Journal of Theoretical and Applied Information Technology
15th October 2021. Vol.99. No 19

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

BIG DATA TECHNOLOGY FOR VILLAGE STATUS CLASSIFICATION BASED ON VILLAGE INDEX BUILDING INVOLVING K-MEANS ALGORITHM IN PROGRAMS TO SUPPORT THE WORK OF THE MINISTRY OF THE VILLAGE

[1]PASKA MARTO HASUGIAN, [2]HARVEI DESMON HUTAHAEAN, [1]BOSKER SINAGA, [2]SRIADHI, [3]SARONOM SILABAN

[1]Informatics Engineering Study Program, STMIK Pelita Nusantara, Medan, Indonesia

[2]Computer Science and Informatics Education Study Program, Universitas Negeri Medan, Indonesia

[3]Departement of chemistry, Universitas Negeri Medan, indonesia

E-mail: [1]paskamarto86@gmail.com, [2]harvei.hutahaean@gmail.com, [1]boskersinaga@gmail.com, [2]sriadhi@unimed.ac.id, [3]saronomsilaban@unimed.ac.id

ABSTRACT

The problem that is used as part of the research is village status data which is only used and utilized in the current year and has been going on since 2014 so that there is an accumulation of data in the database and no in-depth analysis has been carried out to obtain information related to village status. This research aims to analyze the village status group using the k-means algorithm and provide the best cluster information using the Elbow Method by finding the SSE (sum of squared-error) value by utilizing the cluster closest to the elbow on the graph. The Ministry of Villages has described the method of solving problems using secondary data through the website https://idm.kemendesa.go. Id/ followed by the application of the K-means algorithm and determining the best cluster. The results obtained in the study are a comparison of sets from the village ministry with cluster information provided by the K-Means algorithm, namely the status of independent villages has increased by 761 villages, developed villages from 202 experienced an increase in data of 1095 data, growing villages with the K-Means algorithm decreased. With a difference of 1150 villages, underdeveloped villages decreased based on the K-Means Algorithm with a difference of 1158, and very disadvantaged villages increased up to a difference of 452. Testing with the Elbow Method provides information and offers the best cluster for grouping village status. The number of groups is four groups with an independent position, Forward, Develop, lag.

Keywords: *Status Village, K-Means, Elbow, SSE*

1. INTRODUCTION

One of the duties and responsibilities of the village ministry that has been running since 2014 and has been ongoing is the classification of villages based on values developing village index (IDM). The IDM is the basis for village development which is the basis for assessing village progress and independence[1]. The process of determining the IDM with an assessment of several sectors, namely Ecological Resilience (IKL), Economic Resilience Index (IKE), and Social Resilience Index (IKS)[2]. The combination of the three sectors is a combination that will determine the quality of the village. The assessment process is carried out to get the IDM value by forming a team that immediately goes into the field by assessing the three aspects; then, it will be calculated and adjusted to the predetermined classification range[3]. The resulting data will be directed towards the type of classification following the applicable values and provisions with the categories of independent villages, developed villages, developing villages, underdeveloped villages, and very underdeveloped villages[4]. Based on the annual process, there is a considerable accumulation of data with thousands of villages throughout Indonesia stored in the database from 2014 to 2020. No research has been conducted the data mining. From several related studies used as the basis for developing big data with the topic of Clustering, Dewi Pramudi Isma has used K-means to classify data with a feature selection process to reduce the computational load by reducing the size of high-dimensional data. Selecting a feature subset

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

representing all the features used by removing irrelevant data and selecting one element representing a redundant number of features[5]. Agus Perdana's research utilizes the k-means algorithm to support the grouping of rice imported by the leading country of origin by dividing 4 clusters with supporting countries in the world[6]. A. Dharmarajan compares the use of algorithms for grouping between K-means and C-Means with a review of the final results obtained will help as part of decision making[7]. Some of these studies have carried out the grouping process by utilizing the k-means algorithm, but these studies do not offer the most optimal number of clusters or the best group.

Based on this description, the researcher conducted data analysis to explore the potential and knowledge of a collection of village status databases in Indonesia by grouping the data so that new knowledge was generated and provided the best cluster information that could be used as a comparison to village status and was useful as input and determination of village status in next year[8] the excavation process by utilizing knowledge of data mining, namely Clustering with the working principle of dividing data into groups that have objects with the same characteristics[9]–[11]. Clustering is grouping data items into a small number of groups so that each group has something in common[12],[13]. Clustering plays an important role in data mining applications, such as scientific data exploration, information access, text mining, spatial database applications, and web analysis[14]. While the method used to support these activities is K-Means. K-means and its variants are a type of partition-based clustering algorithm that has been widely used in data clustering. K-means groups the data set into k clusters based on the shortest distance between the data and the cluster center[15]. The process of determining the grouping using the K-Means Cluster Analysis algorithm as a solution for classifying the characteristics of objects[16]. The reason for using the K-Means algorithm is because this algorithm has a fairly high accuracy of object size, so this algorithm is relatively more scalable and efficient for processing large numbers of objects[15]. K-Means is one part of the Clustering technique by utilizing unsupervised techniques by partitioning data into two or more groups[17]–[20]. With the basic concept of data getting closer to the center of the cluster by calculation, then the data is included in the predetermined cluster category[21]-[22],[23]. The process in K-Means randomly selects several cluster centers according to the specified number of clusters[17],[18]. In each iteration, the membership of the data to the new cluster center is calculated. The process will stop if the cluster center and data membership do not change[23]. The formula used to determine the classification of the pattern with the description 1) Determine k as the number of clusters to be formed 2). Initialization of the initial k centroids (cluster center points) randomly. 3) Allocate each data or object to the nearest cluster[24]–[26]. The distance between objects and the distance between objects with a certain cluster is determined by the distance between the data and the center of the cluster. To calculate the distance of all data to each cluster center, using the euclidean distance theory. D(i,j)= . With the provisions, D(i,j) = distance of data I to the center of cluster j, Xki = data to I on attribute data to k, Xkj = center point of cluster j on attribute to k The distance of cluster center is recalculated with the current cluster membership. The cluster's center is the average of all data or objects in a particular cluster; if desired, the median value of the cluster can also be used $\sqrt{(x_{Ki}-x_{Kj})^2+(x_{Ki}-x_{Kj})+\cdots+(x_{ki}-x_{kj})^2}$ [27],[28]. With the results of the grouping, measurements will be made of the most optimal cluster that will be offered by utilizing the Elbow method by finding the SSE value[29],[30].

The Elbow method is used to select the optimal number of clusters or groups based on the sum of square error (SSE) using the formula SSE= where K is the number of groups used in the K-MeansXi algorithm is the number of data and Ck is the number of clusters in the k cluster. From the grouping that is developed, the more qualified and able to maximize the more dominant group. The results of different percentages of each cluster value can be shown by using a graph as a source of information. If the value of the first cluster with the weight of the second cluster gives the angle in the graph or the value has decreased the most, then the value of the group is the best $SSE = \sum_{k=1}^{K}\sum_{xi \in sk}]]Xi - Ck]]^2$[31]–[33].

2. RELATED WORK

Manoj Kumar Gupta, Pravin Chandra, with the research topic MP-K-Means: Modified Partition Based Cluster Initialization Method for K-Means Algorithm, stated that the performance and accuracy of K-means are influenced by the selection of the initial cluster centroid by finding the Cluster Initialization Method Based on Modified Partitions for k-means (MP-k-means). In MP-k-means, the data dimensions are partitioned so that if 'd' is the data dimension, then a list of 'd' consisting of equal-sized partitions 'k's based on the mean of the positions is created[34].

Kaile Zhou, Shanlin Yang with the topic Effect of cluster size distribution on Clustering: a comparative study of k-means and fuzzy means Clustering with the main focus on the effect of data distribution on Clustering and presents a comparative analysis of k-means Clustering and FCM with the emphasis that Experiments Extensive analysis of synthetic datasets and real-world datasets show that FCM has more potent uniform effects than k-means[19].

The research of S. Santha Subbulaxmi, G. Arumugam with the topic of K-Means Cluster-Based Undersampling Ensemble for Imbalanced Data Classification with a research study that the K-Means cluster-based undersampling ensemble algorithm is proposed to overcome the problem of unbalanced data classification. The proposed method combines the undersampling and boosting methods based on the K-Means cluster. The experimental results show that the proposed algorithm outperforms other ensemble sampling algorithms from previous studies[35].

