International Journal of Social Science (IJSS)
Vol.2 Issue.2 August 2022, pp: 1513-1522
ISSN: 2798-3463 (Printed) | 2798-4079 (Online)
DOI: https://doi.org/10.53625/ijss.v2i2.3085

GROUPING OF SEWING TOOL ASSISTANCE RECIPIENTS USING K-MEANS CLUSTERING ANALYSIS

by
Nana Suarna[1], Herman[2], Nining Rahaningsih[3], Rini Astuti[4], Yudhistira Arie Wijaya[5]
[1,3,5]STMIK IKMI Cirebon, [2]Universitas Islam Nusantara Bandung, [4]STMIK LIKMI Bandung
Email: [1]st_nana@yahoo.com, [2]Hermanarianto1972@gmail.com, [3]niningr157@yahoo.co.id, [4]riniastuti@likmi.ac.id, [5]yudhistira010471@gmail.com

Article Info

Article history:
Received Juni 04, 2022
Revised Juni 23, 2022
Accepted Juli 21, 2022

Keywords:
sewing equipment assistance
classification
distribution
k-means method

ABSTRACT

Aid programs for underprivileged communities need continuous data collection in overcoming welfare problems, what has happened so far is by providing direct assistance to very poor families in every village in Indonesia. Related to this, the problem that has occurred so far is that the direct assistance program is not right on target, because many deserving families should not receive the assistance. The data obtained from the village government is not accurate, it is found that data is considered invalid. This study aimed to determine the distribution of sewing equipment recipients in the best cluster, for sewing equipment recipients in Greged Village, Cirebon Regency. As one way to improve data accuracy, a computational method or model is needed in the form of a data mining algorithm using the k-means clustering method to generate priority groups among hundreds of citizens or the poor. The stages start from data collection, training data, and testing data that consider several criteria from household information, economic conditions, housing conditions, and the number of household members in Greged Village, Cirebon Regency. The results of the tests carried out using 155 data with the best level of accuracy were in the K3 cluster with the Davies Bouldin Index's value o: - 0.584. With the K-Means method, it is very appropriate to determine the recipient of the sewing equipment program in Greged Village.

Corresponding Author:
Nana Suarna,
STMIK IKMI Cirebon
Email: st_nana@yahoo.com

1. INTRODUCTION

Determination of recipients of sewing equipment assistance for the community still uses conventional methods to collect data on residents or potential beneficiaries of sewing equipment. The collected data is then carried out in the selection stage. The selection carried out by the Village Government is also still using the manual method and is still not integrated, namely by comparing and separating data from residents or prospective recipients who meet the criteria and those who do not meet the criteria. (Hendarman Lubisa, Ratna Salkiawatib, Sudirman Hala, 2022).

The problem that has occurred so far is that the target in the distribution of the sewing tool assistance program is not accurate, many eligible families should not receive the assistance due to inaccurate data obtained from village government officials found data that is considered invalid. As one way to improve data accuracy, a computational model or method is needed in the form of a data mining algorithm using the K-Means Clustering method. Poverty is a multidimensional problem characterized by the population's low average quality of life, education, health, children's nutrition, and drinking water sources. (Mayssara A. Abo Hassanin Supervised, 2014).

The purpose of this study was to determine the distribution of sewing equipment assistance program groups for the community through social assistance. In predicting the acceptance of sewing equipment assistance using predetermined criteria, to determine whether recipients of assistance were eligible or not eligible to receive community assistance. K-Means Method Clustering has classification capabilities such as decision tree and neural network methods. This method can be used to predict the probability of membership in a class (Sugianto & Maulana, 2019).

