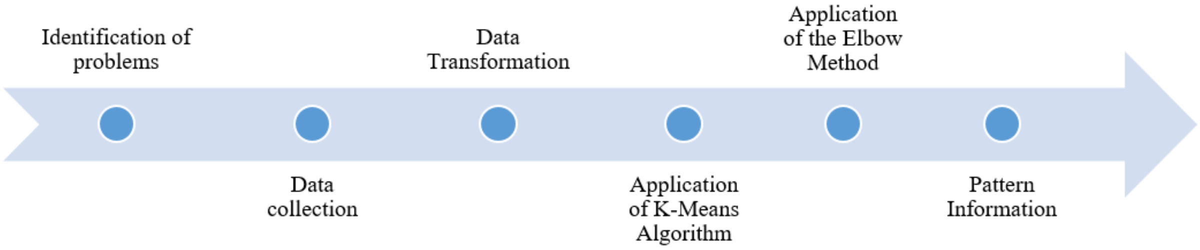

Figure 1. Research Flow

1. The identification stage is the determination of problems and objectives that will be achieved in research, namely rice production nationally ranked third but does not support the population in Indonesia so that a grouping can be used as a priority for the development of rice products in low-category areas and maintain production with high categories. For grouping category in three categories namely Cluster 3 (C3) Is High Production, Cluster 2 (C2) Medium Production, Cluster 1 (C1) Low Production[4]
2. Data Collection
   The study used a total of 34 data sourced from rice production data based on regions in Indonesia from 2018-2020 with a breakdown of supporting attribute data is the area of harvest (ha), productivity (Ku / Ha) and Production (tons). The data description used is described in table 1.

Table 1. Rice Production Data for 2018-2020 [20]

No	Harvest Area (ha)			Produktivitas (Ku/ha)			Production (tons)		
	2018	2019	2020	2018	2019	2020	2018	2019	2020
1	329.516	310.012	317.869	56	55	55	861567.10	1.714.438	1.757.313
2	110.978	95.319	90.981	60	61	58	667.069	579.321	532.168
3	344.836	303.732	325.333	49	48	51	1.687.783	1.470.503	1.655.170
4	65.891	64.407	64 137.28	44	46	46	288.811	296.472	292.834
5	93.956	111.477	110 548.12	55	48	47	514.935	533.477	523.396
6	673	623	915	73	54	50	4.899	3.359	4.544
7	56.632	49.010	48.686	48	47	47	269.540	231.211	227.627
8	86.203	69.536	84.773	44	45	46	383.046	309.933	386.413
9	1.707.254	1.578.836	1.586.889	57	58	57	9.647.359	9.084.957	9.016.773
10	1.821.983	1.678.479	1.666.931	58	58	57	10.499.588	9.655.654	9.489.165
11	1.751.192	1.702.426	1.754.380	58	56	57	10.203.213	9.580.934	9.944.538
12	286.476	290.048	256.575	28	29	30	799.715	847.875	778.170
13	323.091	356.246	289.836	41	38	40	1.327.492	1.342.862	1150 306.66
14	147.572	146.145	143.275	35	30	32	514.769	443.561	457.952
15	64.961	69.708	73.568	40	36	36	262.774	253.818	262.435
16	13.707	10.295	9.883	33	32	34	45.064	33.357	33.574
17	17.234	17.088	17.841	27	29	32	45.725	48.806	57.324
18	376	356	299	29	32	29	1.097	1.151	853
19	511.941	464.103	545.149	49	47	49	2 488 641.91	2 164 089.33	2.650.290
20	29.052	25.977	28.668	40	38	39	116 228.86	98 254.75	110.447
21	13.413	11 700.50	10 301.91	37	32	42	49 047.11	37 945.64	43.383
22	289.243	281 666.04	273 460.82	50	50	48	1 460 338.81	1 402 182.39	1.317.190
23	218.233	198 867.41	181 690.63	41	41	40	899 935.88	811 724.18	725.024
24	52.412	54 131.72	52 727.52	43	43	31	223 119.42	235 339.51	166.002
25	7.767	7 192.15	7 570.63	32	42	32	24 967.13	29 943.56	24.378
26	71.448	63 142.04	64 733.13	37	37	38	266 375.53	230 873.97	243.685
27	65.304	62 581.47	64 826.18	48	48	53	316 478.37	300 142.22	345.050
28	1.185.484	1 010 188.75	976 258.14	50	50	48	5 952 616.45	5 054 166.96	4.708.465
29	201.279	186 100.44	178 066.94	46	45	44	926 978.66	844 904.30	792.249
30	136.674	132 343.86	133 697.15	39	39	40	538 876.14	519 706.93	532.773
31	70.353	62 020.39	61 827.86	46	45	40	326 929.74	277 776.31	248.879
32	313.051	311 671.23	295 664.47	47	48	47	1 483 076.48	1 482 996.01	1.387.269
33	581.575	539 316.52	551 320.76	51	48	50	2 994 191.84	2 603 396.24	2.743.060
34	408.176	413 141.24	388 591.22	52	50	53	2 108 284.72	2 078 901.59	2.040.500