Research by PV Sankar Ganesh, P. Sripriya with Fuzzy Bat-based Cluster Center Selection Algorithm (FBCCSA) Improved K-Means Algorithm for Type 2 Diabetes Mellitus Prediction proposes an algorithm to perform clusters with improved K-Means algorithm with cluster center selection (CCS) selection. And Logistic Regression (LR) algorithms In the first level, a K-means algorithm enhanced by FBCCSA is proposed to remove incorrectly clustered data[17].

Riski Annisa et al., with the publication title, Improved point center algorithm for K-Means clustering to increase software defect prediction, provide a discussion that the proposed algorithm overcomes random centroid values in k-means and then applies it to predict software defect module errors. The point center algorithm is proposed to determine the initial centroid value for optimization of the k-means algorithm. These findings are helpful and contribute to developing clustering models to handle data, such as to predict software defect modules more accurately[36].

Salvatore Leonardi, Natalia Distefano, Giulia Pulvirenti, with the topic Identification Of Road Safety Measures For Elderly Pedestrians Based On K-Means Clustering And Hierarchical Cluster Analysis, suggested that Hierarchical Clustering and K-Means are used to explore which solutions are proposed by elderly pedestrians to improve pedestrian safety. foot[37].

3. PROPOSED METHODOLOGY

This research focuses on group data and providing the best group information from the groups formed. For the grouping process by utilizing the K-means algorithm and providing information on the best cluster or the most optimal cluster by using the Elbow method. The stages of the process carried out are 1) Data collection by summarizing the data that has been recorded in the village ministry database in IDM management 2) Preprocessing stages to ensure data suitability for calculations using the K-Means algorithm, 3) Grouping based on algorithms is tested to provide exposure the best cluster and describe the description of the grouping as part of knowledge 4) Utilization of the Elbow method. Flowchart K-Means algorithm by following the following flow:

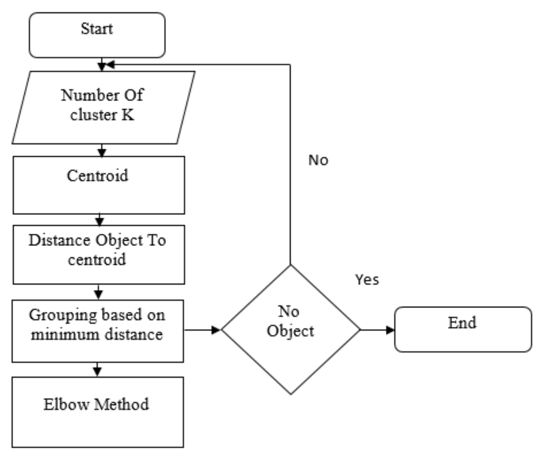


Figure 1. Flowchart the K-Means algorithm

To ensure the process of grouping village status, relevant data is needed. The data is generated from the village ministry's website page with management data devoted to the IDM (Development Village Index) on pagehttps://idm.kemendesa.go.id/. The data used as test data is 2019 data with a village status group of 5439 villages spread across North Sumatra with the data group in the following table.

Table 1. Grouping of Village Status in 2019.

Village Status	Number of Villages
Independent Village	4
Advanced Village	202
Developing Village	2443
Disadvantaged Village	2040
Very Disadvantaged Village	727
Total	**5416**

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

The grouping of villages follows the rules and regulations that the Ministry of Villages PDTT has issued with a table description [3].

Table 2. Village Status and Threshold Value.

Village Status	Value
Independent Village	Village Build Index is greater (>) than 0.8155.
Advanced Village	The Village Index Build is less and equal to (≤) 0.8155 and greater (>) than 0.7072.
Developing Village	The Develop Village Index is less and equal to (≤) 0.7072 and greater (>) than 0.5989.
Disadvantaged Village	The Develop Village Index is less and equal to (≤) 0.5989 and greater (>) than 0.4907.
Very Disadvantaged Village	Village Build Index less and smaller (≤) than 0.4907.

At the preprocessing stage in determining village status based on the village development index with three main supporting attributes, namely IKS (Social Resilience Index), IKE (Economic Resilience Index), IKS (Environmental Resilience Index), and IDM (Development Village Index) with the amount of data to be tested as many as 5416. The status that has been defined by IDM data (Development Village Index) will be transformed into a status code, namely Independent Village (1), Advanced Village (3), Developing Village (3), Disadvantaged Village (4), Very Disadvantaged Village (5). The description of the data used is described below:

Table 3. Data Processing

No	IKS	IKE	IKL	IDM
1	0.7029	0.5167	0.6	0.606533
2	0.7029	0.45	0.6	0.5843
3	0.7829	0.6333	0.6	0.672067
4	0.6171	0.5	0.6667	0.5946
5	0.6343	0.45	0.6	0.561433
6	0.68	0.45	0.6	0.576667
7	0.6114	0.4167	0.6	0.5427
8	0.6	0.3667	0.6667	0.544467
9	0.4743	0.3833	0.6	0.485867
10	0.5886	0.5	0.6	0.562867
11	0.5657	0.35	0.5333	0.483
12	0.5829	0.4833	0.6667	0.577633
13	0.5486	0.4333	0.6667	0.549533
14	0.4971	0.3667	0.6	0.487933
15	0.68	0.55	0.6	0.61
…	0.5029	0.4167	0.6	0.506533
5410	0.52	0.5	0.8667	0.6289
5411	0.6343	0.4167	0.8	0.617
5412	0.7029	0.5333	0.6667	0.6343
5413	0.5657	0.5333	0.8667	0.655233
5414	0.5429	0.35	0.6667	0.519867
5415	0.5886	0.35	0.6	0.512867
5416	0.5486	0.3333	0.6667	0.5162

Statistics from the data used with IKS, IKE, and IKL attributes are descriptions of data types, missing data with statistical details with IDM min, IDM max, and the average of IDM. The report is in table 3 below:

Table 4. Data statistics based on IDM.

Name	Type	Missing	Statistic		
			Min	Max	Average
IKS	Real	0	0.303	0.971	0.658
IKE	Real	0	0.150	0.950	0.456
IKL	Real	0	0	1	0.637

4. EXPERIMENT AND RESULTS

4.1 Grouping With K-Means Algorithm

To perform the clustering process using the K-Means algorithm, the initialization or determination of the cluster number is determined as the initial cluster center. Determination of the number of clusters with several cluster studies with parameter values (K) from cluster 2 to the number of sets as many as 10. The process of determining the center point randomly in the first iteration and subsequent iterations is according to the data in the same group. The distribution process of The total number of data, the iteration used with a maximum number of 10 rounds with the basic principle iteration, will stop if the values in the previous round have the same data group. The results of the clustering process in 2 clusters up to 10 sets with the number of iterations each with the following description:

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 5. Center Points and Cluster Data Groups 2.

I	Center Points						Data groups	
	C1			C2			C1	C2
	IKS	IKE	IKL	IKS	IKE	IKL		
1	0.3657	0.4333	0	0.8857	0.7	1	4350	1066
2	0.528881	0.340937	0.555684	0.689823	0.494901	0.657125	1669	3747
3	0.708683	0.526034	0.645813	0.544785	0.326762	0.617789	1876	3540
4	0.712907	0.533239	0.646532	0.554898	0.323062	0.599516	3500	1916
5	0.713174	0.534776	0.647864	0.557709	0.32172	0.593681	3480	1936
6	0.713374	0.477107	0.577026	0.558955	0.321512	0.591129	3622	1794
7	0.711637	0.289614	0.351159	0.550241	0.321283	0.609491	341	5075
8	0.71826	0.345796	0.299898	0.654138	0.296282	0.445056	197	5219
9	0.744081	0.393995	0.263615	0.654933	0.2961	0.442599	264	5152
10	0.762298	0.38201	0.26843	0.65284	0.295548	0.444784	250	5166

Table 6. Center Points and Cluster Data Groups 3.

I	Center Points									Data groups		
	C1			C2			C3			C1	C2	C3
	IKS	IKE	IKL	IKS	IKE	IKL	IKS	IKE	IKL			
1	0.36	0.43	0	0.70	0.55	0.6	0.88	0.7	1	39	5271	106
2	19.71	13.23	11.53	19.71	13.23	11.53	0.65	0.45	0.63	0	4749	667

Table 7. Center Points and Cluster Data Groups 4.