International Journal of Social Science (IJSS)
Vol.2 Issue.2 August 2022, pp: 1513-1522
ISSN: 2798-3463 (Printed) | 2798-4079 (Online)

The research conducted by Dede Witan, et al, entitled Classification of Determination of Beneficiaries of the Expected Family Program (PKH) Using the C5.0 Algorithm concluded that the C5.0 Algorithm can be used as a classification method in supporting the decision of beneficiaries of the Expected Family Program by taking into account the gain value. The highest (reinforcement) of the four attributes is ownership of school children, type of house walls, ownership of toddlers, and types of work usually done. The decision tree generated by the classification technique Algorithm C5.0 from the variable that has the highest gain, namely having school children can be used as a reference in determining the beneficiaries of the family program of hope in the future.

From the process of applying data mining using the K-Means method in determining the recipient of sewing equipment, using RapidMiner Software, because it has a high level of accuracy and provides predictions of the level of community welfare based on data updating an integrated database. This study resulted in the best level of accuracy in the K-3 cluster with the Davies Bouldin Index of -.584%.

2. LITERATURE REVIEW

2.1. Data Mining

Data mining is a term used to describe the process of extracting values from the database and is one solution to explain the process of extracting information in a large-scale database and the process of case automation classification based on data patterns obtained from datasets. Data mining is also defined as a set of techniques that are used automatically to thoroughly explore and bring to the surface complex relationships in very large data sets. (Apolinaria Ifon Purnama1), Abdul Aziz2), Anggri Sartika Wiguna3), 2019).

2.2. K-Means method

The k-means method is a method that is included in the distance-based clustering algorithm that divides data into several clusters and this algorithm only works on Numeric attributes. (Windha Mega Pradnya Dhuhita, 2015).

2.3. Clustering

Cluster analysis is finding a collection of objects so that objects in one group are the same or have a relationship with others and are different or not related to objects in other groups. (Moh. Fadel Asikin1, Dian Eka Ratnawati2, Mochammad Ali Fauzi3, 2018).

3. RESEARCH METHOD

K-Means is the most famous clustering method and is widely used in various fields because it is simple, easy to implement, can cluster large data, and has linear time complexity where n is the number of documents, K is the number of clusters, and T is the number of clusters. iteration. K-means is a partitioning method of clustering that separates data into different groups. By iterative partitioning, K-Means can minimize the average distance of each data to its cluster. (1M. Nanda Variestha Waworuntu, 2Muhammad Faisal Amin, 2018)

The basic K-means algorithm is as follows:

$$v = \sqrt{\frac{\sum_{i=1}^{n} Xi}{n}}$$

Where:
i: 1, 2, 3,... n
v: centroid on cluster
xi: object i
n: the number of objects/number of objects that are members of the cluster
Calculate the distance of each data to each centroid using the formula.

$$d(x, y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$$

-Group each data based on the shortest distance between the data and its centroid (C).
-Return to step 3 if the position of the new centroid with the old centroid is not the same.
Information:
dij: object distance between x and y objects

Vol.2 Issue.2 August 2022, pp: 1513-1522
ISSN: 2798-3463 (Printed) | 2798-4079 (Online)
DOI: https://doi.org/10.53625/ijss.v2i2.3085

n: Number of Attributes
x_i: Data Object
y_i: Data Cluster

4. RESULTS AND ANALYSIS

Data Collection Results

The purpose of the grouping carried out in this study is how to group the data on sewing equipment recipients for their dataset sourced from the Central Statistics Agency (BPS) with analysis using the K-Means method.

This study uses raw data from the Greged Village Government in 2017 which has 7 attributes and 155 records. The researcher's sample dataset presents 20 records from 155 records as shown in Table 4.1

