



Vol.17 No.1 | 2023

# DATA MINING ANALYSIS TO DETERMINE THE PROSPER FAMILY AREA KECAMATAN SIMPANG EMPAT WITH ALGORITHM METHOD K-MEANS CLUSTERING

Ricky Rizaldi<sup>\*1)</sup>, Fauriatun Helmiah<sup>2)</sup>, Sudarmin<sup>3)</sup>

<sup>1,2,3</sup>Sekolah Tinggi Manajemen Informatika dan Komputer Royal

\*Corresponding Email: [rickyrizaldi49@gmail.com](mailto:rickyrizaldi49@gmail.com)

**Submit :**  
30/09/2022

**Accept :**  
19/01/2023

**Publish :**  
28/02/2023



## Abstract

**Background :** The long-term development phase II of Pelita VI is that development is aimed at improving the quality of Indonesia's fully developed and independent human resources. Community development is very dependent on family life which is the core part of the community, so that the family has a strategic value in national development and becomes the foundation of whole human development. The problem we face today is that there are still many families in Indonesia who are in a pre-prosperous condition, it is our obligation to improve their welfare, so as to achieve prosperous families in Simpang Empat District. **Method :** using data mining technique K-Means Clustering algorithm method. K-Means is a calculation used in discrete sets that isolate information into various sets. Clustering is a method for finding and grouping data that have similar characteristics between one data and another. **Result :** Implementation of the interface is carried out with each program that is built. The following is the implementation of the system interface to determine the area of a prosperous family in Kecamatan Simpang Empat the data is taken based on the hamlet in Kecamatan Simpang Empat. **Conclusion :** In this study, the users of the application of the Clustering Method with the K-Means Algorithm to determine the Prosperous Family Area of the Simpang Empat District Office. The purpose of using data mining is to make it easier for the Secretary of the Simpang Empat sub-district office to determine the Prosperous Family Areas which are KS I, KS II and KS III so that data mining can be used properly by users, so education and training are needed in using the system.

**Keywords:** Determining Prosperous Family Area, K-Means Method, Web-Based

© 2022 Lembaga Layanan Pendidikan Tinggi Wilayah X. This is an open access article under the CC Attribution 4.0 license (<https://creativecommons.org/licenses/by/4.0/>).



## INTRODUCTION

Community development is very dependent on family life which is the core part of the community, so that the family has a strategic value in national development and becomes the foundation of whole human development. The problem we face today is that there are still many families in Indonesia who are in a pre-prosperous condition, it is our obligation to improve their welfare, so as to achieve prosperous families in Simpang Empat District. It is recorded from the data that the Simpang Empat Village Hall Office has a population of 52175 people spread over twelve (12) Kelurahan with such a large population, so that there is so much data that data analysis techniques are needed to solve the problem. The number of Heads of Families who received capital assistance, the number of praKS (Pra (Keluarg Sejahtera), and the number of KS1 using data mining techniques using the K-Means Clustering algorithm method. K-Means is a calculation used in discrete collections that isolates information into various groups (Sadewo et al., 2017). The basic idea of the K-Means algorithm is very simple, namely minimizing the Sum of Squared Error (SSE) between information objects with various centroids (M. G. Sadewo, et al, 2017). Clustering is a method for finding and grouping data that has the similarity of characteristics (similarity) between one data and another (Sulastri et al., 2017). Data partitioning is done by finding the closest distance value between the data and the centroid value that has been set either randomly or with the Initial Set of Centroids, to determine the centroid value based on sequential objects (H. Sulastri and A. I. Gufroni et al., 2017 ). The system is a network of

procedures created according to the integrated management to carry out the company's main activities (R. S. Prahara, et al, 2017). The usefulness of information is to reduce uncertainty in the decision-making process about a situation (R. Asmara, et al., 2017). An information system is an integrated system that works together to collect, process, store and disseminate information to assist an organization in controlling the system to achieve organizational goals (S. S. Romadhon et al., 2017).