I	Center Points												Data groups			
	C1			C2			C3			C4			C1	C2	C3	C4
	IKS	IKE	IKL	IKS	IKE	IKL	IKS	IKE	IKL	IKS	IKE	IKL				
1	0,37	0,43	0	0,46	0,33	0,53	0,74	0,58	0,67	0,89	0,70	1	8	1841	3507	60
2	0,53	0,48	0,16	1,93	1,37	1,87	0,71	0,53	0,65	0,84	0,75	0,76	226	0	4861	329

Table 8. Center Points and Cluster Data Groups 5.

I	Center Points						Data groups				
	C1			C2..C5			C1	C2	C3	C4	C5
	IKS	IKE	IKL	IKS	IKE	IKL					
1	0,37	0,43	0	0,57	0,38	0,47	3	2531	637	2034	211
2	0,41	0,42	0,16	0,58	0,36	0,62	30	2169	703	2172	342
3	0,49	0,38	0,30	0,57	0,35	0,63	116	1853	816	2203	428
4	0,52	0,36	0,39	0,56	0,34	0,64	203	1646	956	2093	518
5	0,53	0,36	0,43	0,56	0,33	0,64	301	1456	1053	2002	604
6	0,53	0,37	0,48	0,55	0,32	0,65	514	1198	1145	1889	670
7	0,53	0,40	0,55	0,55	0,30	0,65	741	1045	1186	1720	724
8	0,55	0,41	0,60	0,54	0,28	0,63	949	959	1224	1538	746
9	0,57	0,41	0,62	0,53	0,28	0,63	1084	913	1268	1393	758
10	0,58	0,41	0,62	0,52	0,28	0,62	1179	882	1293	1297	765

Table 9. Center Points and Cluster Data Groups 6.

I	Center Points						Data groups					
	C1			C2..C6			C1	C2	C3	C4	C5	C6
	IKS	IKE	IKL	IKS	IKE	IKL						
1	0,37	0,43	0	0,47	0,38	0,60	10	1442	2480	1353	127	4
2	0,53	0,39	0,15	0,53	0,33	0,62	56	1446	2278	1393	216	27
3	0,56	0,39	0,30	0,53	0,32	0,63	149	1366	2139	1449	235	78
4	0,57	0,40	0,40	0,53	0,31	0,64	351	1220	1976	1455	244	170

I	Center Points						Data groups					
	C1			C2..C6			C1	C2	C3	C4	C5	C6
	IKS	IKE	IKL	IKS	IKE	IKL						
5	0,57	0,42	0,51	0,53	0,30	0,64	709	1068	1690	1399	255	295
6	0,57	0,42	0,59	0,53	0,29	0,63	922	985	1467	1358	272	412
7	0,57	0,42	0,61	0,52	0,28	0,63	1026	950	1287	1342	287	524
8	0,58	0,42	0,62	0,52	0,28	0,62	1082	931	1188	1283	313	619
9	0,58	0,42	0,62	0,52	0,28	0,62	1097	921	1128	1257	340	673
10	0,59	0,41	0,63	0,52	0,28	0,62	1120	902	1080	1230	358	726

Table 10. Center Points and Cluster Data Groups 7.

I	Center Points						Data groups						
	C1			C2..C7			C1	C2	C3	C4	C5	C6	C7
	IKS	IKE	IKL	IKS	IKE	IKL							
1	0,37	0,43	0	0,54	0,18	0,47	10	100	1264	3126	475	395	46
2	0,47	0,42	0,24	0,51	0,25	0,46	57	322	1241	2189	874	475	258
3	0,52	0,42	0,37	0,49	0,23	0,58	164	468	1140	1724	984	485	451
4	0,55	0,42	0,45	0,48	0,24	0,62	337	541	998	1474	1001	472	593
5	0,54	0,43	0,53	0,48	0,25	0,62	490	587	899	1326	964	438	712
6	0,55	0,43	0,59	0,49	0,26	0,62	587	621	824	1241	953	412	778
7	0,55	0,43	0,61	0,49	0,26	0,62	651	634	782	1177	944	398	830
8	0,55	0,43	0,61	0,49	0,26	0,62	689	646	739	1144	966	387	845
9	0,56	0,43	0,62	0,49	0,27	0,62	716	648	713	1418	1281	326	314
10	0,56	0,43	0,62	0,49	0,27	0,62	799	647	773	1199	1178	502	318

Table 11. Center Points and Cluster Data Groups 8.

I	Center Points						Data groups							
	C1			C2..C8			C1	C2	C3	C4	C5	C6	C7	C8
	IKS	IKE	IKL	IKS	IKE	IKL								
1	0,37	0,43	0	0,45	0,20	0,27	5	37	93	853	2707	1575	120	26
2	0,45	0,44	0,20	0,53	0,30	0,32	33	96	185	929	2257	1655	191	70
3	0,51	0,46	0,35	0,55	0,31	0,40	149	130	278	908	1943	1575	227	206
4	0,53	0,46	0,48	0,58	0,30	0,43	376	165	355	821	1601	1493	202	394
5	0,55	0,46	0,57	0,59	0,30	0,47	533	248	74	1101	1333	1394	195	538
6	0,57	0,47	0,61	0,59	0,28	0,51	742	355	160	827	1208	1315	188	621
7	0,59	0,48	0,62	0,61	0,29	0,56	880	497	257	628	1026	1272	198	658
8	0,60	0,48	0,63	0,61	0,31	0,60	905	626	348	522	883	1247	211	674
9	0,60	0,48	0,63	0,61	0,33	0,62	891	674	432	481	837	1187	225	689
10	0,61	0,49	0,63	0,61	0,33	0,63	945	679	474	468	794	1121	242	693

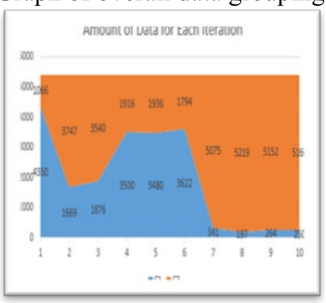
Table 12. Center Points and Cluster Data Groups 9.

I	Center Points						Data groups								
	C1			C2..C9			C1	C2	C3	C4	C5	C6	C7	C8	C9
	IKS	IKE	IKL	IKS	IKE	IKL									
1	0,37	0,43	0	0,44	0,35	0,13	1	10	241	140	1252	2716	977	63	16
2	0,37	0,43	0	0,45	0,39	0,33	1	41	518	228	1015	2293	1129	138	53
3	0,37	0,43	0	0,48	0,41	0,42	2	172	530	463	958	1742	1220	175	154
4	0,40	0,39	0,07	0,51	0,41	0,55	7	342	541	723	859	1308	1133	161	342
5	0,50	0,34	0,15	0,52	0,41	0,60	38	456	539	839	762	1129	1043	157	453
6	0,54	0,33	0,28	0,53	0,41	0,62	82	533	530	909	690	1004	946	158	564
7	0,55	0,24	0,25	0,54	0,41	0,63	48	598	545	951	675	910	913	169	607
8	0,55	0,28	0,33	0,54	0,41	0,62	87	622	540	994	639	854	868	180	632
9	0,55	0,30	0,39	0,55	0,41	0,63	126	627	537	1026	602	833	845	191	629
10	0,55	0,31	0,43	0,55	0,42	0,63	166	647	519	1018	584	818	823	197	644

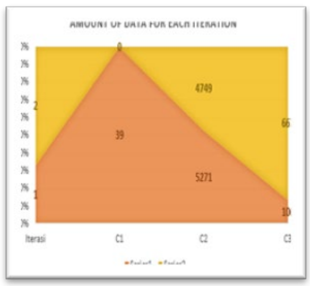
Table 13. Center Points and Cluster Data Groups 10.