Table 4.1. Raw data samples

No	Name	Dusun	Rt	Rw	Desa	Districts
1	Elah	Dusun Pahing	1	3	Greged	Greged
2	Erlianawati	Dusun Pahing	1	3	Greged	Greged
3	Iin	Dusun Pahing	2	3	Greged	Greged
4	Asinah	Dusun Pahing	2	3	Greged	Greged
5	Ucih	Dusun Pahing	3	3	Greged	Greged
6	Robiah	Dusun Pahing	4	3	Greged	Greged
7	Kuniah	Dusun Pahing	4	3	Greged	Greged
8	Yati Maryati	Dusun Pahing	2	3	Greged	Greged
9	Wasniah	Dusun Pahing	1	3	Greged	Greged
10	Ela Jamila	Dusun Pahing	1	3	Greged	Greged
11	Rina Rinawati	Dusun Pahing	2	3	Greged	Greged
12	Ajeung	Dusun Pahing	2	3	Greged	Greged
13	Anisah	Dusun Pahing	3	3	Greged	Greged
14	Feti	Dusun Pahing	4	3	Greged	Greged
15	Neti	Dusun Pahing	2	3	Greged	Greged
16	Nia	Dusun Pahing	2	3	Greged	Greged
18	Wawat	Dusun Pakuwon	1	2	Nanggela	Greged
19	Iis	Dusun Dukuhsari	2	2	Nanggela	Greged
20	Endang Susanti	Dusun Cikaramat	1	2	Nanggela	Greged

Data Mining

In this data mining process, the k-means algorithm is used to classify sewing equipment assistance in Greged Village, Cirebon Regency. In this research, data processing of sewing equipment uses RapidMiner software using the k-means algorithm, the stages is as follows:

1. Open RapidMiner, then it looks like in Figure 4.1:

International Journal of Social Science (IJSS)
Vol.2 Issue.2 August 2022, pp: 1513-1522
ISSN: 2798-3463 (Printed) | 2798-4079 (Online)

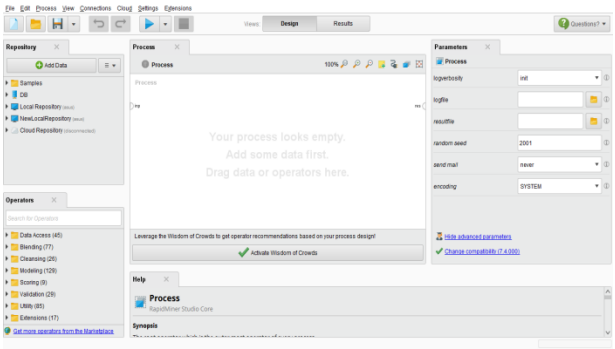

Figure 4.1 RapidMiner Main Screen

Figure 4.1 shows the main page of RapidMiner, which is the display for creating a work page on RapidMiner. Meanwhile, data import is done to enter data to be tested in .xls or .xlsx format. The steps for importing data are:

- Right-click on the main screen of RapidMiner
- Insert Operator
- Data Access
- Files
- Read
- Read Excel, it looks like Figure 4.2:

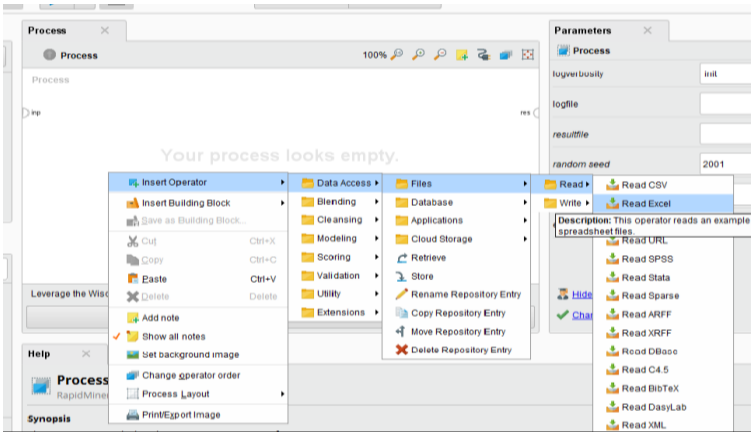

Figure 4. 2 Calling the Read Excel Operator

Read Excel Operators

Excel read display

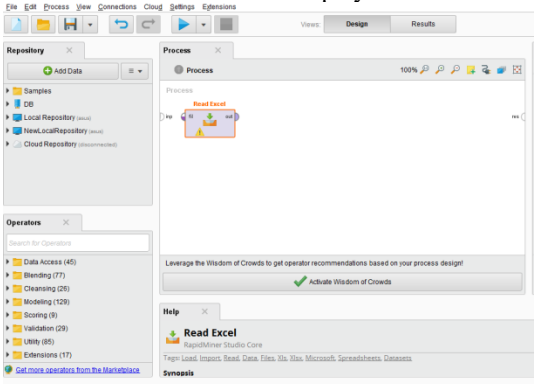

Figure 4. 3 Read Excel operators that have been called

The next stage after the read excel operator has been called, the next step is to import data for processing by:

- click Import Configurations Wizard
- click the file you want to use. If the data has been successfully imported and there are no further errors,

Vol.2 Issue.2 August 2022, pp: 1513-1522
ISSN: 2798-3463 (Printed) | 2798-4079 (Online)
DOI: https://doi.org/10.53625/ijss.v2i2.3085

- click finish.

The following is a display of import data as shown in Figure 4.4 and Figure 4.5

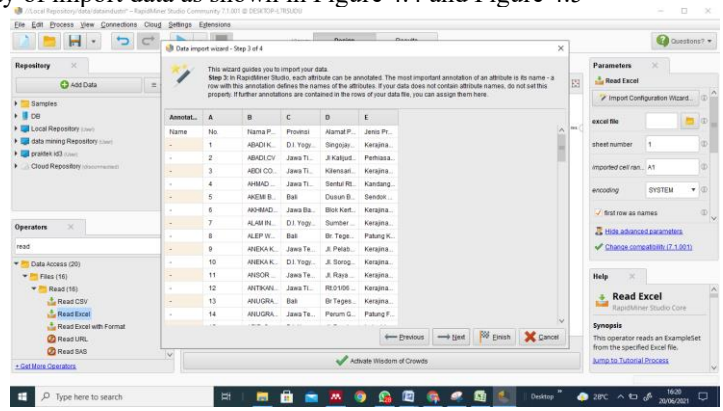

Figure 4. 4 Data import wizard

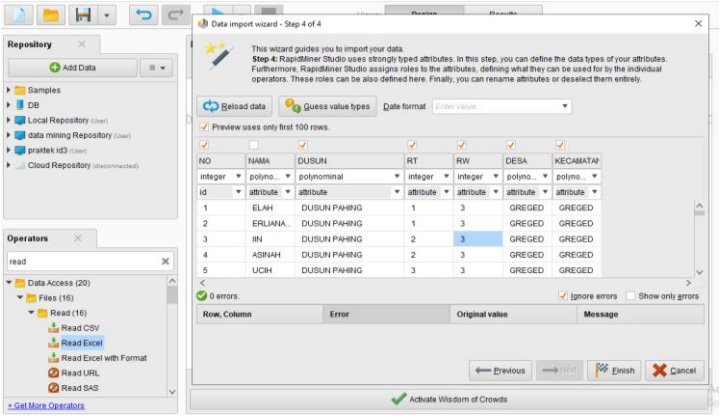

Figure 4. 5 Data import wizard

Nominal to numerical operators

The nominal to numerical data operator is to carry out simple validation by dividing the dataset randomly into two separate data, namely training data and test data along with the split data operator display as shown in Figure 4.6:

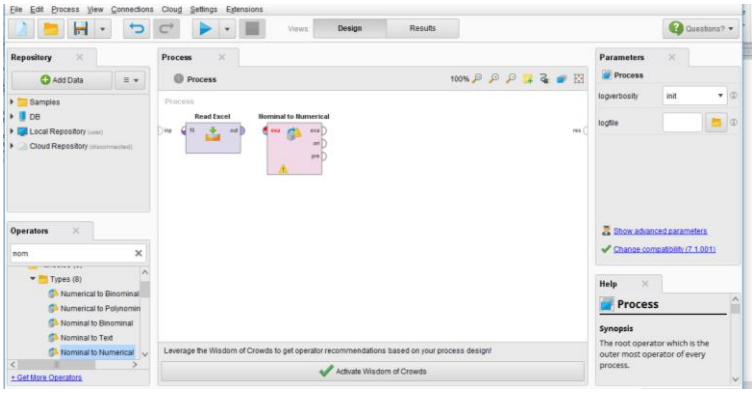

Figure 4.6 Nominal to the numerical operator

Clustering operators

The clustering operator is to specify or set the k:1 to cluster k= to the nth cluster. is to determine the accuracy of the Davies Boulding Index (dbi) data as shown in Figure 4.7

International Journal of Social Science (IJSS)
Vol.2 Issue.2 August 2022, pp: 1513-1522
ISSN: 2798-3463 (Printed) | 2798-4079 (Online)

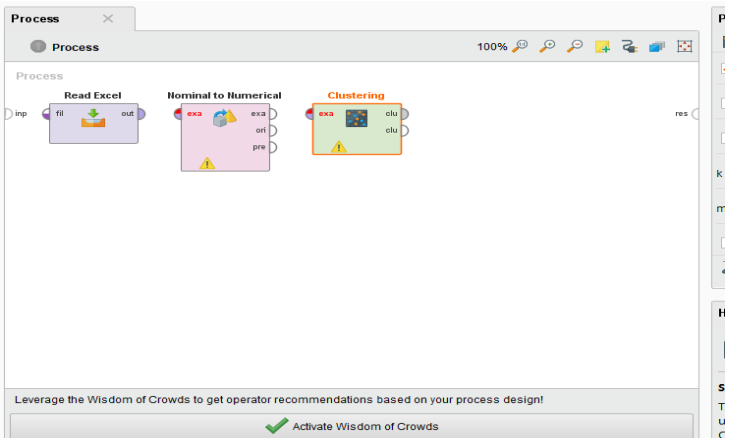

Figure 4. 7. Clustering operators

Cluster distance performance

Cluster distance performance is this operator reads the Example Set from the specified excel file. Display cluster operator distance performance is presented as shown in Figure 4.8

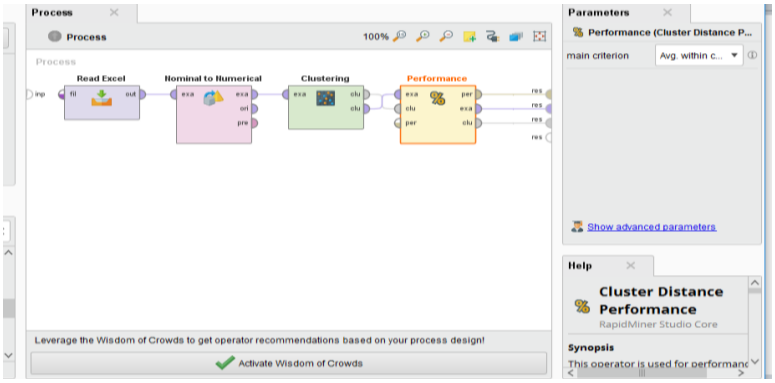

Figure 4.8 Display of cluster distance performance

Cluster model

The k-means process is the process stage for testing the read excel dataset which consists of the training data set and the testing data set connected to the read excel operator. The data model cluster is displayed as shown in Figure 4.9 below:

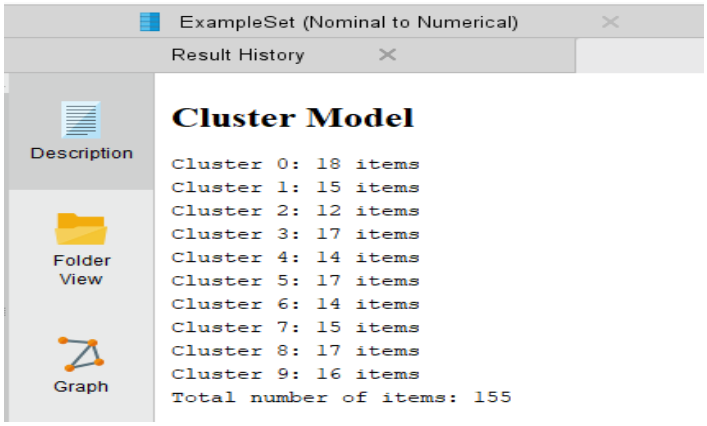

Figure 4.9 Process Cluster model

The explanation of the cluster model from Figure 4.9 which produces a sequence of clusters includes cluster 0: totaling 18 items, cluster 1: totaling 15 items, cluster 2: totaling 12 items, cluster 3: totaling 17 items, cluster 4: totaling 14 items, cluster 5: totaling 17 items, cluster 6: totaling 14 items, cluster 7: totaling 15 items, cluster 8: totaling 17 items, cluster 9: totaling 14 items and cluster 9: totaling 14 clusters.

Vol.2 Issue.2 August 2022, pp: 1513-1522
ISSN: 2798-3463 (Printed) | 2798-4079 (Online)
DOI: https://doi.org/10.53625/ijss.v2i2.3085

Data example set

The data example set is used to apply a previously trained model using training data on unlabeled data (data testing). The goal is to get predictions on unlabeled data (testing data) that do not yet have a label. What needs to be considered is that the testing data must have the same order, type, and attribute roles as the training data. The following is a sample data set table display as shown in Figure 4.10.

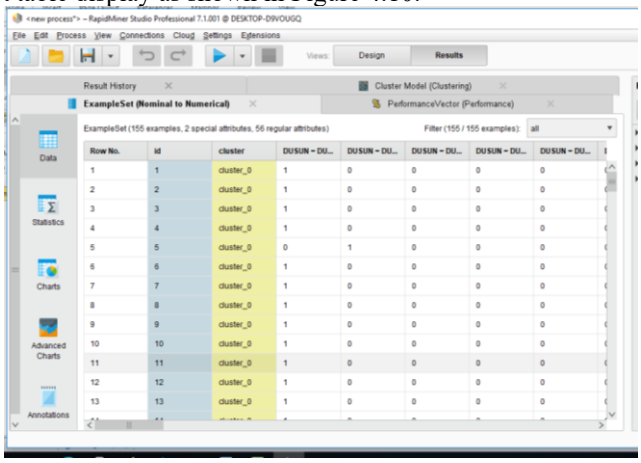

Figure 4.10 Data example set

Display Average centroid distance

The display of the average centroid distance as shown in Figure 4.11

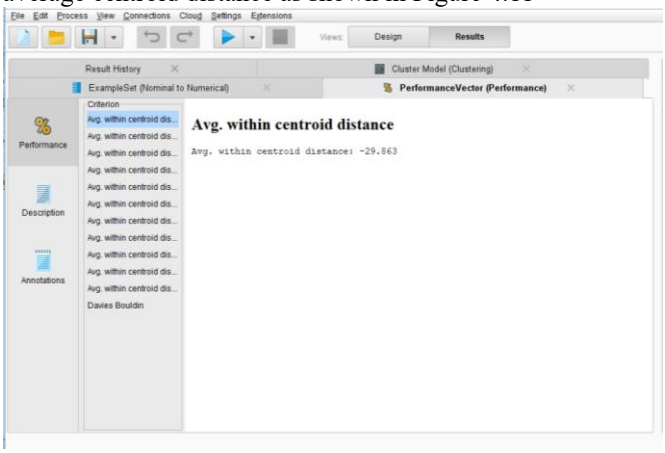

Figure 4. 11 Data example set

Cluster Model Davies Bouldin Index

Display of the Davies Bouldin Index Cluster Model as shown in Figure 4.12

International Journal of Social Science (IJSS)
Vol.2 Issue.2 August 2022, pp: 1513-1522
ISSN: 2798-3463 (Printed) | 2798-4079 (Online)

