Priority Program Selection of Village Fund Using the K-Means Method

Ati Sulastri[a,1,*], Sisti Khalifah [a2], Dhea Lestari [a,3], Dudih Gustian [a,4], Muhamad Muslih [a,5], Kirishchieva Irina Rafaelevna [b,6]

[a,]Nusa Putra University, Jl. Raya Cibolang Kaler No. 21, Kab. Sukabumi 43152, Sukabumi, Indonesia
[b] Rostov State Transport University, Rostovskogo Strelkovogo Polka Narodnogo Opolcheniya Sq., 2, Rostov-on-Don, 344038
[1]ati.sulastri_si17@nusaputra.ac.id*; [2] Siti.khalifah_si17@nusaputra.ac.id; [3] Dhea.Lestari17@nusaputra.ac.id, [4]dudih@nusaputra.ac.id, [5] muslih@nusaputra.ac.id, [6]Irina_raf@list.ru
* Corresponding Author

Received 06 November 2020 ; revised 10 November 2020 ; accepted 12 November 2020

ABSTRACT

The allocation of funds is certainly a big village also demands a big responsibility for the village government. In managing public funds, a public sector organization is required to be able to provide accountable financial reports. To achieve the effectiveness of village financial management, a decision support system is needed to assist village officials in determining which village programs will be prioritized. This research focuses on village programs that use village fund allocations. The purpose of this study is to create a decision support system to determine priority village fund programs using a web-based kmeans clustering algorithm. The k-means clustering algorithm can perform the modeling process without supervision (unsupervised) and is one of the methods that performs data grouping using the partition system. This is in accordance with the desired end result in the form of grouping the village work program into 3 priority levels, where the village road repair program is included in the high-level priority program in cluster 1, the BUMDES program is included in the medium level priority in cluster 2 and the program for the construction of a reservoir for incoming rainwater. in low priority cluster 3. This research was conducted by considering the aspects of urgency and usefulness developed using the programming language PHP version 5 and MySQL database version 3.2.1.

KEYWORDS
Decision Support Systems
K-Means Clustering
Village Programs
Village Funds

1. Introduction

The Village Fund is a fund sourced from the State Revenue and Expenditure Budget allocated for Villages and Traditional Villages which is then transferred through the district / city Regional Revenue and Expenditure Budget and will be used to finance the implementation of village government programs. The view of Rosalinda et al [1]. Dana Village is by the Village Fund which put emphasis on the development of rural communities, is expected to push the handling of some of the problems faced by rural communities independently without having a long wait for the programs of the district government [2].

Funding is a big village would require responsibility which is also big for the village government. The central government allocated village funds in 2015 amounting to IDR 20.67 trillion, in 2016 IDR 46.98 trillion, in 2017 IDR 60 trillion, in 2018 it was 60 trillion and in 2019 IDR 70 trillion for all villages spread across Indonesia [3].

Regulation of the Minister of Home Affairs (Permendagri) number 113 of 2014 concerning village financial management explains that village financial management is all activities that include planning, implementation, administration, reporting, and village financial accountability. Village financial management, managed within 1 (one) fiscal year from January 1 to December 31 [4].