I	Center Points						Data groups									
	C1			C2..C10			C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
	IKS	IKE	IKL	IKS	IKE	IKL										
1	0,37	0,43	0	0,63	0,25	0,20	4	14	196	187	1052	2631	1165	73	82	12
2	0,49	0,51	0,12	0,66	0,28	0,36	30	116	848	323	1068	1432	1044	330	162	63
3	0,55	0,49	0,28	0,64	0,29	0,49	96	252	875	413	905	1046	871	643	181	134
4	0,57	0,49	0,40	0,64	0,30	0,57	138	373	860	471	748	890	753	820	199	164
5	0,57	0,48	0,43	0,63	0,30	0,60	146	436	866	488	683	824	712	875	196	190
6	0,57	0,48	0,44	0,63	0,31	0,62	170	494	860	497	626	779	699	877	203	211
7	0,57	0,48	0,48	0,63	0,32	0,63	217	534	837	504	568	764	679	882	211	220
8	0,56	0,49	0,53	0,63	0,32	0,63	271	561	810	505	530	741	681	871	229	217
9	0,56	0,49	0,57	0,63	0,33	0,63	315	575	769	491	523	742	679	856	251	215
10	0,56	0,49	0,60	0,63	0,33	0,63	350	592	740	476	518	727	672	841	281	219

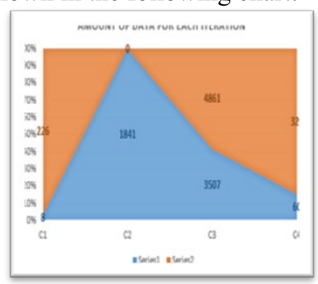
Graph of overall data grouping with a value of K = 2 to a value of K = 10 as shown in the following chart:



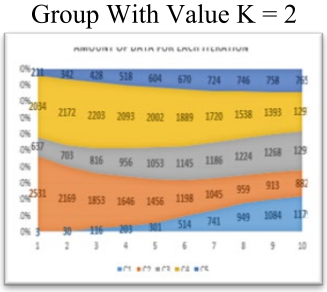
Group With Value K = 2



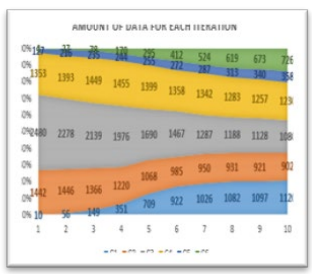
Group With Value K = 3



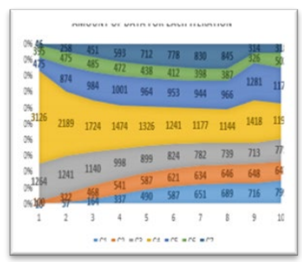
Group With Value K = 4



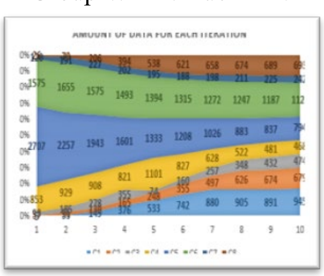
Group With Value K = 5



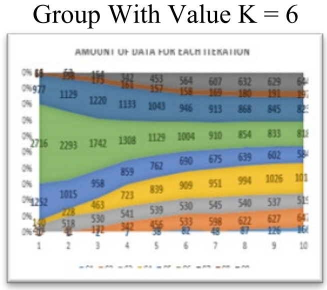
Group With Value K = 6



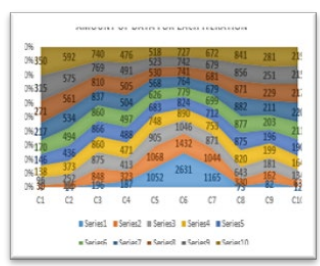
Group With Value K = 7



Group With Value K = 8



Group With Value K = 9



Group With Value K = 10

Figure 2. Graph of Village Status Distribution Based on K = 2-10 Parameters

The results of the clustering process in 2 Clusters up to 10 Clusters with the number of iterations each with the following description:

1. Grouping with a value of K = 2 utilizes ten iterations with the data group in C1, totaling 4350 villages and C2 destroying 1066 villages. The second iteration, in a new group as membership, namely C1 as many as 1669 villages, and C2 as many as 3747 villages. A third iteration is new group membership, namely C1 as many as 1876 villages, and C2 as many as 3540 villages. The fourth iteration in a

new group as membership, namely C1 as many as 3500 villages, and C2 as many as 1916 villages. A fifth iteration is a new group membership, namely C1 as many as 3480 towns, and C2 as many as 1936 villages. The sixth iteration produces a new group as membership, namely C1 as many as 3622 villages, and C2 as many as 1794 villages. The seventh iteration makes a new group as membership, namely C1 as many as 341 villages, and C2 as many as 5075 villages. The eighth iteration in a new group as membership, namely C1 as many as 197 towns and C2 as many as 5219 villages. The ninth iteration, in a new group as membership, namely C1 as many as 264 cities, and C2 as many as 5152 towns. The tenth iteration produces a new group as membership, namely C1 as many as 250 villages, and C2 as many as 5166 villages.

2. The grouping with a value of K = 3 utilizes two iterations with the data group in Cluster 1 (C1) as many as 39 villages, in Cluster 2 (C2) with the number of village data as much as 5271 while the data group in Cluster 3 (C3) with a total of 106 data. The second iteration produces a new group as membership, namely cluster 1 does not have a membership, in Cluster 2 with a total of 4749 data, in C3 with a total of 667.
3. The grouping with a value of K = 4 utilizes two iterations with the data group in C1 as many as eight villages, C2 as many as 1841 villages, C3 as 3507 villages, while the data group in C4 as many as 60 villages. The second iteration produces new groups as membership, namely C1 as many as 226 villages, C2 does not have a membership, C3 as many as 4861 towns, and C4 as many as 329 villages.
4. The grouping with a value of K = 5 utilizes ten iterations with data groups in C1 as many as 3 villages, C2 as 2531 villages, C3 as many as 637 villages, C4 as many as 2034 villages, while the data group in C5 as many as 211 villages. The second iteration resulted in a new group as membership, namely C1 as many as 30 villages, C2 as many as 2169 villages, C3 as many as 703 villages, C4 as many as 2172 villages, while the data group in C5 as many as 342 villages. Villages, C2 as many as 1853 villages, C3 as many as 816 villages, C4 as many as 2203 villages, while the data group in C5 as many as 428 villages. The fourth iteration resulted in new groups as membership, namely C1 as many as 203 villages, C2 as many as 1646 villages, C3 as many as 956 villages, C4 as many as 2093 villages, while the data group in C5 is 518 villages. In the fifth iteration, new groups as membership were produced, namely C1 as many as 301 villages, C2 as many as 1456 villages, C3 as many as 1053 villages, C4 as many as 2002 villages, while the data group in C5 as many as 604 villages. The sixth iteration resulted in new groups as membership, namely C1 as many as 514 villages, C2 as many as 1198 villages, C3 as many as 1145 villages, C4 as many as 1889 villages, while the data group in C5 as many as 670 villages. The seventh iteration resulted in new groups as membership, namely C1 as many as 741 villages, C2 as many as 1045 villages, C3 as many as 1186 villages, C4 as many as 1720 villages, while the data group in C5 as many as 724 villages. The eighth iteration resulted in new groups as membership, namely C1 as many as 949 villages, C2 as many as 959 villages, C3 as many as 1224 villages, C4 as many as 1538 villages, while the data group in C5 as many as 746 villages. The ninth iteration resulted in new groups as membership, namely C1 as many as 1084 villages, C2 as many as 913 villages, C3 as many as 1268 villages, C4 as many as 1393 villages, while the data group in C5 as many as 758 villages. The tenth iteration resulted in new groups as membership, namely C1 as many as 1179 villages, C2 as many as 882 villages, C3 as many as 1293 villages, C4 as many as 1297 villages, while the data group in C5 as many as 765 villages.
5. The grouping with a value of K = 6 utilizes ten iterations with data groups in C1 as many as ten villages, C2 as many as 1442 villages, C3 as many as 2480 villages, C4 as many as 1353 villages, C5 as 127 villages, while the data group in C6 as many as four villages. In the second iteration, it produces new groups as membership, namely C1 as many as 56 villages, C2 as many as 1446 villages, C3 as many as 2278 villages, C4 as many as 1393 villages, C5 as many as 216 villages, while the data group in C6 as many as 27 villages. Membership is C1 as many as 149 villages, C2 as many as 1366 villages, C3 as many as 2139 villages, C4 as many as 1449 villages, C5 as many as 235 villages, while the data group in C6 as many as 78 villages. The fourth iteration resulted in new groups as membership, namely C1 as many as 351 villages, C2 as many as 1220 villages, C3 as many as 1976 villages, C4 as many as 1455 villages, C5 as many as 244 villages, while the data group in C6 as many as 170 villages. In the fifth iteration, new groups were formed as