3. Data transformation is an activity that is done to make changes that have been collected by adjusting the needs of the k-means algorithm, in this process used the average data of each attribute.
4. Stages of Application of K-Means Algorithm is an activity carried out for the process of grouping rice production with the following activities:

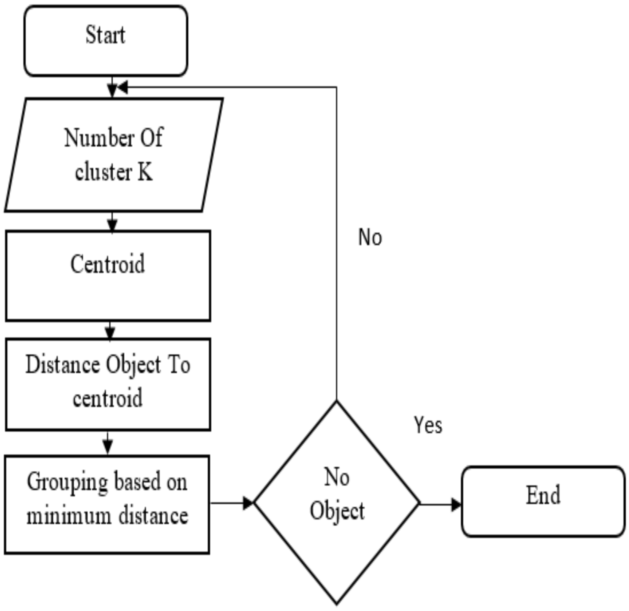

Figure 2. Flowchart Algorithm K-Means [12]

To calculate the distance of all data to each cluster center, using Euclidean distance theory.

D(i,j)= $\sqrt{(x_{Ki} - x_{Kj})^2 + (x_{Ki} - x_{Kj}) + \cdots + (x_{ki} - x_{kj})^2}$……… 1. [21]

Provided D(i,j) = distance of data I to the center of cluster j, Xki = data to I on the data attribute to k, Xkj = the center point of cluster j on the attribute to k The distance of the center of the cluster is recalculated by the current cluster membership. The center of the cluster is the average of all data or objects in a particular cluster if desired can also be used as the median value of the cluster [22].

5. Application of Elbow Method to determine the best number of clusters by looking at the percentage of comparison results between the number of clusters (K) that will form elbows at a point by utilizing SSErules (*Sum of Square Error*) with formula
SSE=$\sum_{K=1}^{K} \sum_{K=1\, X1 \in Sk} ||Xi - Ck||_2^2$……… 2 [7]

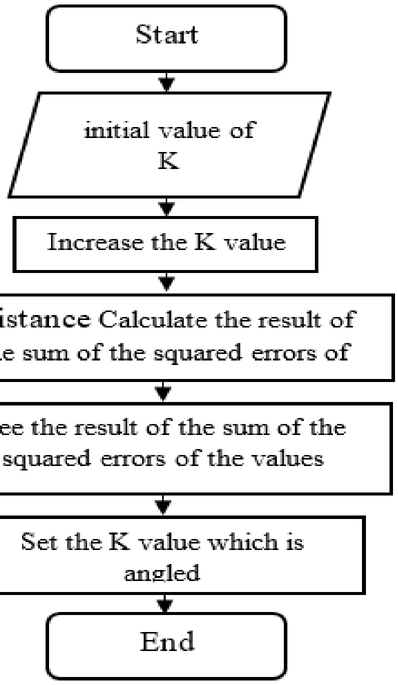

Figure 3. Flowchart Elbow Method[23]

6. Exposure of clustering results to rice production as part of decision-making solutions with graph descriptions and groupings based on predetermined attributes.