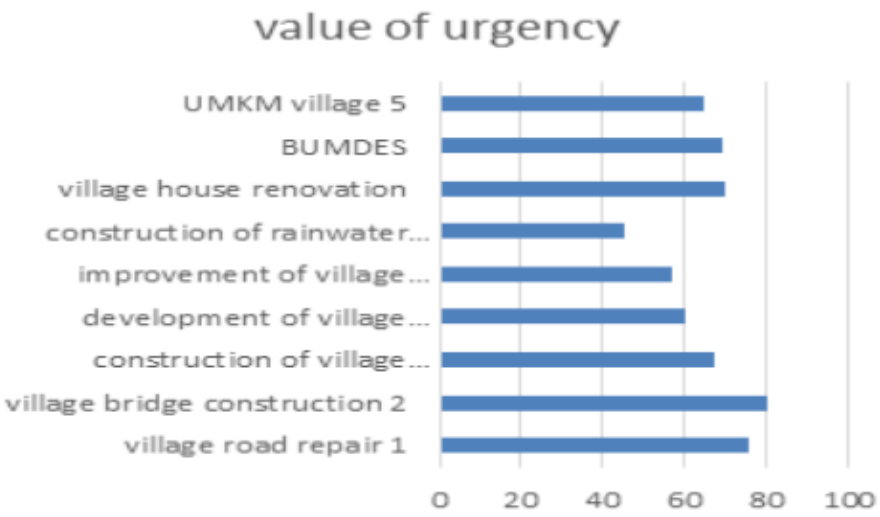

Fig. 1. Figure 1. The Urgency Value of Village Fund

Distribution xyz From Figure 1 above there is a problem that the allocation of village funds is not very good. For example, the construction of rainwater tanks which should be prioritized but has received less attention, even though the conditions in the area often flood when the rainfall is high. On the other hand, the existing road and bridge infrastructure is still in good condition, but still prioritized. This is reinforced by the results of observations with the community in general that the distribution of village funds has so far been less important. This means that the distribution of village funds that occurs is not in accordance with the existing priority scale, such as an example of development that is not important, around 65% is still carried out by the local government, which means that the distribution of Village Funds in XYZ Village is too large for development that is not important, while what is very important can be called neglected. This is shown in Figure 2 below.

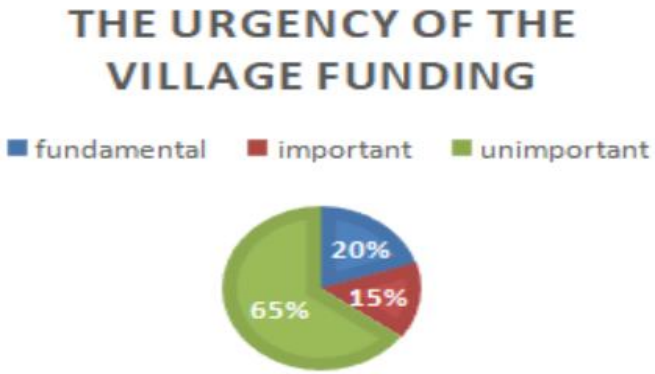

Fig. 2. Figure 2. The Urgency Value of Village Fund Donors XYZ

Clustering k-means with 5 attributes and 3 clusters, namely the results of the first cluster of 46 population data with a percentage of 21%, the second cluster with 143 data results. population and a percentage of 66%, and the third cluster with 28 population data results with a percentage of 13%. (Noviana et al., 2019) [5]. The value of the silhouette coeffcient for validating data from clustering with KMedoids results in village / kelurahan areas in Kukar district based on indicators of underdeveloped villages with 2 clusters with a value of 0.430 which states that the resulting cluster structure in this grouping is weak structure [6].

Data grouping uses the fuzzy c-means method, which generates random numbers as the initial partition matrix, calculates the cluster center, calculates the objective function and calculates the change in each partition matrix. The iteration stops when the conditions are met, after which the cluster center is obtained. Each cluster will be sorted based on the proximity of the data element to the center of the cluster to obtain a ranking [7]. This research provides a solution. In classifying the prioritized village program to be realized which is divided into 3 priority scales, namely the level of high, medium and low needs. So that there is no mistake in allocating village funds in accordance with the priority cluster degrees of existing needs. In addition, research also provides benefits for village officials to be more focused in allocating village funds according to the existing priority scale, so that the funds channeled by the Government to the Village are allocated properly and directed.

Said abdul aziz, Sarjon Defit, Yuhandri Yunus (2021). Clusterization of aid funds in the family hope program (PKH)uses the k-means method. Object. In this study, the researchers used a system analysis with the k-means methodfor clustering the distribution of funds for the Hope Family Program (PKH) based on expectations into 3 groups, namely the C1 group which is the near-poor household category group (RTHM), C2 is the house category group. poor households (RTM), and C3 is a very poor household category group (RTSM) [8]. Tia Noviana, Jasmir, Yudi Novianto (2018). Application of data mining determines the priority groups of recipients of RASTRA rice assistance with k-means clustering. In this research, the researcher produced a priority assessment system for the poor population that uses k-means clustering data mining which can be used as a decision support system tool in determining priority population groups who receive prosperous rice assistance (RASTRA) [9].