membership, namely C1 as many as 709 villages, C2 as many as 1068 villages, C3 as many as 1690 villages, C4 as many as 1399 villages, C5 as many as 255 villages, while the data group in C6 as many as 295 villages. In the sixth iteration, new groups as membership were produced, namely C1 as many as 922 villages, C2 as many as 985 villages, C3 as many as 1467 villages, C4 as many as 1358 villages, C5 as many as 272 villages, while the data group in C6 as many as 412 villages. In the seventh iteration, new groups as membership were produced, namely C1 as many as 1026 villages, C2 as many as 950 villages, C3 as many as 1287 villages, C4 as many as 1342 villages, C5 as many as 287 villages, while the data group in C6 as many as 524 villages. The eighth iteration resulted in new groups as membership, namely C1 as many as 1082 villages, C2 as many as 931 villages, C3 as many as 1188 villages, C4 as many as 1283 villages, C5 as many as 313 villages, while the data group in C6 as many as 619 villages. The ninth iteration resulted in new groups as membership, namely C1 as many as 1097 villages, C2 as many as 921 villages, C3 as many as 1128 villages, C4 as many as 1257 villages, C5 as many as 340 villages, while the data group in C6 as many as 673 villages. The tenth iteration resulted in new groups as membership, namely C1 as many as 1120 villages, C2 as many as 902 villages, C3 as many as 1080 villages, C4 as many as 1230 villages, C5 as many as 358 villages, while the data group in C6 as many as 211 villages. In comparison, the data group in C6 is 619 villages. The ninth iteration resulted in new groups as membership, namely C1 as many as 1097 villages, C2 as many as 921 villages, C3 as many as 1128 villages, C4 as many as 1257 villages, C5 as many as 340 villages, while the data group in C6 as many as 673 villages. The tenth iteration resulted in new groups as membership, namely C1 as many as 1120 villages, C2 as many as 902 villages, C3 as many as 1080 villages, C4 as many as 1230 villages, C5 as many as 358 villages, while the data group in C6 as many as 211 villages. In comparison, the data group in C6 is 619 villages. The ninth iteration resulted in new groups as membership, namely C1 as many as 1097 villages, C2 as many as 921 villages, C3 as many as 1128 villages, C4 as many as 1257 villages, C5 as many as 340 villages, while the data group in C6 as many as 673 villages. The tenth iteration resulted in new groups as membership, namely C1 as many as 1120 villages, C2 as many as 902 villages, C3 as many as 1080 villages, C4 as many as 1230 villages, C5 as many as 358 villages, while the data group in C6 as many as 211 villages.

6. The grouping with a value of K=7 utilizes ten iterations with data groups in C1 as many as 100 villages, C2 as 1264 villages, C4 as many as 3126 villages, C5 as many as 475 villages, C6 as many as 395 while the data group in C7 as many as 46 villages. The second iteration resulted in new groups as membership, namely C1 as many as villages, C2 as many as 322 villages, C3 as many as 1241 villages, C4 as many as 2189 villages, C5 as many as 874 villages, C6 as many as 475 while the data group in C7 as many as 258 villages. In the third iteration, it produces new groups as membership, namely C1 as many as villages, C2 as many as 468 villages, C3 as many as 1140 villages, C4 as many as 1724 villages, C5 as many as 984 villages, C6 as many as 485 while the data group in C7 as many as 451 villages. The fourth iteration produces a new group as membership, namely C1 as many as villages, C2 as many as 541 villages, C3 as many as 998 villages, C4 as many as 1474 villages, C5 as many as 1001 villages, C6 as many as 472 while the data group in C7 as many as 593 villages. In the fifth iteration, new groups as membership were produced, namely C1 as many as villages, C2 as many as 587 villages, C3 as many as 899 villages, C4 as many as 1326 villages, C5 as many as 964 villages, C6 as many as 438 while the data group in C7 was 712 villages. The sixth iteration resulted in new groups as membership, namely C1 as many as villages, C2 as many as 621 villages, C3 as many as 824 villages, C4 as many as 1241 villages, C5 as many as 953 villages, C6 as many as 412 while the data group in C7 was 778 villages. The seventh iteration resulted in new groups as membership, namely C1 as many as villages, C2 as many as 634 villages, C3 as many as 782 villages, C4 as many as 1177 villages, C5 as many as 944 villages, C6 as many as 398 while the data group in C7 as many as 830 villages. The eighth iteration resulted in new groups as membership, namely C1 as many as villages, C2 as many as 646 villages, C3 as many as 739 villages, C4 as many as 1144 villages, C5 as many as 966 villages, C6 as many as 387 while the data group in C7 as many as 845 villages. The ninth iteration resulted in new groups as membership,

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

namely C1 as many as villages, C2 as many as 648 villages, C3 as many as 713 villages, C4 as many as 1418 villages, C5 as many as 1281 villages, C6 as many as 326 while the data group in C7 as many as 314 villages. The tenth iteration resulted in new groups as membership, namely C1 as many as villages, C2 as many as 647 villages, C3 as many as 773 villages, C4 as many as 1199 villages, C5 as many as 1178 villages, C6 as many as 502 while the data group in C7 as many as 318 villages. The eighth iteration resulted in new groups as membership, namely C1 as many as villages, C2 as many as 646 villages, C3 as many as 739 villages, C4 as many as 1144 villages, C5 as many as 966 villages, C6 as many as 387 while the data group in C7 as many as 845 villages. The ninth iteration resulted in new groups as membership, namely C1 as many as villages, C2 as many as 648 villages, C3 as many as 713 villages, C4 as many as 1418 villages, C5 as many as 1281 villages, C6 as many as 326 while the data group in C7 as many as 314 villages. The tenth iteration resulted in new groups as membership, namely C1 as many as villages, C2 as many as 647 villages, C3 as many as 773 villages, C4 as many as 1199 villages, C5 as many as 1178 villages, C6 as many as 502 while the data group in C7 as many as 318 villages. The eighth iteration resulted in new groups as membership, namely C1 as many as villages, C2 as many as 646 villages, C3 as many as 739 villages, C4 as many as 1144 villages, C5 as many as 966 villages, C6 as many as 387 while the data group in C7 as many as 845 villages. The ninth iteration resulted in new groups as membership, namely C1 as many as villages, C2 as many as 648 villages, C3 as many as 713 villages, C4 as many as 1418 villages, C5 as many as 1281 villages, C6 as many as 326 while the data group in C7 as many as 314 villages. The tenth iteration resulted in new groups as membership, namely C1 as many as villages, C2 as many as 647 villages, C3 as many as 773 villages, C4 as many as 1199 villages, C5 as many as 1178 villages, C6 as many as 502 while the data group in C7 as many as 318 villages. C2 is 646 villages, C3 is 739 villages, C4 is 1144 villages, C5 is 966 villages, C6 is 387, while the data group in C7 is 845 villages. The ninth iteration resulted in new groups as membership, namely C1 as many as villages, C2 as many as 648 villages, C3 as many as 713 villages, C4 as many as 1418 villages, C5 as many as 1281 villages, C6 as many as 326 while the data group in C7 as many as 314 villages. The tenth iteration resulted in new groups as membership, namely C1 as many as villages, C2 as many as 647 villages, C3 as many as 773 villages, C4 as many as 1199 villages, C5 as many as 1178 villages, C6 as many as 502 while the data group in C7 as many as 318 villages. C2 is 646 villages, C3 is 739 villages, C4 is 1144 villages, C5 is 966 villages, C6 is 387, while the data group in C7 is 845 villages. The ninth iteration resulted in new groups as membership, namely C1 as many as villages, C2 as many as 648 villages, C3 as many as 713 villages, C4 as many as 1418 villages, C5 as many as 1281 villages, C6 as many as 326 while the data group in C7 as many as 314 villages. The tenth iteration resulted in new groups as membership, namely C1 as many as villages, C2 as many as 647 villages, C3 as many as 773 villages, C4 as many as 1199 villages, C5 as many as 1178 villages, C6 as many as 502 while the data group in C7 as many as 318 villages. The ninth iteration resulted in new groups as membership, namely C1 as many as villages, C2 as many as 648 villages, C3 as many as 713 villages, C4 as many as 1418 villages, C5 as many as 1281 villages, C6 as many as 326 while the data group in C7 as many as 314 villages. The tenth iteration resulted in new groups as membership, namely C1 as many as villages, C2 as many as 647 villages, C3 as many as 773 villages, C4 as many as 1199 villages, C5 as many as 1178 villages, C6 as many as 502 while the data group in C7 as many as 318 villages. The ninth iteration resulted in new groups as membership, namely C1 as many as villages, C2 as many as 648 villages, C3 as many as 713 villages, C4 as many as 1418 villages, C5 as many as 1281 villages, C6 as many as 326 while the data group in C7 as many as 314 villages. The tenth iteration resulted in new groups as membership, namely C1 as many as villages, C2 as many as 647 villages, C3 as many as 773 villages, C4 as many as 1199 villages, C5 as many as 1178 villages, C6 as many as 502 while the data group in C7 as many as 318 villages.