III. Result and Discussion

A. Data Transformation

The transformation stage is one of the components of data mining for data adjustment that will be used as part of the training data, in this study the data used is the average value of each attribute of harvest area, productivity, and production, description of data transformation in thousands to the harvest area, by utilizing data hundreds of Ku / Ha on productivity and thousands of tons on production data. As part of the calculation of the entire data is used by calculating the rerate from 2018-2020 for each attribute used.

Table 2. Data Transformation

Provinsi	Rerate		
	Harvest Area (ha)	Produktivitas (ku/ha)	Production (tons)
Aceh	319	56	1,735
Bali	99	60	592
Banten	324	49	1,604
Kalimantan utara	11	33	37
Kep. Bangka belitung	17	29	51
Kep. Riau	344	30	1
Lampung	507	48	2,650
Maluku	27	39	110
Maluku utara	4	37	43
Nusa tenggara barat	96	49	1,317
Sulawesi Tenggara	45	40	532
Sulawesi Utara	23	44	248
Sumatera barat	104	47	1,387

Provinsi	Rerate		
	Harvest Area (ha)	Produktivitas (ku/ha)	Production (tons)
Sumatera selatan	193	50	2,743
Sumatera utara	136	51	2,040

Based on the transformation data determined statistical values that can be taken into consideration in the determination of centroid points, in this process produced a minimum value, maximum, and average of each attribute, which is generated from the rapid miner application with the content in it is the attribute stuffing, type of data used, data missing, min, max, and average.

Table 3. Statistical Data

Atribut	Type	M	Min	Max	Average
Land	Int	0	344	1735999	254545.235
Productivity	Int	0	29	60	44.500
production	Int	0	1033	9909562	1630605.912

B. Implementation Of K-Means Algorithm

To group by utilizing the cluster test value with parameter K as much as 3 with the number of iterations stopped at round 3 with a random centroid process based on the provisions of the K-Means algorithm, the description of each centroid value, and the value for the closest distance of each iteration is outlined in the following table:

Table 4. Centroid Value

Iteration	Harvest Area (ha)			Productivity (ku/ha)			Production (tons)		
	C1	C2	C3	C1	C2	C3	C1	C2	C3
1	344	507	1.735	30	48	57	1	2.650	9.909
2	101.65	287.625	1693.67	41.434	48.652	57	355.86	2275.25	9679.66
3	101.90	266.333	1693.67	41.090	48.666	57	312.18	1268.778	967,66

Based on the centroid value that has been generated in table 4, it is continued with the calculation process to determine the distance and placement of the cluster. calculation and determination of the closest distance in iteration 1 are described in table 5.

Table 5. Closest Distance to Iteration 1

No	C1	C2	C3	Distance	Cluster
1	3,006,792	837,413	66,815,692	837,413	C2
2	349,527	4,235,772	86,808,125	349,527	C1
3	2,569,636	1,094,299	68,974,436	1,094,299	C2
4	84,982.	5,560,628	92,488,381	84,982	C1
…					
32	1,921,236	1,595,572	72,626,115	1,595,572	C2
33	7,518,716	8,963.006	51,353,098	8,963	C2
34	4,157,730	372,471	61,922,760	372,471	C2

The calculation process is performed on the second iteration and produces the shortest distance and the latest cluster with the information or knowledge described in table 6 below:

Table 6. Closest Distance to Iteration 2

No	C1	C2	C3	Distance	Cluster
1	1902219	291902.3	63119103	291902.3	C2
2	55776.34	2833520	82587280	55776.34	C1