Reasonable evaluation of building energy performance can provide great benefit for government (policy makers) to put forward effective energy conservation policies, and building regulators to make suitable energy-saving strategies. Traditional approaches employ the annual total energy consumption (or the annual total energy use intensity) to evaluate and rank the energy performance of buildings. However, these approaches ignore the energy consumption fluctuation of buildings in different seasons. As a consequence, this paper proposes an improved data-driven-based building energy performance evaluating and ranking approach for office building in city scale using the Simple-normalization, Entropy-based TOPSIS and K-means method. The proposed approach chooses the monthly energy use intensity in single year of office buildings as evaluation indicators, and consists of three steps: (1) The building energy use intensity are calculated by Simple-normalization method. (2) The Entropy-based TOPSIS method is applied to score and evaluate the building energy performance. (3) Finally, the K-means method is employed to rank the evaluated buildings. To validate the proposed approach, 24 office buildings in Tongling city of China are served as case study to present the evaluating and ranking procedure. Simultaneously, the comparison analysis between the proposed approach and two traditional approaches are conducted considering consistency and rationality synchronously. The comparison results demonstrate the proposed approach can more effectively score the energy performance of office buildings, and achieve the reasonable order and grade. In addition, to demonstrate the application, the proposed approach is also applied on Urban-area Building Energy Consumption Monitoring System of Tongling city in China. As a result, this study provides convenient and effective tool to present the energy efficiency gap and energy-saving potential between different buildings in city scale.[13]

Learning class is a collection of several students in an educational institution. Every beginning of the school year the educational institution conducts a grouping class test. However, sometimes class grouping is not in accordance with the ability of students. For this reason, a system is needed to be able to see the ability of students according to the desired parameters. Determination of the weight of test scores is done using the K-Means method as a grouping method. Iteration or repetition process in the K-Means method is very important because the weight value is still very possible to change. Therefore, the repetition process is carried out to produce a value that does not change and is used to determine the ability level of students. The results of the class grouping test scores affect the ability of students. Application of K-Means method is used in building an information system grouping student admissions in an educational institution. Acceptance of students will be grouped into 3 groups of learning classes. The results of testing the system that applies K-Means method and based on data on the admission of prospective students from educational institutions have very high accuracy with an error rate of 0.074. [14]

Cluster analysis aims to classify data objects into two categories: objects that are similar in characteristics in one cluster and objects that are different in characteristics with the other objects of another cluster. K-Means is a method included in the distance-based clustering algorithm that starts by

determining the number of desired clusters. Malnutrition is one of the biggest concerns in Indonesia. According to Riskesdas 2018 data, as many as 17.7% infants under 60-month-old are still having problems with nutrition intake while 3.9% are having malnutrition. This might result in higher death rate. This research was conducted to classify the nutritional status of infants under 60-month-old conducted by the C-Means Clustering method. This research is non-reactive, using secondary data in Ponkesdes Mayangrejo, Bojonegoro without direct interaction with the subject. This study concluded that the grouping of nutritional status is possible by using K-Means with 4 clusters formed which are 23 malnourished toddlers, 17 undernourished toddlers, 7 nourished toddlers, and 10 over-nourished toddlers.[15]

The presidential election is one of the political events that occur in Indonesia once in five years. Public satisfaction and dissatisfaction with political issues have led to an increase in the number of political opinion tweets. The purpose of this study is to examine the performance of the k-means and k-medoids method in the Twitter data and to tweet about the presidential election in 2019. The data used in this study are primary data taken from Muhyi's research, then mining the text against data obtained. Because this data has been processed by Muhyi to analyze the electability of the 2019 presidential candidate pairs, for this journal needs a preprocessing was carried out to analyze the tendency of tweets to side with the candidate pairs of one or two. The difference in the pre-processing of this research with previous research is that there is a cleaning of duplicate data and normalizing. The results of this study indicate that the optimal number of clusters resulting from the k-means method and the k-medoid method are different.[16]

A case study of a passive components company in Taiwan is presented to assess the supplier performance evaluations in accordance with nine criteria. Except for coordination, the rest of eight criteria have the objective assessments. In order to establish a more objective assessment in supplier performance evaluations, a monitoring system is to be set up by considering the eight criteria to determine if a supplier performance is either underestimated or overestimated. K-means method is employed to classify all of the suppliers into three categories. The results based on the data from four quarters in 2017 and the first quarter in 2018 show that 4 of 43, 13 of 57, 24 of 58, 13 of 57, and 15 of 57 indicate the supplier performance seems to be abnormal, i.e., either underestimate or overestimate, for the first quarter, second quarter, third quarter, and fourth quarter of 2017, and the first quarter of 2018, respectively. Therefore, further investigations on coordination criterion can be conducted to understand if the judgment on coordination is reasonable.[17]