7. The grouping with a value of K = 8 utilizes ten iterations with data groups in C1 as many as five villages, C2 as 37 villages, C3 as many as 93 villages, C4 as many as 853 villages, C5 as many as 2707 villages, C6 as many as 1575 villages, C7 as many as 120 while the data group in C8 as many as 26 villages. In the second iteration resulted in new groups as membership, namely C1 as many as 33 villages,

C2 as many as 96 villages, C3 as many as 185 villages, C4 as many as 929 villages, C5 as many as 2257 villages, C6 as many as 1655 villages, C7 as many as 191 while the data group in C8 as many as 70 villages. . In the third iteration resulted in new groups as membership, namely C1 as many as 149 villages, C2 as many as 130 villages, C3 as many as 278 villages, C4 as many as 908 villages, C5 as many as 1943 villages, C6 as many as 1575 villages, C7 as many as 227 while the data group in C8 as many as 206 villages. . In the fourth iteration resulted in new groups as membership, namely C1 as many as 376 villages, C2 as many as 165 villages, C3 as many as 355 villages, C4 as many as 821 villages, C5 as many as 1601 villages, C6 as many as 1493 villages, C7 as many as 202 while the data group in C8 as many as 394 villages. . In the fifth iteration resulted in new groups as membership, namely C1 as many as 533 villages, C2 as many as 248 villages, C3 as many as 74 villages, C4 as many as 1101 villages, C5 as many as 1333 villages, C6 as many as 1394 villages, C7 as many as 195 while the data group in C8 as many as 538 villages. . In the sixth iteration resulted in new groups as membership, namely C1 as many as 742 villages, C2 as many as 355 villages, C3 as many as 160 villages, C4 as many as 827 villages, C5 as many as 1208 villages, C6 as many as 1315 villages, C7 as many as 188 while the data group in C8 as many as 621 villages. . In the seventh iteration resulted in new groups as membership, namely C1 as many as 880 villages, C2 as many as 497 villages, C3 as many as 257 villages, C4 as many as 628 villages, C5 as many as 1026 villages, C6 as many as 1272 villages, C7 as many as 198 while the data group in C8 as many as 658 villages. . In the eighth iteration resulted in new groups as membership, namely C1 as many as 905 villages, C2 as many as 626 villages, C3 as many as 348 villages, C4 as many as 522 villages, C5 as many as 883 villages, C6 as many as 1247 villages, C7 as many as 211 while the data group in C8 as many as 674 villages. . In the ninth iteration resulted in new groups as membership, namely C1 as many as 891 villages, C2 as many as 674 villages, C3 as many as 432 villages, C4 as many as 481 villages, C5 as many as 837 villages, C6 as many as 1187 villages, C7 as many as 225 while the data group in C8 as many as 689 villages .

8. The grouping with a value of K=9 utilizes two iterations with data groups in C1 as many as one village, C2 as many as ten villages, C3 as many as 241 villages, C4 as many as 140 villages, C5 as many as 1252 villages, C6 as many as 2716 villages, C7 as many as 977 villages, C8 as many as 63 while the data group in C9 is 16 villages. The second iteration resulted in a new group as membership, namely C1 as many as one village, C2 as many as 41 villages, C3 as many as 518 villages, C4 as many as 228 villages, C5 as many as 1015 villages, C6 as many as 2293 villages, C7 as many as 1129 villages, C8 as many as 138 while the data group in C9 as many as 53 villages. The third iteration resulted in new groups as membership, namely C1 as many as two villages, C2 as many as 172 villages, C3 as many as 530 villages, C4 as many as 463 villages, C5 as many as 958 villages, C6 as many as 1742 villages, C7 as many as 1220 villages, C8 is 175 while the data group in C9 is 154 villages. The fourth iteration resulted in new groups as membership, namely C1 as many as seven villages, C2 as many as 342 villages, C3 as many as 541 villages, C4 as many as 723 villages, C5 as many as 859 villages, C6 as many as 1308 villages, C7 as many as 1133 villages, C8 as many as 161 while the data group in C9 as many as 342 villages. The fifth iteration resulted in new groups as membership, namely C1 as many as 38 villages, C2 as many as 456 villages, C3 as many as 539 villages, C4 as many as 839 villages, C5 as many as 762 villages, C6 as many as 1129 villages, C7 as many as 1043 villages, C8 as many as 157 while the data group in C9 as many as 453 villages. The sixth iteration resulted in new groups as membership, namely C1 as many as 82 villages, C2 as many as 533 villages, C3 as many as 530 villages, C4 as many as 909 villages, C5 as many as 690 villages, C6 as many as 1004 villages, C7 as many as 946 villages, C8 as many as 158 while the data group in C9 as many as 564 villages. The seventh iteration resulted in new groups as membership, namely C1 as many as 48 villages, C2 as many as 598 villages, C3 as many as 545 villages, C4 as many as 951 villages, C5 as many as 675 villages, C6 as many as 910 villages, C7 as many as 913 villages, C8 as many as 169 while the data group in C9 as many as 607 villages. The eighth iteration resulted in new groups as membership, namely C1 as many as 87 villages, C2 as many as 622 villages, C3 as many as 540 villages, C4 as

many as 994 villages, C5 as many as 639 villages, C6 as many as 854 villages, C7 as many as 868 villages, C8 as many as 180 while the data group in C9 as many as 632 villages. The ninth iteration resulted in new groups as membership, namely C1 as many as 126 villages, C2 as many as 627 villages, C3 as many as 537 villages, C4 as many as 1026 villages, C5 as many as 602 villages, C6 as many as 833 villages, C7 as many as 845 villages, C8 as many as 191 while the data group in C9 as many as 629 villages. The tenth iteration resulted in new groups as membership, namely C1 as many as 166 villages, C2 as many as 647 villages, C3 as many as 519 villages, C4 as many as 1018 villages, C5 as many as 584 villages, C6 as many as 818 villages, C7 as many as 823 villages, C8 as many as 197 while the data group in C9 as many as 644 villages.

9. The grouping with a value of K = 10 utilizes two iterations with data groups in C1 as many as four villages, C2 as 14 villages, C3 as many as 196 villages, C4 as many as 187 villages, C5 as many as 1052 villages, C6 as many as 2631 villages, C7 as many as 1165 villages, C8 as many as 73 villages, C9 as many as 82 while the data group in C10 as many as 12 villages. The second iteration resulted in new groups as membership, namely C1 as many as 30 villages, C2 as many as 116 villages, C3 as many as 848 villages, C4 as many as 323 villages, C5 as many as 1068 villages, C6 as many as 1432 villages, C7 as many as 1044 villages, C8 as many as 330 villages, C9 as many as 162 while the data group in C10 as many as 63 villages. The third iteration resulted in new groups as membership, namely C1 as many as 96 villages, C2 as many as 252 villages, C3 as many as 875 villages, C4 as many as 413 villages, C5 as many as 905 villages, C6 as many as 1046 villages, C7 as many as 871 villages, C8 as many as 643 villages, C9 as many as 181 while the data group in C10 as many as 134 villages. The fourth iteration resulted in new groups as membership, namely C1 as many as 138 villages, C2 as many as 373 villages, C3 as many as 860 villages, C4 as many as 471 villages, C5 as many as 748 villages, C6 as many as 890 villages, C7 as many as 753 villages, C8 as many as 820 villages, C9 as many as 199 while the data group in C10 as many as 164 villages. The fifth iteration resulted in new groups as membership, namely C1 as many as 146 villages, C2 as many as 436 villages, C3 as many as 866 villages, C4 as many as 488 villages, C5 as many as 683 villages, C6 as many as 824 villages, C7 as many as 712 villages, C8 as many as 875 villages, C9 as many as 196 while the data group in C10 as many as 190 villages. The sixth iteration resulted in a new group as membership, namely C1 as many as 170 villages, C2 is 494 villages, C3 is 860 villages, C4 is 497 villages, C5 is 626 villages, C6 is 779 villages, C7 is 699 villages, C8 is 877 villages, C9 is 203. In comparison, the data group in C10 is 211 villages. The seventh iteration resulted in new groups as membership, namely C1 as many as 217 villages, C2 as many as 534 villages, C3 as many as 837 villages, C4 as many as 504 villages, C5 as many as 568 villages, C6 as many as 764 villages, C7 as many as 679 villages, C8 as many as 882 villages, C9 as many as 211 while the data group in C10 as many as 220 villages. The eighth iteration resulted in new groups as membership, namely C1 as many as 271 villages, C2 as many as 561 villages, C3 as many as 810 villages, C4 as many as 505 villages, C5 as many as 530 villages, C6 as many as 741 villages, C7 as many as 681 villages, C8 as many as 871 villages, C9 as many as 229 while the data group in C10 as many as 217 villages. The ninth iteration resulted in new groups as membership, namely C1 as many as 315 villages, C2 as many as 575 villages, C3 as many as 769 villages, C4 as many as 491 villages, C5 as many as 523 villages, C6 as many as 742 villages, C7 as many as 679 villages, C8 as many as 856 villages, C9 as many as 251 while the data group in C10 as many as 215 villages. The tenth iteration resulted in new groups as membership, namely C1 as many as 350 villages, C2 as many as 592 villages, C3 as many as 740 villages, C4 as many as 476 villages, C5 as many as 518 villages, C6 as many as 727 villages, C7 as many as 672 villages, C8 as many as 841 villages, C9 as many as 281 while the data group in C10 as many as 219 villages. C7 as many as 679 villages, C8 as 856 villages, C9 as many as 251, while the data group in C10 as many as 215 villages. The tenth iteration resulted in new groups as membership, namely C1 as many as 350 villages, C2 as many as 592 villages, C3 as many as 740 villages, C4 as many as 476 villages, C5 as many as 518 villages, C6 as many as 727 villages, C7 as many as 672 villages, C8 as many as 841 villages, C9 as many as 281 while the data group in C10 as many as 219 villages. C7 as