## RESEARCH METHODS

Data mining is a method used to explore knowledge from data sets. Collections or piles of data that are only stored are processed by data mining to produce knowledge that can be utilized for clinical prediction needs (D. Jollyta, et al., 2017). K-Means is a non-hierarchical data grouping method that tries to partition existing data into two or more groups. This method partitions data into groups so that data with the same characteristics are included in the same group and data with different characteristics are grouped into other groups. The K-Means algorithm is one of the algorithms in the clustering function or grouping. Clustering refers to grouping data, observations or cases based on the similarity of the object under study (R. Risnawati and R. Rohminati et al., 2020). Clustering is part of the science of data mining which is unsupervised. Clustering is the process of dividing data into classes or clusters based on their level of similarity. In clustering, data that have similarities are included in the same cluster, while those that do not have similarities are included in different





clusters. Clustering will group the data into a number of cluster groups based on the similarity of the characteristics of each data in the existing groups (R. Risnawati and R. Rohminatin et al., 2020).

In general, clustering data using the K-Means method is carried out with the following basic algorithm:

- a. Determine the number of clusters to be formed
- b. Allocate data into clusters randomly, for the selection of random numbers the initial center of the cluster must be in the order of the data.
- c. Calculate the group center of each group data. From each centroid will be taken the average value. If the mean is expressed as a group, then  $i$  is the feature,  $p$  is the dimension of the data whose equation is to calculate the centroid of feature  $i$  using equation.

$$= \frac{1}{M} \sum_{j=1}^M X_j = 1 \dots \dots \dots (1)$$

Equation 1 is carried out as many as  $p$  dimensions  $i=1$  to  $i=p$

- d. Allocate each data to the nearest centroid by measuring the distance of space that can be found the equation.

$$= \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \dots \dots \dots (2)$$

Allocate data to each K-Means method based on the comparison of the distance between the centroids to each group.

$$d = \min\{D(x_1, C_1)\}$$

$$1 \dots \dots \dots (3)$$

lainnya

$a_{i1}$  is the value of the member point  $x_i$  to the center of the group  $c_1$ ,  $d$  is the shortest

distance from the data  $x_i$  to the group after comparison, and  $c_1$  is the center of the group 1.

$$= \sum_{k=1}^k \sum_{i=1}^n a_{ki}^2 \dots \dots \dots (4)$$

$n$  is the number of data,  $k$  is the number of groups,  $a_{i1}$  is the value of the member data point  $x_i$  to the  $c_i$  group that is followed.  $a$  has a value of 0 or 1, the data is a member of the group.

- e. Perform the calculation again if there are still changes to the data until it is complete

## RESULTS AND DISCUSSION

Nm	Data Flow	Data Flow	Source
1	Dusun Data at the Simpang Empat sub-district office	Dusun Data at the Simpang Empat sub-district office	Simpang Empat sub-district office

Tabel 1. Output Document

The researcher conducted an analysis on the data of the Prosperous Family Area by taking samples from the Simpang Empat sub-district office, the following researchers made an illustration of how the researchers processed the existing data which would later become a recapitulation per hamlet for further inclusion in the calculation formula for the K-Means Clustering Analysis algorithm.

Nm	Criteria	Data source	Condition (Value)
1	Income	Simpang Empat sub-district office	If income is 3,000,000, then the value is 1, If income is 2.9 million – 2.5 million, then the value is 2 and if income is < 2,400,000 then the value is 3.





2	Dependent	Simpang Empat sub-district office	If the dependent is 1 person, then the value is 1, if the dependent is 2 - 3 people, then the value is 2, if the dependent is more than 3 people, then the value is 3.
3	Help Recipient	Simpang Empat sub-district office	If you have not received assistance, then the score is 3, if you receive 1 time assistance, then the value is 2, if you receive 2 times more assistance, then the value is 1.

Table 2. Process Analysis

Nm.	Name Alternativ	K1	K2	K3	Cluster
1	Dusun I	3	2	3	
2	Dusun II	3	2	3	
3	Dusun III	3	2	2	
4	Dusun IV	2	3	3	
5	Dusun V	3	3	3	Cluster 1
6	Dusun VI	2	3	3	
7	Dusun VII	3	3	2	
8	Dusun VIII	3	3	3	
9	Dusun IX	3	2	2	
10	Dusun Hessa I	3	2	3	
11	Dusun Hessa II	3	2	2	
12	Dusun Hessa III	2	3	2	
13	Dusun Hessa IV	3	3	2	
14	Dusun Hessa V	2	2	2	Cluster 2
15	Dusun Hessa VI	2	3	2	
16	Dusun Hessa VII	3	2	3	
17	Dusun Hessa VIII	2	3	3	
18	Dusun Hessa IX	3	3	3	
19	Dusun Hessa X	2	3	3	
20	Dusun I	2	2	3	
21	Dusun II	2	2	3	
22	Dusun III	2	3	3	
23	Dusun IV	3	3	2	
24	Dusun V	3	3	3	
25	Dusun VI	2	3	3	
26	Dusun VII	2	3	3	
27	Dusun VIII	3	3	2	
28	Dusun I	3	3	2	
29	Dusun II	3	3	2	
30	Dusun III	3	3	2	
31	Dusun IV	2	2	2	
32	Dusun V	3	2	2	
33	Dusun VI	2	2	3	
34	Dusun I	3	2	3	
35	Dusun II	2	1	2	Cluster