There are various methods of objects' clusterization used in different areas of machine learning. Among the vast amount of clusterization methods, the K-means method is one of the most popular. Such a method has as pros as cons. Speaking about the advantages of this method, we can mention the rather high speed of objects clusterization. The main disadvantage is a necessity to know the number of clusters before the experiment. This paper describes the new way and the new method of clusterization, based on the K-means method. The method we suggest is also quite fast in terms of processing speed, however, it does not require the user to know in advance the exact number of clusters to be processed. The user only has to define the range within which the number of clusters is located. Besides, using suggested method there is a possibility to limit the radius of clusters, which would allow finding objects that express the criteria of one cluster in the most distinctive and accurate way, and it would also allow limiting the number of objects in each cluster within the certain range [18]

Driving cycle plays a vital role in the production and evaluating the performance of the vehicle. Driving cycle is a representative speed-time profile of driving behavior of specific region or city. Many countries has developed their own driving cycle such as United State of America, United Kingdom, India, China, Ireland, Slovenia, Singapore, and many more. The objectives of this paper are to characterize and develop driving cycle of Kuala Terengganu city at 8.00 a.m. along five different routes using k-means method, to analyze fuel rate and emissions using the driving cycle developed and to compare the fuel rate and emissions with conventional engine vehicles, parallel plug-in hybrid electric vehicle, series plug-in hybrid electric vehicle and single split-mode plug-in hybrid electric vehicle. The methodology involves three major steps which are route selection, data collection using on-road measurement method and driving cycle development using k-means method. Matrix Laboratory software (MATLAB) has been used as the computer program platform in order to produce the best driving cycle and Vehicle System Simulation Tool Development (AUTONOMIE) software has been used to analyze fuel rate and gas emission. Based on the findings, it can be concluded that, Route C and single spilt-mode PHEV powertrain used and emit least amount of fuel and emissions.[19]

Clustering of the load patterns from distribution network customers is of vital importance for several applications. However, the growing number of advanced metering infrastructures (AMI) and a variety of customer behaviors make the clustering task quite challenging due to the increasing amount of load data. K-means is a widely used clustering algorithm in processing a large dataset with acceptable computational efficiency, but it suffers from local optimal solutions. To address this issue, this paper presents a hierarchical K-means (H-K-means) method for better clustering performance for big data problems. The proposed method is applied to a large-scale AMI dataset and its effectiveness is evaluated by benchmarking with several existing clustering methods in terms of five common adequacy indices, outliers' detection, and computation time.[20]

The surface river water quality in Banjarmasin city tends to decline constantly as the result of direct and indirect waste disposal from various human activities along the river body. This study aimed to determine the vulnerability points against pollution in the rivers of Banjarmasin using clustering techniques with K-means algorithm. The parameters observed include Biological Oxygen Demand (BOD), Chemical Oxygen Demand (COD), Total Suspend Solid (TSS) and Dissolved Oxygen (DO). The data were collected at eight water monitoring stations on various rivers in Banjarmasin city. With the K-means method, four water quality status were clustered. The result showed that 6 stations observed during the period April to October 2016 were catagorized into the heavy polluted cluster with major pollution point of sources came from the domestic and industrial activities.[21]

Among many clustering algorithms, the K-means clustering algorithm is widely used because of its simple algorithm and fast convergence. However, the K-value of clustering needs to be given in advance and the choice of K-value directly affect the convergence result. To solve this problem, we mainly analyze four K-value selection algorithms, namely Elbow Method, Gap Statistic, Silhouette Coefficient, and Canopy; give the pseudo code of the algorithm; and use the standard data set Iris for experimental verification. Finally, the verification results are evaluated, the advantages and disadvantages of the above four algorithms in a K-value selection are given, and the clustering range of the data set is pointed out.[22]

This study focuses on the analysis of sentiments on Indonesian twitter data. Twitter data on Indonesian simultaneous pilkada used to get its sentiments using Naïve Bayes Classifier method as a method of classification and K-means method to get Label on the data train process. Combining the two methods is expected to get high accuracy results. The results obtained from the research shows a pretty good accuracy of 74.5%.[23]

Time-of-day interval partition (TIP) at a signalized intersection is of great importance in traffic control. There are two shortcomings of the traditional clustering algorithms based on traditional distance definitions (such as Euclidean distance) of traffic flows. First, some continuous time intervals are usually divided into small segments. Second, 0 o'clock (24 o'clock) is usually selected as the breakpoint. It follows that the relationship between TIP and traffic signal control is neglected. To this end, a novel cyclic distance of traffic flows is defined, which can make the end of the last cycle (24 o'clock of the last day) and the beginning of the current cycle (0 o'clock of the current day) cluster into one group. Next, a cyclic weighted k-means method is proposed, with centroid initialization, cluster number selection, and breakpoint adjustment. Lastly, the proposed method is applied to a real intersection to evaluate the benefits of traffic signal control. The conclusion of the empirical study confirms the feasibility and effectiveness of the method.[24]