Journal of Theoretical and Applied Information Technology
15th October 2021. Vol.99. No 19

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

many as 679 villages, C8 as 856 villages, C9 as many as 251, while the data group in C10 as many as 215 villages. The tenth iteration resulted in new groups as membership, namely C1 as many as 350 villages, C2 as many as 592 villages, C3 as many as 740 villages, C4 as many as 476 villages, C5 as many as 518 villages, C6 as many as 727 villages, C7 as many as 672 villages, C8 as many as 841 villages, C9 as many as 281 while the data group in C10 as many as 219 villages.

4.2 Determining the Best Cluster

The process of determining the best cluster uses the Elbow Method by finding the SSE (sum of squared-error) value with information the closer to the elbow in the graph, the better the Clustering. The SSE value for the entire cluster from the cluster with a value of K=2 to K=10 is described in the following table.

Table 14. Sum of Squared Error Table with K=2 Value.

No	IKS	IKE	IKL	Distance	STD	SSE
1	0.7029	0.5167	0.6	0.274783912	-0.02782	0.000773679
2	0.7029	0.45	0.6	0.224618403	-0.07798	0.006080973
3	0.7829	0.6333	0.6	0.393806978	0.091208	0.008318896
4	0.6171	0.5	0.6667	0.303849852	0.001251	1.56464E-06
5	0.6343	0.45	0.6	0.219752566	-0.08285	0.006863531
5412	0.7029	0.5333	0.6667	0.329057336	0.026458	0.000700044
5413	0.5657	0.5333	0.8667	0.492069763	0.189471	0.035899171
5414	0.5429	0.35	0.6667	0.253571808	-0.04903	0.002403665
5415	0.5886	0.35	0.6	0.176589589	-0.12601	0.015878371
5416	0.5486	0.3333	0.6667	0.2480686	-0.05453	0.002973564
Sum				1638.876		
Mean					0.302599	
SSE						48.95251035

Table 15. Sum of Squared Error Table with K=4 Value.

No	IKS	IKE	IKL	Distance	STDV	SSE
1	0.7029	0.5167	0.6	0.015624669	-0.13805	0.019057
2	0.7029	0.45	0.6	0.079920487	-0.07375	0.005439
3	0.7829	0.6333	0.6	0.129913961	-0.02376	0.000564
4	0.6171	0.5	0.6667	0.098490981	-0.05518	0.003045
5	0.6343	0.45	0.6	0.111190226	-0.04248	0.001805
5412	0.7029	0.5333	0.6667	0.010645236	-0.14303	0.020457
5413	0.5657	0.5333	0.8667	0.194650036	0.040977	0.001679
5414	0.5429	0.35	0.6667	0.245018244	0.091345	0.008344
5415	0.5886	0.35	0.6	0.217724762	0.064052	0.004103
5416	0.5486	0.3333	0.6667	0.253693243	0.10002	0.010004
Sum				832.2923198		
Mean					0.15367288	
SSE						51.35904

Table 16 Sum of Squared Error Table with K = 10.

No	IKS	IKE	IKL	Distance	STDV	SSE
1	0.7029	0.5167	0.6	0.050222	-0.00898	8.06E-05
2	0.7029	0.45	0.6	0.042728	-0.01647	0.000271
3	0.7829	0.6333	0.6	0.075259	0.016057	0.000258
4	0.6171	0.5	0.6667	0.054193	-0.00501	2.51E-05
5	0.6343	0.45	0.6	0.010413	-0.04879	0.00238
5412	0.7029	0.5333	0.6667	0.036482	-0.02272	0.000516
5413	0.5657	0.5333	0.8667	0.116994	0.057792	0.00334
5414	0.5429	0.35	0.6667	0.023505	-0.0357	0.001274
5415	0.5886	0.35	0.6	0.04867	-0.01053	0.000111
5416	0.5486	0.3333	0.6667	0.040502	-0.0187	0.00035
Sum				320.6359		
Mean					0.059202	
SSE						6.305328

Based on the SSE value of each cluster, the overall value is generated from the value of K = 2 to the value of K = 10 with the description in the following table:

Table 17. Sum of Squared Error.

Nilai K	SSE
2	48.95251
3	29.2833
4	51.35904
5	9.770379
6	8.210143
7	7.947435
8	7.466088
9	6.251741
10	6.305328

With the SSE chart with the following image:

Figure 3. SSE Graph

From the village status cluster analysis that has been developed, a comparison process is carried out on the data groups that the village ministry has poured with a total of five village status groups with the results of the application of the K-means algorithm with clusters with a value of K = 5, group descriptions in the following table:

Table 18. Group Descriptions In The Following

Village Status	Village Status Group	
	Ministry of Village	K-Means
Independent Village	4	765
Advanced Village	202	1297
Developing Village	2443	1293
Underdeveloped Village	2040	882
Very Underdeveloped Village	727	1179
Total	5416	5416

Based on the description of the table, the difference between the calculation of k-means and the selection of the value of K=5 has a very significant difference where the calculation of K-means provides information according to the iteration used to produce a new status, namely the difference in the level of independent villages reaching an increase of 761 villages, advanced villages from 202 experienced an increase in data of 1095 data, developing villages with the K-Means algorithm decreased by a difference of 1150 villages, underdeveloped villages fell based on the K-Means algorithm with a difference of 1158. Very underdeveloped villages increased up to a difference of 452. Graph of grouping and contrast with the following chart this :

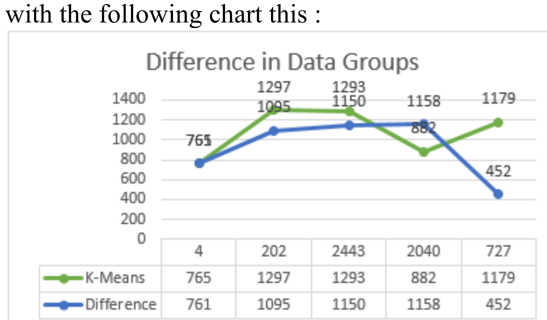


Figure 4. Group Difference Image With K-Means Algorithm

By utilizing the elbow method, the SEE values for each Clustering are found sequentially with parameter values 2 to 10 are 48.95251, 29.2833, 51.35904, 9.770379, 8.210143, 7.947435, 7.466088, 6.251741, 6.305328, the SSE value is used as part of the graph formation. It provides information that the value closest to the elbow with the SSE value is 51.35904, so that the information provided through the use of the elbow method is that the 4th cluster is the best.

5. CONCLUSION

Following the tests and discussions that have been described in the research, the k-means algorithm can group village status by experimenting with cluster selection starting from the number of clusters 2 to the number of sets 10. The comparison obtained is based on the village status of the village ministry of village groups with the k algorithm. -means by selecting the value of K = 5 with various groups and the difference in the value of independent village status reaching an increase of 761 villages, developed villages from 202 experiencing an increase in data by 1095 data, developing villages using the K-Means algorithm decreased by a difference of 1150 villages,

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

underdeveloped villages decreased based on the K-Means Algorithm with a difference of 1158. Very underdeveloped villages increased to a difference of 452. Testing with the Elbow Method provides information and offers the best cluster for grouping village status. The number of groups is four groups with an independent status, Advanced, Developing, lagging.

