International Journal of Software Engineering and Informatics
Vol. 1, No.1 March 2023 Pp 63-70.
ISSN: xxxx xxxxx,

K-Means Clustering in Determining the Eligibility of Recipients of Assistance for the Poor Case Study of Village Sukoharjo III

Aviv Fitria Yulia[1], Ratnasari[2] ,Panji Bintoro[3]

[1,2,3] Aisyah Pringsewu University, Indonesia

Article Info

Keywords:

Data Mining
K-Means
Village

ABSTRACT

Poor people are residents who have an average per capita income below the poverty line. Assistance funds for the poor are one of the government's program efforts to reduce the level of community poverty. The purpose of this research is to test the eligibility of beneficiaries so that they are right on target - really poor people who deserve to receive this assistance. This is because there are many criteria that must be considered in determining the eligibility of recipients of aid funds with a total of 4072 residents, the researchers used data utilization techniques or also called Data Mining. One of the data mining methods is quite popular, namely clustering using the K-Means algorithm. Processing the selection data for receiving aid funds using the K-Means algorithm results in a Davies bouldin index of 0.738. These results are considered quite good because the closer the results are to zero, this study produces three groups: 1218 residents are eligible to receive assistance, 2514 residents are considered to receive assistance, 1040 are not eligible to receive assistance.

Author Correspondence:

Aviv Fitria Yulia,
Software Engineering Study Program
Aisyah Pringsewu University, Indonesia
Email: avivfitriayulia0@gmail.com

1. INTRODUCTION

There are three important roles of government efforts in creating improved welfare in society, namely the role in public services, the role in development/empowerment and the role in protection according to Wasistiono (2001).

The Government's efforts strive to improve the welfare of its people, namely by making programs that can support welfare, improve people's living standards in various fields, for example, in the fields of education, health and community economics that can be used as a benchmark for community welfare.

Based on data released by the Central Statistics Agency (BPS) In the People's Welfare statistics, namely in November 2022, the current social assistance that the community receives is in the form of non-cash or family hope assistance in accordance with the criteria of the beneficiaries.

Based on the Presidential Regulation No. 63 of 2017 concerning the distribution of non-cash social assistance provided to the community, for example, assistance in the form of money, goods or services to individuals, families, groups or poor people, or poor people who in this case are vulnerable to social risks (financial, 2017).
Poverty is a condition where an individual or group of people who in this case are unable to meet or maintain a decent life.so it can be called poverty, which is a condition of economic inability to provide basic food and non-food needs.
Poor people are people who have an average per capita income below the poverty line, are unable to meet the needs of life and are vulnerable to the risk of malnutrition
The assistance fund for the poor is one of the government's efforts to reduce the poverty rate of the community and ensure people's lives to live a decent and secure life in food and non-food needs.

The purpose of this study is to test the eligibility of aid recipients to be right on target, so that it is really poor people who are eligible to receive the assistance, this is because of the many criteria that must be considered in determining the eligibility of recipients of the aid funds.

Based on the above problems that have been described, the purpose of this study is to conduct testing using Data Mining Algorithm K-Means clustering to make it easier for the government to predict the receipt of aid funds for poor people in pekon sukoharjo III using criteria, to find out the recipients of aid funds are really feasible and on target

2. RESEARCH METHODS

Stages in completing feasibility research on Receiving Aid Funds for the Poor Using the K-Means Clustering Algorithm Case study: Sukoharjo III can be seen from the picture below:

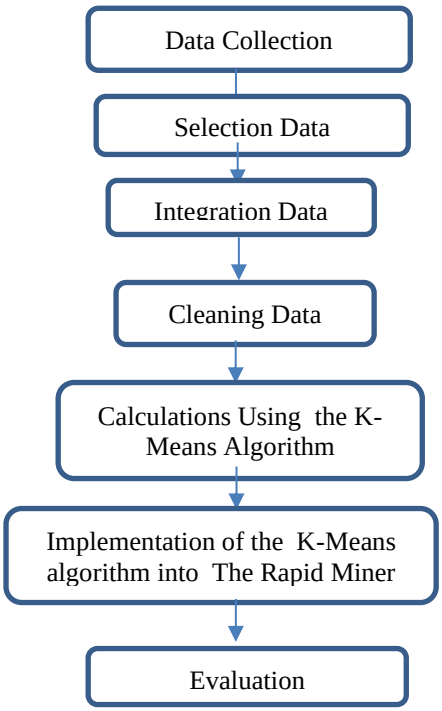

Figure 1.Research Framework

a.Data Collection
is the initial stage to find information related to the application of the K-Means Clustering method to determine the eligibility of recipients of aid funds for the poor by interviews, observations and studies literature to the government of pekon sukoharjo III.At this stage researchers obtain some data that will be processed such as: name, address, hamlet, rw, rt, gender, place of birth, date of birth, religion, occupation, marital status, income with Total data 4772.

b.Selection Data
Data Selection is choosing good data to be used at this stage researchers use feature selection to reduce less relevant data.in this selection researchers use feature selection techniques using the Information Gain method.