A farmer's welfare classification can be performed to accommodate all significant issues that will assist policymakers, government, and scientists. This study aims to compare K-Nearest Neighbor (K-NN) and K-Means methods for clustering Indonesian farmers' welfare using the fifth wave of Indonesia Family Life Survey (IFLS 5) data. The K-Means method is an unsupervised learning algorithm by classifying the data according to the closest distance between observed and centroids. The K-NN method is a supervised learning algorithm by classifying most of the nearest neighbour data. This study used fifteen factors affecting farmers' welfare including land area, type of water, type of rice, income, expenditure, loan, mobile phone use, harvest frequency, crop failure, land ownership, gender, age, level of education, home ownership, and ownership of health insurance. The K-NN performed well to classify farmers' welfare as the K-Means methods in the district data, with an accuracy of 89.8% compared to 53.7%. The K-NN classification results in provinces data showed that the provinces of Bali, East Java, South Kalimantan, Lampung, West Nusa Tenggara, South Sulawesi, and South Sumatra

were included as prosperous provinces; while the provinces of Banten, DI Yogyakarta, West Java, Central Java, West Sumatra, and North Sumatra were included as non-prosperous provinces.[25]

Fig. 3. Figure 3. Framework thingking created

2. Method

2.1 Data Collection Stage

In the application of k-means clustering to determine priority program for village funds based on the level of need, related data is needed. Sources of research data were obtained from the village work program design data submitted by each head. hamlet and village government in 2020. The criteria used are the value of urgency and the value of the benefits of a work program to be implemented.

2. 2 Data Processing Stage

Data that has been processed must be processed first so that it can be clustered [10]. So at this stage it will produce a value of urgency and value of benefit from each proposed work program which is then processed at thestage clustering.

2.3 Stage Clustering

Clustering is unsupervised classification and is the process of partitioning a set of objects from a set of data into several *clusters.*This can be done by applying equations and steps regarding the distance algorithm, namely *Euclidean Distance* (Venkateswarlu & Raju, nd). The three *clusters,* namely *cluster* level high priority, *clusters* medium priority leveland *clusters of* low priority level. The stages in determining the *clusters* of each proposed work program are described in the *flowchart* following. Clustering is a process of classification into several equal parts according to the predefined categories [11].

2.4 Analysis Phase

At this stage, village program data analysis is carried out. The data obtained is then processed by using the weight calculation for each index. In the previous stage, it has been determined that the results will be *clustered*into 3 *clusters,* then at this stage the results will be analyzed. Work program data that has been collected and has been given a weight from each index, namely the value of urgency and value of benefit as shown in Table 1 below.

Table 1. Table 1. Data on the Value of Criteria for Each Work Program

No	Village Program Proposal	Value of Urgency	Value of Use
1	Repair of hamlet roads	76	47
2	Construction of hamlet bridges 2	80	56
3	Construction of village reservois	67	57
4	Construction of hamlet madrasah diniyah 3	60	31
5	Repair of village sports facilities	57	27
6	Construction of rainwater tanks	45	38
7	Hamlet house surgery 4	70	40
8	Vumdesinputted	69	32
8	Village UMKM 5	65	34

Then, the data into thestage *clustering* by applying the *K-Means algorithm* todata into three *cluster theclusters*. *Kmeans* is amethod of clustering *partitioning* that is able to separate data into different groups. By *partitioning* iteratively. *K-Means* can minimize the average distance of each data to the cluster. In the *K Means algorithm*, each data must belong to *acluster* certainat one stage of the process, at the next stage of the process it can move to *cluster* another [12]. In the K-Means method, there are several steps that must be carried out when performing calculations to find cluster data until the iteration is the same as the previous calculation, here are the K-Means stages:

a. Determining K as the number of *clusters* to be formed, in this study the researchers formed 3 *clusters.* (C1, C2, C3).