3	1558052	450612.9	65217762	450612.9	C2
4	4138.082	3933525	88129936	4138.082	C1
…					
32	1063236	789171.7	68769910	789171.7	C2
33	5698483	218884.7	48118845	218884.7	C2
34	2836331	55494.21	58366064	55494.21	C2

the iteration resumed and the process has ceased by the rules and conditions of the K-means Algorithm stating that the calculation will stop if the group on the iteration of the process runs the same as the following group, so that in the calculation of the recommended iteration is the 3rd iteration with the description of the low production group (C1) with the dominance of the region of Bali, Bengkulu, Yogyakarta, DKI Jakarta, Gorontalo, Jambi, West Kalimantan, Central Kalimantan, East Kalimantan, North Kalimantan, kepulauan. Bangka Belitung, kep. Riau, Maluku, North Maluku, West Nusa Tenggara, East Nusa Tenggara, Papua, West Papua, Riau, West Sulawesi, Central Sulawesi, Southeast Sulawesi, North Sulawesi. Medium Production (C2) with the dominance of Aceh, Banten, South Kalimantan, Lampung, South Sulawesi, West Sumatra, South Sumatra, North Sumatra. High Production (C3) with the dominance of West Java, Central Java, and East Java.grouping description in table 7 below.

Table 7. Closest Distance to Iteration 3

No	C1	C2	C3	Distance	Cluster
1	2024629	188216.3	63119103	188216.3	C2
2	78317.35	2486396	82587280	78317.35	C1
3	1669016	319031.6	65217762	319031.6	C2
4	466.3444	3522518	88129936	466.3444	C1
5	44479.37	2708783	83846170	44479.37	C1
32	1155240	611338.8	68769910	611338.8	C2
33	5908969	329804.5	48118845	329804.5	C2
34	2985391	16714.07	58366064	16714.07	C2

The distribution of each iteration of the regions included in cluster 1, cluster 2, and cluster 3 based on the table description is shown with the following graph:

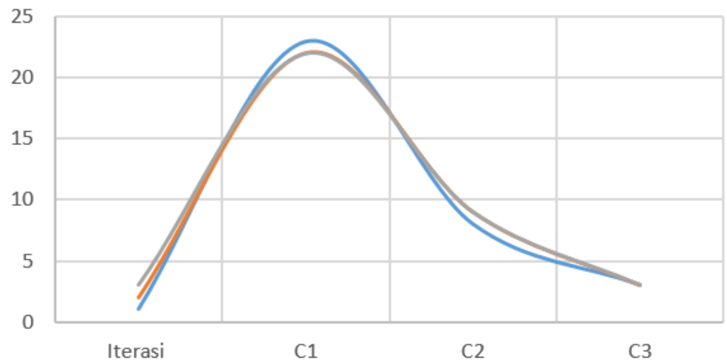

Figure 4. Data Group of Each Criterion

C. Elbow Method Utilization

Referring to the results of iterations 1 to iterations 3 by tables 5 to table 6 conducted tests to ensure the strength of each iteration by utilizing the Elbow method according to the topic discussed. The process is done by forming a graph utilizing the SSE formula. Based on the trials conducted against 34 in the first iteration resulted in a Sum of Square Errorvalue that was presented in table 8 below.

Table 8. Sum of Square Error Value On Iteration 1

No	**Distance**	**STDV**	**SSE**

1	837413.2	378718.7	1.43428E+11
2	349527.8	-109167	11917351345
3	1094299	635604.6	4.03993E+11
4	84982.37	-373712	1.39661E+11
…	…	…	…
34	372471	-86223.4	7434481555
Sum	15595611		
Mean	458694.5		
SSE			2.30204E+13

for the Sum of Square Error value in the second iteration returns a SEE value of 3.39 calculated based on the SEE value for each region, described in table 9 below:

Table 9. Sum of Square Error Value On Iteration 2

No	Distance	STDV	SSE
1	291902.3	-43612.7	1902069986
2	55776.34	-279739	78253731348
3	450612.9	115097.9	13247531064
4	4138.082	-331377	1.09811E+11
…	…	…	…
34	55494.21	-280021	0
Sum	**11407511**		
Mean	**335515**		
SSE			**3.39466E+13**