According to Verghese & Sushmita, Information gain is one method of feature selection where in the Information Gain process features will be ranked, and the result of the largest ranking is the most feature. relevant and have a strong connection with the associated data set.

Researchers use the implementation of information gain in determining attributes that are suitable for use, in this case researchers use machine learning, namely rapid miners and use data on residents of Pekon Sukoharjo III taken on Date 2023-03-01.:D Below are the results of feature selection with the information Gain method using machine learning:

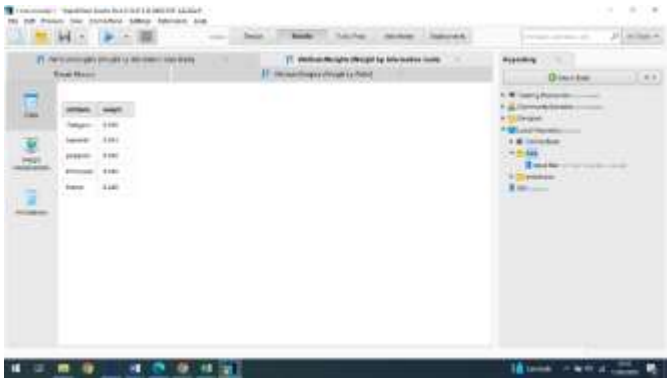

Figure 1. Implementation Results

Figure 1 is the result of the implementation of the information gain method dataset using rapid minner, obtained the results of attribute cracking, namely dependents have a weight value of .0000, the attribute of home ownership has a weight value of 0.004, the income attribute has a weight value of 0.006, the job attribute has an attribute value of 0.025, and the name attribute has a weight value of 0.220. And researchers took the top 4 scores.

c.Data Integration

In this data integration stage, a process is carried out to give value weights to the original data where this data is given value weights, namely nominal data is changed into numerical form to make it easier to Researchers calculated K-Mean clustering with 4772 data.

Below is the Data weighting process

Table 1 Dependents conversion

Dependents	Weight
9->10	5
7-8	4
5-6	3
3-4	2
1-2	1

Table 1 is a process table that converts data from dependents into value weights if dependents 1-2 then given a value weight of 1, if dependents 3-4 then weights will be given 2 , if Dependents 5-6 will be given a value weight of 3, if dependents 7-8 will be given a value weight of 4, and if more than 9 then a value weight of 5 will be given.

Table 2 Ownership Conversion

Home Ownership	Weight
Ngontrak,	3
Living with parents,	2
own own,	1

Table 2 is a process table that explains the process of converting home ownership data into value weights with the process , namely if you have your own house, you will be given a value weight of 1 , if you live with your parents, you will be given a value weight of 2, if Ngontrak it will be given a value weight of 3,

Table 3 Earnings Conversion

Earnings (/month)	Weight
• <IDR 500,000 • IDR 500,000-IDR 999,999 • IDR 1,000,000 – IDR 1,999,999 • IDR 2,000,000 – IDR 2,999,999 • IDR 3,000,000 -<IDR 4,000,000	• 5 • 4 • 3 • 2 • 1

Table 3 is a process table that explains the process of converting income data into value weights with the process, namely if income with an amount of >IDR 4,000,000 will be given a value weight of 1, if income with a total of IDR 2,000,000 – IDR 2,999,999 will be given a value weight of 2, if income with a total of IDR 1,000,000 – IDR 1,999,999 will be given a value weight of 3 , if income with an amount of IDR 500,000-IDR 999,999 will be given a value weight of 4, if income with an amount of <IDR 500,000 will be given a value weight of 5.

4. Job Conversion Table

Earnings (/month)	Weight
• Not Yet/Not Working Students/STUDENTS,Freelance Daily Laborers • Farmer,Livestock,Electrician, • Honorary Teacher,Honorary Staff • Merchant,Self-employed,Private employee, • Doctor, Civil Servant, Nurse	• 5 • 4 • 3 • 2 • 1

Table 4 is a job conversion table that explains the process of converting work data into value weights with the process, namely if the work of doctors, civil servants, nurses will be given a value weight of 1, and if the work of traders, self-employed, private employees will be given a value weight of 2, if the work of honorary teachers, honorary staff will be given a value weight of 3, If farmers, livestock, electricians, will be given a value weight of 4, and if the work has not been / does not work, students / students, freelance daily laborers will be given a value weight of 5.