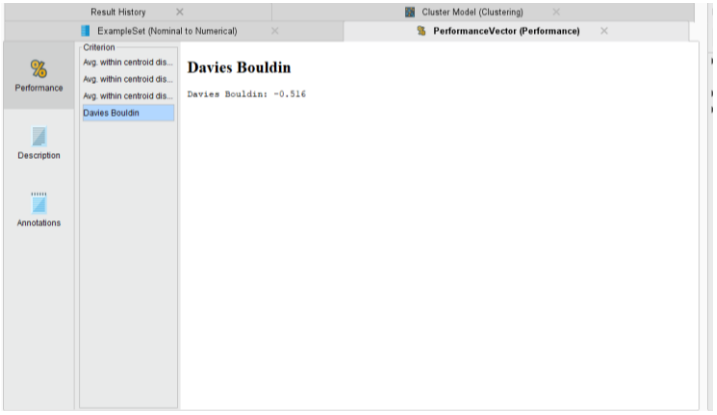

Figure 4.12 Results of Davies Bouldin index

Display cluster graph

Cluster graph display as shown in Figure 4.13

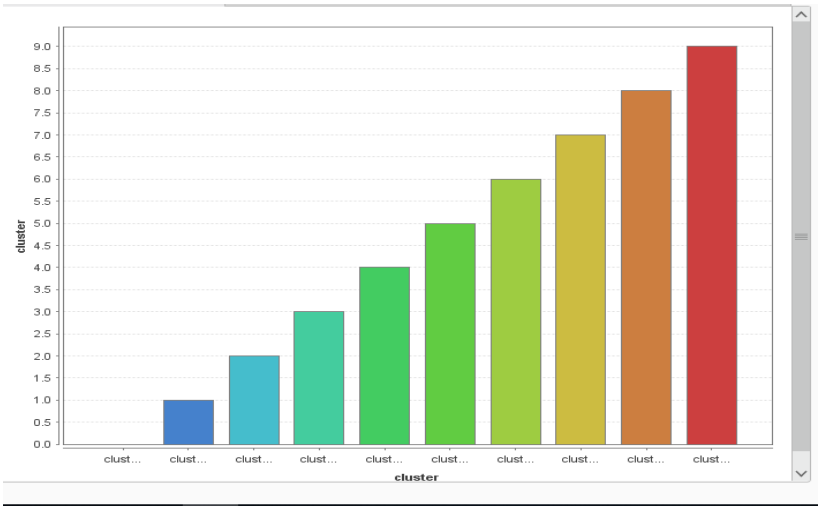

Figure 4.13 Data example set cluster

Vol.2 Issue.2 August 2022, pp: 1513-1522
ISSN: 2798-3463 (Printed) | 2798-4079 (Online)
DOI: https://doi.org/10.53625/ijss.v2i2.3085

Davies Bouldin Index (dbi) process

k	Dbi
2	0.587
3	0.584
4	0.738
5	0.713
6	0.737
7	0.724
8	0.851
9	0.887
10	0.644
11	0.787

Figure 4.14 Davies Bouldin Index Process

Explanation: Figure 4.14

After carrying out the Cluster distance performance process, the resulting K-means clustering k=10, can produce davies bouldin with sequence k2=0,587, sequence k3=0,584, sequence k4=0,588, sequence k5=0,581, sequence k6=0,529, sequence k7=0,612 , the order of k8 = 0.653, the order of k9 = 0.639, the order of k10 = 0.669, then the conclusion of the Cluster distance performance process produces the best dbi of 0.584 in the 3rd order.