The process of calculating the SEE value continues up to the limit of the number of iterations that are instigated by the calculation of the K-Means algorithm, for this test is done only for iterations of 3 with the SEE value in table 10 follows:

Table 10. Sum of Square Error Value On Iteration 3

No	Distance	STDV	SSE
1	188216.3	-143064	20467197186
2	78317.35	-252963	63990077606
3	319031.6	-12248.3	150021875.2
4	466.3444	-330814	1.09438E+11
5	44479.37	-286801	82254573792
34	16714.07	-314566	98951691204
Sum	11263518		
Mean	331279.9		
SSE			3.97621E+13

From the calculation produced SSE value against iteration 1 with a value of 2.3 iterations 2 with a value of 3.39 and iteration 3 with a value of 3.39, the final result in testing formed SEE value of the overall iteration summarized in table 11 follows.

Table 11. SSE Values For Each Iteration

iteration	SSE
1	2.3
2	3.39
3	3.97

After the calculation process for each iteration with a SEE value is described elbow chart. By the principle of elbow clusters closest to the angle, it is used as the best cluster recommendation, in this case, is the second cluster. The cluster determination graph is described below:

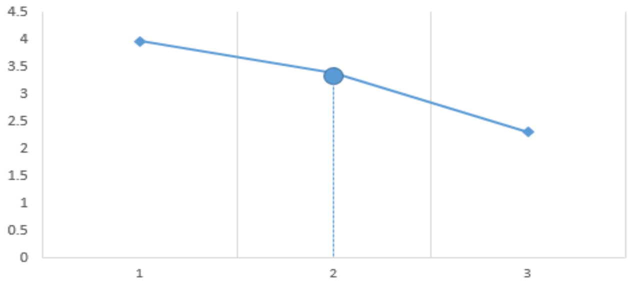

Figure 5. Elbow Chart

D. Discussion

Implementation of K-means form a data group by the specified number of clusters that is with the number of clusters as many as 3 groups, with the iteration process stopped at the third iteration because the provisions in the k-means process will stop when the group on the iteration is done the same as the previous data group, Based on the calculation, the data group produced by the production group on each iteration is iteration 1 with the data group on C1 with jumla h data as many as 23 regions with a dominance of the region namely Bali, Bengkulu, in Yogyakarta, DKI Jakarta, Gorontalo, Jambi, West Kalimantan, Central Kalimantan, East Kalimantan, North Kalimantan, Kep. Bangka belitung, Kep. Riau, Maluku, North Maluku, West Nusa Tenggara, East Nusa Tenggara, Papua, West Papua, Riau, West Sulawesi, Central Sulawesi, Southeast Sulawesi, North Sulawesi with the lowest product category. C2 with the number of data 8 Regions with the distribution of aceh, Banten, South Kalimantan, Lampung, South Sulawesi, West Sumatra, South Sumatra, North Sumatra with the category of medium or medium production and the highest production in the group C3 with the number of data as many as 3 regions with group dominance namely West Java, Central Java, East Java. In iteration 2 and iteration 3 by producing the same data group namely C1 with the number of data as many as 22 regions with the dominance of Bali, Bengkulu, in Yogyakarta, DKI Jakarta, Gorontalo, Jambi, West Kalimantan, Central Kalimantan, East Kalimantan, North Kalimantan, Kep. Bangka belitung, Kep. Riau, Maluku, North Maluku, East Nusa Tenggara, Papua, West Papua, Riau, West Sulawesi , Central Sulawesi, Southeast Sulawesi, North Sulawesi with the lowest product category, C2 as much as 9 regional data namely Aceh, Banten, South Kalimantan, West Nusa Tenggara, Lampung, South Sulawesi, West Sumatra, South Sumatra, North Sumatra and C3 as many as 3 Regions of West Java, Central Java, East Java.