Table 5 Conversion Results

	Nama	Tanggungan	Kepemilikan	Penghasilan	Pekerjaan	Tanggungan	Kepemilikan	Penghasilan	Pekerjaan
2	RASHYA CITRA MAY ONE	2	NGONTRAK	Rp. 1.000.000 - Rp. 1.999.999	BELUM/TIDAK BEKERJA	1	3	3	5
3	NAKIDO PERTAI	2	HILIR SENDIRI	Rp. 500,000 - Rp. 999,999	BURUH HARIAN LEPAS	1	1	4	5
4	MAULIA FANA ZAHRA	3	NGONTRAK	Rp. 2,000,000 - Rp. 4,999,999	PELAJAR/MAHASISWA	2	3	4	5
5	NOVA KARDILA	1	MILIK SENDIRI	Rp. 500,000 - Rp. 999,999	MENGURUS RUMAH TANGGA	1	1	4	5
6	[illegible]	1	HILIR SENDIRI	Kurang dari Rp. 500,000	BURUH HARIAN LEPAS	1	1	5	5
7	MUHAMMAD ELFARIZKI	1	NGONTRAK	Rp. 500,000 - Rp. 999,999	BELUM/TIDAK BEKERJA	1	3	5	5
8	MUHAMMAD FADEL KHAM	1	NGONTRAK	Kurang dari Rp. 500,000	PELAJAR/MAHASISWA	1	3	5	5

3.RESULTS AND DISCUSSION

3.1 Results

a. Clustering Process

At this stage, it is a process of grouping recipients of aid funds for the poor of pekon sukoharjo III, by applying the K-Means algorithm with input parameters as follows: the number of datasets as follows: n data and number of centroid initializations k = 3 according to the study. The data to be processed amounted to 4772 residents with the application of the K-Means algorithm.The experiment will be carried out using the following parameters:

Number of Clusters : 3

Total Data : 4772

Number of attributes: 4

b.Testing Tools

In this research process , the author utilizes the rapid miner tool as a tool in testing data sets. With the steps below:

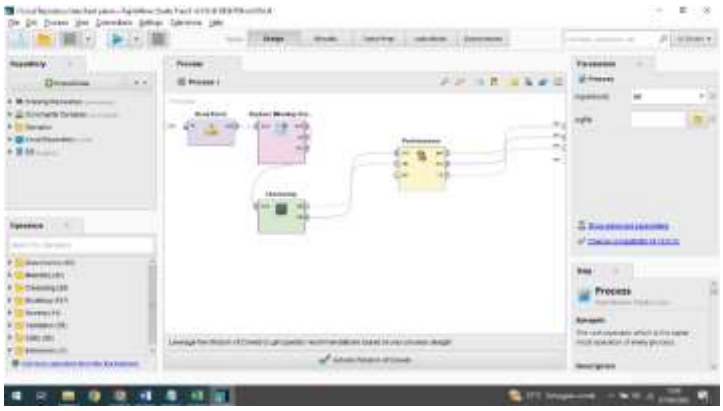

Figure 3.1 Process Design

Figure 4.1 is the process design stage at this stage carried out using 4 processes, namely:

a.Read excel

What is done at this stage is the operation of inputting datasets in the form of extension files.xls with data on all residents of Pekon Sukoharjo 3 totaling 4772 poor.

b.Replace missing value

At the Replace Missing Value stage, what is done is to fill in the missing value with the maximum value.

c.Clustering

At the clustering stage as the algorithm used in this study.

d.Performance

At this stage, a davies bouldin index value search operation is carried out.

Row No.	id	cluster	Tanggungan	Kepemilikan	Penghasilan	Pekerjaan
1	1	cluster_2	1	3	3	5
2	2	cluster_1	1	1	4	5
3	3	cluster_1	2	3	4	5
4	4	cluster_1	1	1	4	5
5	5	cluster_1	1	1	5	5
6	6	cluster_2	1	3	3	5

Figure 4.2 Example set result

Figure 3.2 is an example set result where at this stage the results of data clustering are displayed. Cluster labels are divided into three groups, namely cluster 0, cluster 1, cluster 2. This division is based on the results of the proximity of each data to the closest distance (k).