ACKNOWLEDGEMENT

The researcher would like to thank the Ministry of Education, Culture, Research, and Technology. In particular, RISTEK-BRIN facilitated the implementation of this research.

REFERENCES:

[1] 2015 Kementrian Desa. Pembangunan Daerah Tertinggal dan Transmigrasi, "Indeks Desa Membangun 2015," *Kementrian Desa. Pembang. Drh. Tertinggal dan Transm.*, 2015.

[2] 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. https://doi.org/10.1088/1742-6596/1641/1/012058

[3] D. J. P. dan P. M. D. Kementerian Desa Pembangunan Daerah Tertinggal dan Transmigrasi, *Indeks Desa Membangun (IDM)*. 2020.

[4] S.- Suroso, "Kebijakan Pembangunan Desa Tertinggal Berbasis Indeks Desa Membangun (IDM) dan Potensi Lokal," *J. Litbang Media Inf. Penelitian, Pengemb. dan IPTEK*, vol. 16, no. 1, pp. 47–62, 2020. https://doi.org/10.33658/jl.v16i1.167

[5] D. P. Ismi, S. Panchoo, and Murinto, "K-means clustering based filter feature selection on high dimensional data," *Int. J. Adv. Intell. Informatics*, 2016. https://doi.org/10.26555/ijain.v2i1.54

[6] 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. https://doi.org/10.29099/ijair.v1i2.17

[7] 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. https://doi.org/10.5373/JARDCS/V11/20192597

[8] W. A. Günther, M. H. Rezazade Mehrizi, M. Huysman, and F. Feldberg, "Debating big data: A literature review on realizing value from big data," *J. Strateg. Inf. Syst.*, 2017. https://doi.org/10.1016/j.jsis.2017.07.003

[9] E. Rendón, I. Abundez, A. Arizmendi, and E. M. Quiroz, "Internal versus External cluster validation indexes," *Int. J. Comput. Commun.*, 2011.

[10] S. Rani and G. Sikka, "Recent Techniques of Clustering of Time Series Data: A Survey," *Int. J. Comput. Appl.*, 2012. https://doi.org/10.5120/8282-1278

[11] K. Hammouda, "A Comparative Study of Data Clustering Techniques," *Int. J. Res. Comput. Sci.*, 2011.

[12] P. Rai and S. Singh, "A Survey of Clustering Techniques," *Int. J. Comput. Appl.*, 2010. https://doi.org/10.5120/1326-1808

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

[14] L. Waltman, N. J. van Eck, and E. C. M. Noyons, "A unified approach to mapping and clustering of bibliometric networks," *J. Informetr.*, 2010. https://doi.org/10.1016/j.joi.2010.07.002

[15] D. Meister and J. Bittner, "Parallel BVH construction using k-means clustering," *Vis. Comput.*, vol. 32, no. 6–8, pp. 977–987, 2016. https://doi.org/10.1007/s00371-016-1241-0

[16] W. Hardi, W. A. Kusuma, and S. Basuki, "Clustering topic groups of documents using K-Means algorithm: Australian Embassy Jakarta media releases 2006-2016," *Berk. Ilmu Purpose. Dan Inf.*, vol. 15, no. 2, p. 226, 2019. https://doi.org/10.22146/bip.36451

[17] 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. https://doi.org/10.35940/ijrte.D1003.1284S519

[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. https://doi.org/10.35940/ijeat.A2708.109119

[19] K. Zhou and S. Yang, "Effect of cluster size distribution on clustering: a comparative study of k-means and fuzzy c-means clustering," *Pattern Anal. Appl.*, vol. 23, no. 1, pp. 455–466, 2020. https://doi.org/10.1007/s10044-019-00783-6

[20] J. Li, Y. Horiguchi, and T. Sawaragi, "Cluster Size-Constrained Fuzzy C-Means with Density Center Searching," *Int. J. Fuzzy Log. Intell. Syst.*, vol. 20, no. 4, pp. 346–357, 2020. https://doi.org/10.5391/IJFIS.2020.20.4.346

[21] L. Morissette and S. Chartier, "The k-means clustering technique: General considerations and implementation in Mathematica," *Tutor. Quant. Methods Psychol.*, vol. 9, no. 1, pp. 15–24, 2013. https://doi.org/10.20982/tqmp.09.1.p015

[22] P. M. Hasugian, H. D. Hutahaean, and B. Sinaga, "Design of big data technology prototype for classification of village status based on village development index involve k-means algorithm to support village ministry Pdtt work programs Design of big data technology prototype for classification of the village," 2021. https://doi.org/10.1088/1742-6596/1811/1/012012

[23] C. Series, "Review the Utilization of Big Data and K-Means Algorithm in Supporting The Determination of Village Status As Support To The Ministry of Village PDTT Review the Utilization of Big Data and K-Means Algorithm in Supporting The Determination of Village Statu," 2021.

[24] S. Zhong, "Efficient online spherical k-means clustering," in *Proceedings of the International Joint Conference on Neural Networks*, 2005.

[25] B. Purnima and K. Arvind, "EBK-Means: A Clustering Technique based on Elbow Method and K-Means in WSN," *Int. J. Comput. Appl.*, 2014.

[26] I. Bin Mohamad and D. Usman, "Standardization and its effects on K-means clustering algorithm," *Res. J. Appl. Sci. Eng. Technol.*, 2013. https://doi.org/10.19026/rjaset.6.3638

[27] M. Verma, M. Srivastava, N. Chack, A. K. Diswar, and N. Gupta, "A Comparative Study of Various Clustering Algorithms in Data Mining," *Int. J. Eng. Res. Appl. www.ijera.com*, 2012.

[28] T. Velmurugan, "Efficiency of k-Means and K-Medoids Algorithms for Clustering Arbitrary Data Points," *Int. J. Comput. …*, 2012.

[29] J. D'Silva and U. Sharma, "Unsupervised Automatic Text Summarization of Konkani Texts using K-means with Elbow Method," *Int. J. Eng. Res. Technol.*, vol. 13, no. 9, pp. 2380–2384, 2020. https://doi.org/10.37624/IJERT/13.9.2020.2380-2384

[30] H. Khalid and V. Wade, "Topic detection from conversational dialogue corpus with parallel latent Dirichlet allocation model and elbow method," *arXiv*. 2020. https://doi.org/10.5121/csit.2020.100508

[31] X. Li, S. Liu, Y. Chang, S. Li, Y. Fan, and H. Yu, "A Human Joint Torque Estimation Method for Elbow Exoskeleton Control," *Int. J. Humanoid Robot.*, vol. 17, no. 3, 2020. https://doi.org/10.1142/S0219843619500397

[32] W. Wang, D. Wang, M. Wesseling, B. Xue, and F. Li, "Comparison of modeling and tracking methods for analyzing elbow and forearm kinematics," *Proc. Inst. Mech. Eng. Part H J. Eng. Med.*, vol. 233, no. 11, pp. 1113–1121, 2019. https://doi.org/10.1177/0954411919872400

[33] D. Aggarwal and D. Sharma, "Application of clustering for student result analysis," *Int. J. Recent Technol. Eng.*, vol. 7, no. 6, pp. 50–53, 2019.

[34] M. K. Gupta and P. Chandra, "MP-K-Means: Modified Partition Based Cluster Initialization Method for K-Means Algorithm," *Int. J. Recent Technol. Eng.*, vol. 8, no. 4, pp. 1140–1148, 2019. https://doi.org/10.35940/ijrte.D6837.118419

[35] S. S. Subbulaxmi and G. Arumugam, "K-Means Cluster-Based Undersampling Ensemble for Imbalanced Data Classification," *Int. J. Eng. Adv. Technol.*, vol. 9, no. 3, pp. 2074–2079, 2020. https://doi.org/10.35940/ijeat.C5188.029320

[36] 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. https://doi.org/10.26555/ijain.v6i3.484

[37] S. Leonardi, N. Distefano, and G. Pulvirenti, "Identification of road safety measures for elderly pedestrians based on k-means clustering and hierarchical cluster analysis," *Arch. Transp.*, vol. 56, no. 4, pp. 107–118, 2020. https://doi.org/10.5604/01.3001.0014.5630