From the testing of the K-means Algorithm followed by the determination of the best iteration that will be recommended against the determination of rice production group by utilizing elbow method. The optimal determination of the number of clusters in this study used one of the cluster analysis methods, the Elbow method, taking into account the comparison value (from the SSE calculation for each cluster value) between the number of clusters that would form an elbow at a point. The result of SSE value for each iteration with iteration value 1 with SEE value 2.3, iteration 2 with a value of 3.39 and iteration 3 with a value of 3.97 and information generated from the calculation of SSE provides recommendations cluster used is iteration 2 ie low product group of 22 regions, medium production as much as 9 regions and high production.

From the results of testing conducted by utilizing the K-Means algorithm and Elbow Method can complete the study of previous research that has given the calculation that the cluster in the last round is a cluster that is used as a proposed knowledge, while the results of this study provide the knowledge that clusters at each round have the same opportunity to be used as a pattern or knowledge in decision making evidenced by p New knowledge that when used K-Means Algorithm then the recommended information is iteration 3 after assessment for each iteration produced a new recommendation that the 2nd iteration is the most Optimal group. The benefit of this study is the determination of rice clusters by region can help and provide solutions to the ministry of agriculture in determining the priority of land development and can minimize the shortage of pro rice production nationally. Lack and weakness Research that has been developed for subsequent research where the clustering process only dominates against regions with a larger concept so as not to define every area that exists in a particular region.

IV. Conclusion

The implementation of the K-Means algorithm has resulted in the grouping of data according to the number of clusters as many as 3 and the process stops at the 3rd iteration, followed by the calculation of the Elbow method with SSE assessment on iteration 2.3, iteration 2 with a value of 3.39 and iteration 3 with a value of 3.97 so that at the time of forming elbow in the form of a graph and the closest to the elbow with cluster recommendations is iteration 2 which is the lowest rice production group with the number of 22 Regions, rice production with a medium category of 9 regions and production in the highest category with three regions. From the process of testing the K-Means algorithm followed by Elbow gives consideration Information that with K-Means with knowledge used is the 2nd iteration while in the implementation of Elbow knowledge given is the 2nd iteration. This research can be developed by mapping groups to areas by the cluster that has been determined.

Acknowledgment

Thank you to STMIK Pelita Nusantara for providing facilities and funding in carrying out this research, hopefully, this research is useful and can be used as a reference for the next research.

References

[1] S. Suharyanto, J. Rinaldy, and N. Ngurah Arya, "Analisis Risiko Produksi Usahatani Padi Sawah," *Agrar. J. Agribus. Rural Dev. Res.*, vol. 1, no. 2, pp. 70–77, 2015, doi: 10.18196/agr.1210.

[2] M. Maulana, "Produktifitas sebagai Sumber Pertumbuhan Padi Sawah di Indonesia 1980 – 2001," *Agro Ekon.*, vol. 22, no. 1, pp. 74–95, 2004.

[3] BPS and BPPT, *Luas Panen dan Produksi Padi di Indonesia 2019*. 2020.

[4] A. P. Windarto, "Implementation of Data Mining on Rice Imports by Major Country of Origin Using Algorithm Using K-Means Clustering Method," *Int. J. Artif. Intell. Res.*, vol. 1, no. 2, p. 26, 2017, doi: 10.29099/ijair.v1i2.17.

[5] A. Dharmarajan and T. Velmurugan, "Performance analysis on K-means and fuzzy C-means clustering algorithms using CT-DICOM images of lung cancer," *J. Adv. Res. Dyn. Control Syst.*, 2019, doi: 10.5373/JARDCS/V11/20192597.

[6] R. Annisa, D. Rosiyadi, and D. Riana, "Improved point center algorithm for k-means clustering to increase software defect prediction," *Int. J. Adv. Intell. Informatics*, vol. 6, no. 3, pp. 328–339, 2020, doi: 10.26555/ijain.v6i3.484.

[7] D. A. I. C. Dewi and D. A. K. Pramita, "Analisis Perbandingan Metode Elbow dan Silhouette pada Algoritma Clustering K-Medoids dalam Pengelompokan Produksi Kerajinan Bali," *Matrix J. Manaj. Teknol. dan Inform.*, 2019, doi: 10.31940/matrix.v9i3.1662.

[8] S. S. Subbulaxmi and G. Arumugam, "K-Means Cluster Based Oversampling Algorithm for Imbalanced Data Classification," no. 5, pp. 3436–3440, 2020, doi: 10.35940/ijrte.E6535.018520.