b. Determining the *centroid* data, in the application of the Kmeans algorithm, a midpoint orvalue is generated *centroid* from the data obtained provided that the desired number of clusters is 3. Determination of *clusters is* divided into three parts, namely the high-level priority cluster (C1), *cluster* the medium level priorityC2) and *cluster* the low-level priority(C3), then the midpoint or centroid also has 3 points. Determination of the cluster point is done by taking the largest (maximum) value for the high-level priority cluster (C1), the average (*averagevalue*)for the mediumlevel priority cluster (C2) and the smallest (minimum) value for the low-level priority cluster (C3). . The point value can be seen in Table 2 below:

Table 2. Centroid Initial Data

Attribute	Urgency Value	Benefit Value (b)
C1	80	57
C2	65.44	40.22
C3	45	27

Calculation of K-Means, after the data is available then it is entered into the formula for the data clustering stage, and the following general formula for calculating K-Means is shown in Eq. (1). So that to determine the distance between the data and the*cluster* using Eq. (2) below:

$$d(x|\ \ |j,y_j) = \sqrt{\sum_{i=1}^{n}(x_j - y_j)^2}$$

So that to determine the distance between the data and the cluster using Eq. (2) below:

$$d(x_1,c_1) = \sqrt{(a_1 - c_{1a})^2 + (b_1 - c_{1b})^2}$$
$$d(x_2,c_1) = \sqrt{(a_2 - c_{1a})^2 + (b_2 - c_{1b})^2}$$
$$d(x_3,c_1) = \sqrt{(a_3 - c_{1a})^2 + (b_3 - c_{1b})^2}$$
$$d(x_4,c_1) = \sqrt{(a_4 - c_{1a})^2 + (b_4 - c_{1b})^2}$$

Below is an example of the calculation of the existing data after being entered into the K-Means formula.

C1 = √ (76 - 80)2 + √ (47 - 57)2

C1 = √16 + √100

C1 = √116 C1 = 10.77

From the above calculation, the distance between the first data and the center of *thecluster* firstis 10.77. Next determine C2 as in the following example:

C2 = √ (76 - 65.44)2 + √ (47 –

40.22)2 C2 = √111.51 + √45.97

C2 = √157.48

C2 = 12.54

From the above calculations, the distance between the first data and thecenter *cluster* secondis

12.55. Next determine C3 as in the following example:

C3 = √ (76 - 45)2 + √ (47 - 27)2

C3 = √961 + √400

C3 = √1,361

C3 = 36.89

From the above calculation results the distance of the first data to the center of the *cluster* third is 36.89. Furthermore, to determine the closest distance by choosing one of C1, C2 and C3 which has the closest distance, namely C1 with a data distance of 10.77. Following are the results of the distance after completing the overall calculation for iteration 1 that it shows on table 3.

Table 3. Table 3. Data Distance in Iteration 1

Program Submission	C1	C2	C3	Shortest Distance
Repair of hamlet 1 roads	10.77	12.54	3 Results of 36.89	10.77 Village
Bridge construction 2	1	2147	45.45	1
Development of village reservoirs	13	16.85	37.30	13
Village program	C1	C2	C3	Shortest Distance
Development of madrasah diniyah hamlet	32.80	10.71	15.52	10.71
Improvement of village sports facilities	37.80	15.69	12	12
Construction of rain water	39.82	2056	11	11
Hamler house 4	19.72	4.56	28.18	4.56
BUMDES	27.31	8.96	24.56	8.96
MSMEs hamlet	27.46	6.24	21.19	6.24

After calculating in iteration 1, the work program data are grouped into 3 *clusters* as shown in Table 4 below.

Table 4. Data grouping in Iteration 1

No	Village Program Proposal	Distance to Centroid		
		C1	C2	C3
1	Improvement of hamlet road 1	1	-	-
2	Contruction of hamlet bridges 2	1	-	-
3	Construction of village reservoirs	1	-	-
4	Construction of madrasah diniyah hamlet 3	-	1	-
5	Improvement of village sport facilities	-	-	1
6	Construction of rain water tanks	-	-	1
7	Hamlet house 4	-	1	-
8	Renovation BUMDES	-	1	-
9	Hamlet UMKM 5	-	1	-

Then the calculation is continued in iteration 2 until the object data does not change *clusters*. In this study, the calculated data stopped at the second iteration, because the results of grouping the data in iteration 1 and iteration 2 did not change.

2.5 Implementation

Stage This stage is the development stage of a-based decision support system *web* using theeditor code *Sublime Text*. 3. The programming language used is PHP version 5 with MySQL database version 3.2.1. In this system, there are 2 actors involved, namely the hamlet head and the village government. To illustrate the interaction and access rights of each actor.