				3
36	Dusun III	2	2	3
37	Dusun IV	3	2	2
38	Dusun V	2	2	2
39	Dusun I	3	3	3
40	Dusun II	2	2	3
41	Dusun III	2	3	2
42	Dusun IV	2	2	3
43	Dusun V	2	3	2
44	Dusun VI	3	2	3
45	Dusun I	2	2	2
46	Dusun II	3	2	2
47	Dusun III	2	3	2
48	Dusun IV	3	3	2
49	Dusun V	3	2	2
50	Dusun VI	3	2	2

Table 3. Alternative Data Tables and Criteria and Prosperous Family Area

- For the initial determination it is assumed:
1. The 1st data is taken as the center of the 1st Cluster, namely Mr. Sadinomo as the cluster that has the highest value: (3, 3, 3)
2. The 2nd data is taken as the center of the 2nd Cluster, namely Ibu Semi as the cluster that has the lowest value: (2, 2, 2).

#### Cluster Center Distance Calculation

$$D_c = \sqrt{(x_i - s_i)^2 + (y_i - t_i)^2}$$

Since C1 is looking for a new cluster calculation to be:

$$A1 = \sqrt{(3 - 3)^2 + (2 - 3)^2 + (3 - 3)^2} = 1,00$$

$$A2 = \sqrt{(3 - 3)^2 + (2 - 3)^2 + (3 - 3)^2} = 1,00$$





$$\begin{aligned}
 A3 &= \sqrt{(3-3)^2 + (2-3)^2 + (3-3)^2} = 1,414 \\
 A4 &= \sqrt{(3-3)^2 + (3-3)^2 + (3-3)^2} = 1,00 \\
 A5 &= \sqrt{(3-3)^2 + (2-3)^2 + (3-3)^2} = 1,00 \text{ and so on until } A50
 \end{aligned}$$

### Determination of New Cluster Center

Because C1 has 26 members, the new cluster calculation becomes: C1

$$\begin{aligned}
 &= \left( \frac{3+3+3+2+3+2+3+3+3+3+3+2+3+2+3+2+3+2+2+2+3+3+3+3+3}{26} \right) \\
 &= 2,7308 \\
 &= \left( \frac{2+2+2+3+3+3+3+3+2+2+2+3+3+3+2+3+3+3+3+3+3+2}{26} \right) \\
 &= 2,7692 \\
 &= \left( \frac{3+3+2+3+3+3+2+3+2+3+2+2+2+2+3+3+3+3+3+3+3+3+3+3+2+2+2}{26} \right) \\
 &= 2,6923
 \end{aligned}$$

N m.	Name Alternativ	Member	Cl 1	Cl 2	Cl 3
1	Dusun I	C1	Blue		
2	Dusun II	C1	Blue		
3	Dusun III	C2		Red	
4	Dusun IV	C1	Blue		
5	Dusun V	C1	Blue		
6	Dusun VI	C1	Blue		
7	Dusun VII	C1	Blue		
8	Dusun VIII	C1	Blue		
9	Dusun IX	C2		Red	
10	Dusun Hessa I	C1	Blue		
11	Dusun Hessa II	C2		Red	
12	Dusun Hessa III	C2		Red	
13	Dusun Hessa IV	C1	Blue		
14	Dusun Hessa V	C2		Red	
15	Dusun Hessa VI	C2		Red	
16	Dusun Hessa VII	C1	Blue		
17	Dusun Hessa VIII	C1	Blue		