[9] T. M. Kodinariya and P. R. Makwana, "Review on determining number of Cluster in K-Means Clustering," *Int. J. Adv. Res. Comput. Sci. Manag. Stud.*, 2013.

[10] M. Mohammed, M. B. Khan, and E. B. M. Bashier, "k-Means Clustering," in *Machine Learning*, 2020.

[11] C. Yuan and H. Yang, "Research on K-Value Selection Method of K-Means Clustering Algorithm," *J*, 2019, doi: 10.3390/j2020016.

[12] P. M. Hasugian, H. D. Hutahaean, B. Sinaga, Sriadhi, and S. Silaban, "Villages Status Classification Analysis Involving K-Means Algorithm to Support Kementerian Desa Pembangunan Daerah Tertinggal dan Transmigrasi Work Programs," *J. Phys. Conf. Ser.*, vol. 1641, no. 1, 2020, doi: 10.1088/1742-6596/1641/1/012058.

[13] J. Karimov and M. Ozbayoglu, "Clustering Quality Improvement of k-means Using a Hybrid Evolutionary Model," 2015, doi: 10.1016/j.procs.2015.09.143.

[14] N. Putu, E. Merliana, and A. J. Santoso, "Analisa Penentuan Jumlah Cluster Terbaik pada Metode K-Means," pp. 978–979.

[15] M. A. Syakur, B. K. Khotimah, E. M. S. Rochman, and B. D. Satoto, "Integration K-Means Clustering Method and Elbow Method for Identification of the Best Customer Profile Cluster," 2018,

doi: 10.1088/1757-899X/336/1/012017.
[16] B. Purnima and K. Arvind, “EBK-Means: A Clustering Technique based on Elbow Method and K-Means in WSN,” *Int. J. Comput. Appl.*, vol. 105, no. 9, pp. 17–24, 2014, [Online]. Available: https://www.ijcaonline.org/archives/volume105/number9/18405-9674.
[17] I. Sudono, W. Utami, and S. Lestari, “Pengelompokan Produksi Padi Nasional dengan Pendekatan Data Mining Konsep K-Means,” *J. Irig.*, vol. 8, no. 2, p. 72, 2016, doi: 10.31028/ji.v8.i2.72-89.
[18] G. F. Nama, L. Hakim, and Junaidi, “Implementation of K-means technique in data mining to cluster researchers google scholar profile,” *Int. J. Eng. Adv. Technol.*, vol. 9, no. 1, pp. 3654–3660, 2019, doi: 10.35940/ijeat.A2708.109119.
[19] A. Chaerudin, D. T. Murdiansyah, and M. Imrona, “Implementation of K-Means ++ Algorithm for Store Customers Segmentation Using Neo4J,” vol. 6, no. April, pp. 53–60, 2021, doi: 10.34818/indojc.2021.6.1.547.
[20] B. P. Statistik, “Badan Pusat Statistik,” *Luas Panen, Produksi, dan Produktivitas Padi Menurut Provinsi 2018-2020*, 2021. https://www.bps.go.id/indicator/53/1498/1/luas-panen-produksi-dan-produktivitas-padi-menurut-provinsi.html.
[21] P. V Sankarganesh and P. Sripriya, “Fuzzy Bat based Cluster Center Selection Algorithm (FBCCSA) Improved K-Means Algorithm for Type 2 Diabetes Mellitus Prediction,” *Int. J. Recent Technol. Eng.*, vol. 8, no. 4S5, pp. 128–133, 2020, doi: 10.35940/ijrte.d1003.1284s519.
[22] U. R. Raval and C. Jani, “Implementing & Improvisation of K-means Clustering Algorithm,” *Int. J. Comput. Sci. Mob. Comput.*, 2016.
[23] E. Muningsih and S. Kiswati, “Sistem Aplikasi Berbasis Optimasi Metode Elbow Untuk Penentuan Clustering Pelanggan,” *Joutica*, vol. 3, no. 1, p. 117, 2018, doi: 10.30736/jti.v3i1.196.