CONCLUSION

From the process of implementing data mining through the cluster distance performance process, the resulting K-means clustering k=10, can produce a davies Bouldin index with the order k2=0.587, the order k3=0.584, the order k4=0.588, the order k5=0.581, the order k6=0.529, sequence k7 = 0.612, sequence k8 = 0.653, sequence k9 = 0.639, sequence k10 = 0.669, then the conclusion of the Cluster distance performance process produces the best dbi of 0.584 in the 3rd order.

REFERENCES

[1] 1M. Nanda Variestha Waworuntu, 2Muhammad Faisal Amin. (2018). Penerapan Metode K-Means Untuk Pemetaan Calon Penerima Jamkesda. *Klik - Kumpulan Jurnal Ilmu Komputer*, 190.

[2] Apolinaria Ifon Purnama1), Abdul Aziz2), Anggri Sartika Wiguna3). (2019). Penerapan Data Mining Untuk Mengklasifikasi Penerima Bantuan Pkh Desa Wae Jare Menggunakan Metode Naïve Bayes. *Router Research*, 22-28.

[3] Hendarman Lubisa, Ratna Salkiawatib, Sudirman Hala. (2022). Penerapan Metode Simple Additive Weighting Untuk Penerimaan Bantuan Langsung Tunai Dana Desa. *Fakultas Ilmu Komputer*.

[4] Moh. Fadel Asikin1, Dian Eka Ratnawati2, Mochammad Ali Fauzi3. (2018). Penentuan Penerima Bantuan Ternak Menggunakan Algoritma K-Means & Penentuan Penerima Bantuan Ternak Menggunakan Algoritma K-Means & Naïve Bayes. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 641-649.

[5] Windha Mega Pradnya Dhuhita. (2015). Clustering Menggunakan Metode K-Mean Untuk Menentukan Status Gizi Balita. *Jurnal Informatika Darmajaya*, 160-174.

[6] [I., & Firdaus, A. A. (2019). Penerapan Algoritma C4.5 Untuk Klasifikasi Penentuan Penerimaan Bantuan Langsung Di Desa Ciomas. Teknois : Jurnal Ilmiah Teknologi Informasi Dan Sains, 8(1), 17–28. https://doi.org/10.36350/jbs.v8i1.18

[7] Aprilawati, R., & Tarigan, P. (2019). Implementasi Data Mining Untuk Menentukan Keluarga Yang Layak Mendapat Kartu Pkh (Program Keluarga Harapan) Dengan Metode K-Means Clustering. Jurnal Pelita Informatika, 18(1), 69–73.

[8] Hofmann, M., & Klinkenberg, R. (2014). Data Mining and Knowledge Discovery Series : RAPID MINER Data Mining Use Cases and Business Analytics Applications. https://doi.org/78-1-4822-0550-3

[9] Rahayu, S., & RMS, A. S. (2018). Penerapan Metode Naive Bayes Dalam Pemilihan Kualitas Jenis Rumput Taman CV. Rumput Kita Landscape. Digital Zone: Jurnal Teknologi Informasi Dan Komunikasi, 9(2), 162–171. https://doi.org/10.31849/digitalzone.v9i2.1942

[10] J. Nasional, S. Informasi, and A. Wanto, "Optimasi Prediksi Dengan Algoritma Backpropagation Dan Conjugate Gradient Beale-Powell Restarts," vol. 03, no. 2017, pp. 370–380, 2018. https://www.cirebonkab.go.id/daftar-penerima-bansos-kab-cirebon.

International Journal of Social Science (IJSS)
Vol.2 Issue.2 August 2022, pp: 1513-1522
ISSN: 2798-3463 (Printed) | 2798-4079 (Online)

THIS PAGE HAS INTENTIONALLY BEEN LEFT BLANK