4. Results and Discussion

4.1 Results

Results of this study are grouping the data into three predefined clusters. This study also proves that the K-Means Clustering method is effective in solving the problems in this study. In addition, this study also resulted in the implementation of a web-based decision support system for selecting priority programs for village funds. This system was built using the PHP programming language and MySQL as the database. The output of this system is the grouping of village priority programs into three *clusters*.

a. Classification results of work programs

From the nine proposed work programs that are calculated using the K-Means *Clustering algorithm*, it is known that there are three proposed programs that are included in *cluster* the high-level priority(C1), namely repairing hamlet 1 roads, building hamlet 2 bridges and building village reservoirs. . Then there are four village programs that are included in *cluster* the medium level priority(C2), namely the construction of madrasah diniyah in hamlet 3,house renovation in hamlet 4, BUMDES and management of MSMEs in hamlet 5.While those included in *cluster* the low-level priority(C3) are two programs. work, namely renovation of village sports facilities and construction of rainwater storage tanks. Fig 4 shows the percentage of work program classification results.

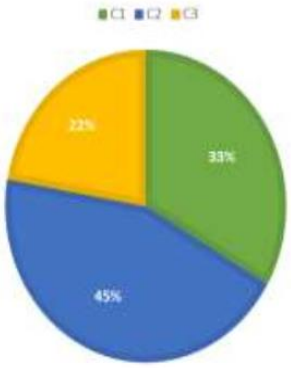

Fig. 4. Percentage of Work Program Classification Results

4.2. Discussion

Data used in this study were data on work programs proposed by the hamlet heads and the village government. The data was obtained through observations and interviews conducted within the village government. The method used is the K-Means Clustering method which is used to classify the proposed work program into three priority level clusters. Through the K-Meanscalculation *Clustering* on the proposed work program data, 3 work programs includingobtained *clusters* high-level priority(C1) were, 4 work programs including *clusters* medium-level priority(C2) and 2 work programs including *clusters* low-level priority(C3). The implementation of this research is a-based decision support system *web* built using thecode editor *Sublime Text 3* with the programming language PHP version 5 and MySQL version

V. CONCLUSION

This research resulted in a decision that the proposed road repair work program in hamlet 1, bridge construction in hamlet 2 and construction of village reservoirs were included in high priority programs. Then the proposed work program for the construction of madrasah diniyah in hamlet 3, house renovation in hamlet 4, BUMDES and management of MSMEs in hamlet 5 are included in medium priority programs. Meanwhile, the work program to improve village sports facilities and the construction of rainwater storage tanks is included in the low priority program. In addition, this study also proves that the K-Means Clustering method can be used to classify proposed work programs and implement them into a-based decision support system web.

REFERENCES

[1] Ahmad Dzauqy Abdur Rabb, "Village Funds Based on the Minister of Finance Regulation Number 93 of 2015 In The Ganra Sub-District, Soppeng Regency, 2016.

[2] S. Gatra, "Total village funds 2019-2024 Rp 400 Trillion", Wired, February 26, 2019, [Online]. Available: https://nasional.kompas.com/read/2019/02/26/17333511/ total-dana-desa-2019-2024-rp-400 triliun?page=all [Accesed: 25 May 2021].

[3] ....., Regulation of the Minister of Home Affairs of the Republic of Indonesia number 113 of 2014- Ministry of Trade", Wired, December 31, 2014, [Online]. Available:http://binapemdes.kemendagri.go.id/produkhu kum/detail/peraturanmenteridalamnegerirepublikindonesi a number 113 tahun2014 [Accesed: May 15, 2021].

[4] F. A. Hermawati, " Data Mining. Publisher : ANDI, 2013. Available at: Book

[5] Noviana, T., Jasmir, J., &Novianto, Y. (2019). The Implementation of Data Mining Determines the Priority Group of Rastra Rice Beneficiaries With K-Means Clustering. Informatics Engineering Study Program, Stikom Dinamika Bangsa, 159–174, 2019. Available at: Google Scholar

[6] Venkateswarlu, B., &Raju, P. G. S. V. P. (n.d.). Mine Blood Donors Information through Improved K- Means Clustering. ArXiv Preprint ArXiv:1309.2597. Available at: Goolge Scholar

[7] Windarto, A. P. (2017). Application of Datamining On Export of Fruits by Destination Country Using K-Means Clustering Method. Techno.Com, 16(4), 348–357, 2017. doi: 10.33633/tc.v16i4.1447