VIII				
18	Dusun Hessa IX	C1	Blue	
19	Dusun Hessa X	C1	Blue	
20	Dusun I	C2		Red
21	Dusun II	C2		Red
22	Dusun III	C1	Blue	
23	Dusun IV	C1	Blue	
24	Dusun V	C1	Blue	
25	Dusun VI	C1	Blue	
26	Dusun VII	C1	Blue	
27	Dusun VIII	C1	Blue	
28	Dusun I	C1	Blue	
29	Dusun II	C1	Blue	
30	Dusun III	C1	Blue	
31	Dusun IV	C2		Red
32	Dusun V	C2		Red
33	Dusun VI	C2		Red
34	Dusun I	C1	Blue	
35	Dusun II	C3		Green
36	Dusun III	C2		Red
37	Dusun IV	C2		Red
38	Dusun V	C2		Red
39	Dusun I	C1	Blue	
40	Dusun II	C2		Red
41	Dusun III	C2		Red
42	Dusun IV	C2		Red
43	Dusun V	C2		Red
44	Dusun VI	C1	Blue	
45	Dusun I	C2		Red
46	Dusun II	C2		Red
47	Dusun III	C2		Red
48	Dusun IV	C1	Blue	
49	Dusun V	C2		Red
20	Dusun VI	C2		Red

Table 4. Cluster Member Grouping

### CONCLUSION

In this study, the users of the application of the Clustering Method with the K-Means Algorithm to determine the Prosperous Family Area of the Simpang Empat District Office. The purpose of using data mining is to make it easier for the Secretary of the Simpang Empat sub-district office to determine the Prosperous Family Areas which are KS I, KS II and KS III so that data mining can be used properly by users, so education and training are needed in using the system.

### BIBLIOGRAPHY

[1] M. G. Sadewo, A. P. Windarto, and D. Hartama, "Penerapan Datamining





- Pada Populasi Daging Ayam Ras Pedaging Di Indonesia Berdasarkan Provinsi Menggunakan K-Means Clustering,” *InfoTekJar (Jurnal Nas. Inform. dan Teknol. Jaringan)*, vol. 2, no. 1, pp. 60–67, 2017, doi: 10.30743/infotekjar.v2i1.164.J. U. Duncombe, “Infrared navigation—Part I: An assessment of feasibility,” *IEEE Trans. Electron Devices*, vol. ED-11, pp. 34–39, Jan. 1959.
- [2] H. Sulastri and A. I. Gufroni, “Penerapan Data Mining Dalam Pengelompokan Penderita Thalassaemia,” *J. Nas. Teknol. dan Sist. Inf.*, vol. 3, no. 2, pp. 299–305, 2017, doi: 10.25077/teknosi.v3i2.2017.299-305.
- [3] R. S. Prahara, H. Wahyono, and S. H. Utomo, “Menentukan Kualitas Pembelajaran Ekonomi Berwawasan Lingkungan Berdasarkan Input-Proses-Output Pembelajaran,” 2017.
- [4] S. Kasus and D. Hwi, “K-MEANS CLUSTERING HWI PRODUCTS,” 2019.
- [5] R. Asmara, S. Kom, and M. Kom, “Sistem Informasi Pengolahan Data Penanggulangan Bencana Pada Kantor Badan Penanggulangan Bencana Daerah (BPBD) Kabupaten Padang Pariaman,” *J. J-Click*, vol. 3, no. 2, 2017.
- [6] S. S. Romadhon, P. Studi, and T. Informatika, “Vol . 3 No . 1 Februari 2019 ISSN : 2597-3673 ( Online ) ISSN : 2579-5201 ( Printed ) ISSN : 2597-3673 ( Online ) ISSN : 2579-5201 ( Printed ),” vol. 3, no. 1, pp. 21–28, 2019.
- [7] D. Jollyta, W. Ramdhan, and M. Zarlis, *Konsep Data Mining Dan Penerapan*. Deepublish, 2020.
- [8] R. A. Indraputra and R. Fitriana, “K-Means Clustering Data COVID-19,” vol. 10, no. 3, pp. 275–282, 2020.
- [9] I. Tanjung, “Perancangan Sistem Informasi Rekam Medis Terpadu Dalam Upaya,” *J. Intra-Tech*, vol. 1, no. 1, pp. 43–54, 2017.
- [10] R. Risnawati and R. Rohminatin, “K-MEANS CLUSTERING HWI PRODUCTS (Case Study: HWI Kisaran Distributor),” in *International Conference on Social, Sciences and Information Technology*, 2020, vol. 1, no. 1, pp. 27–36.