[8] S. Defit, S.A. Azis, Y.Yunus, "Clustering of Assistance Funds in The Family of Hope Program (PKH) Using KMeans Method", Journal of Business Economic Informatics, Vol. 3, No. 2, 2020, Pp.: 53-59. Available at: Google Scholar

[9] T. Noviana, Jasmir, Y. Novianto, "Application of Data Mining Determines The Priority Group of Rastra Rice Beneficiaries with K-Means Clustering", Student Scientific Journal, Informatics Engineering, Vol. 1, No,3, 2019. Available at: Google Scholar

[10] Windarto, "Data Mining Application in Fruits Export by Destination Country Using the K-Means Clustering Method. Techno.Com, 2017. Available at: Google Scholar

[11] S. Lesmana, A. F. Akbari, E. Y. Rahman, D. Gustian, "Implementation of K-Means in the Effectiveness of ELearning Learning during the Covid-19 Pandemic, " National Seminar on Informatics 2020 (SEMNASIF 2020), Yogyakarta Veterans National Development University, 2020. Available at: Google Scholar

[12] F. Sembiring, Octaviana, S. Saepudin, "Implementation of K-Means Method in Illegal Collection Area In Sukabumi Regency (Case Study: Population and Civil Registration Office)", Jurnal Tekno Insentif | ISSN (p):1907-4964 |, Vol. 14, No. 1, April 2020. doi: 10.36787/jti.v14i1.165

[13] Sun FYu J, Improved energy performance evaluating and ranking approach for office buildings using Simple-normalization, Entropy-based TOPSIS and K-means method, Energy Reports, (2021), 7 doi: 10.1016/j.egyr.2021.03.007

[14] D. Indriyanti AR. Prehanto DZ. Vitadiar T, K-means method for clustering learning classes, Indonesian Journal of Electrical Engineering and Computer Science, (2021), 22(2) doi: 10.11591/ijeecs.v22.i2.pp835-841

[15] Nagari SInayati L, IMPLEMENTATION OF CLUSTERING USING K-MEANS METHOD TO DETERMINE NUTRITIONAL STATUS, Jurnal Biometrika dan Kependudukan, (2020), 9(1) doi: 10.20473/jbk.v9i1.2020.62-68

[16] Oktarina CNotodiputro KIndahwati I, COMPARISON OF K-MEANS CLUSTERING METHOD AND K-MEDOIDS ON TWITTER DATA, Indonesian Journal of Statistics and Its Applications, (2020), 4(1) doi: 10.29244/ijsa.v4i1.599

[17] Tseng CWu H, A monitoring system to assess supplier performance of a passive components company in Taiwan by K-Means method, IAENG International Journal of Computer Science, (2020), 47(2) Available at: Google Scholar

[18] Litvinenko NMamyrbayev OShayakhmetova A et al.See Clusterization by the K-means method when K is unknown, ITM Web of Conferences, (2019), 24 doi: 10.1051/itmconf/20192401013

[19] Anida ISalisa A, Driving cycle development for Kuala Terengganu city using k-means method, International Journal of Electrical and Computer Engineering, (2019), 9(3) doi: 10.11591/ijece.v9i3.pp1780-1787

[20] Tian Shi XuHsiao Dong Chiang[...]Chin Woo Tan ,Hierarchical K-means Method for Clustering Large-Scale Advanced Metering Infrastructure Data, IEEE Transactions on Power Delivery (2017) Available at: Google Scholar

[21] Zubaidah TKarnaningroem NSlamet A, K-means method for clustering water quality status on the rivers of Banjarmasin, Indonesia, ARPN Journal of Engineering and Applied Sciences, (2018), 13(11) doi: 10.31227/osf.io/s9n2u

[22]Yuan CYang H, Research on K-Value Selection Method of K-Means Clustering Algorithm,J, (2019), 2(2) doi: 10.3390/j2020016

[23] Susanto AAtho'il Maula MUtomo I et al , Sentiment Analysis on Indonesia Twitter Data Using Naïve Bayes and K-Means Method, Journal of Applied Intelligent System, (2021), 6(1) Available at: Google Scholar

[24] Wang GQin WWang Y, Cyclic weighted k-means method with application to time-of-day interval partition, Sustainability (Switzerland), (2021), 13(9) doi: 10.3390/su13094796

[25] Pawa LKismiantini, Comparing k-nearest neighbor and k-means methods for clustering Indonesian farmers' welfare, Journal of Physics: Conference Series, (2020) doi: 10.1088/1742-6596/1581/1/012020