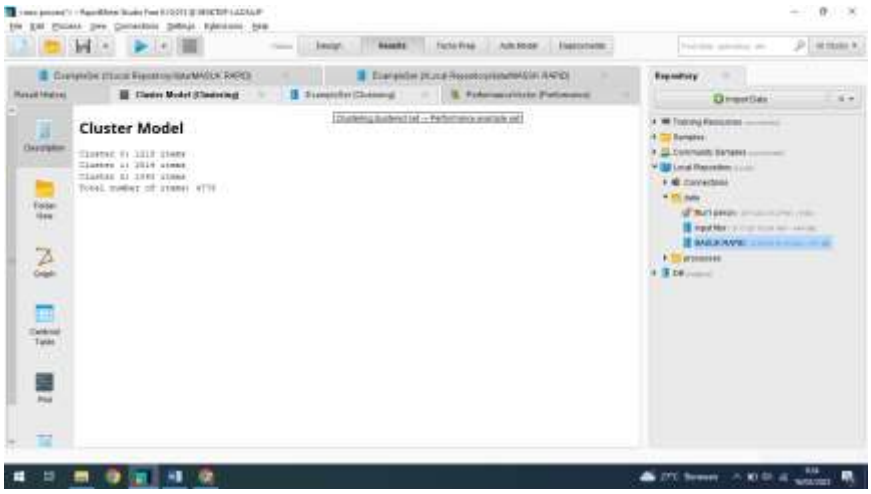

Figure 4.3. Result

Figure 3.3 is a cluster model where at this stage the results of data division between each cluster are displayed.

Cluster 0 has 1218 members, Cluster 1 has 2514 members, Cluster 2 has 1040 members from a total of 4772 datasets tested.

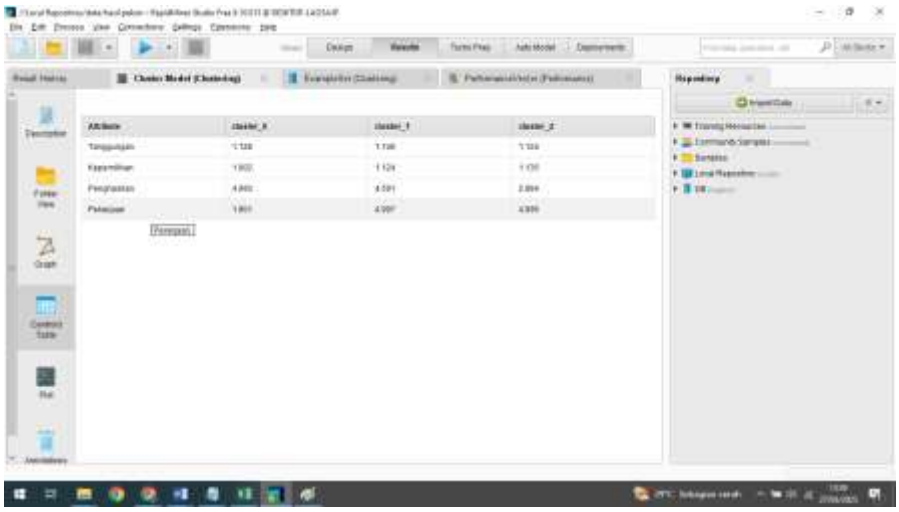

Figure 4.3 Centroid table

Figure 3.4 is a centroid table where at this stage the value of the center point in each cluster is displayed. This value makes a reference for calculations in each dataset by measuring the proximity of values to each central point of the cluster.

4.2 Discussion

a. Custering K-Means Results

After testing with the rapid miner tool, it can be concluded as follows:

- Cluster 0 has 1218 members,
- Cluster 1 has 2514 members,
- Cluster 2 has 1040 members
- from a total of 4772 datasets tested

After the cluster is formed, the following conclusions are obtained:

Table 3.1 Data description based on clusters

Cluster	Description
0	Cluster members deserve help
1	Members of the cluster may be considered receiving assistance
2	Cluster members do not deserve help

Table 4.1 is a description of data based on clusters where After the clustering process is complete, performance operations are carried out to determine the value of the Davies bouldin index which aims to maximize distance measurements between clusters and minimize distances between members in a cluster.

Table 3.2 Performance vector operation results

PerformanceVector	Value
Avg. within centroid distance	0,906
Avg. within centroid distance_cluster_0	1.304
Avg. Within centroid distance_cluster_1	0.788
Avg. Within centroid distance_cluster_2	0.723

Davies Bouldin	0.738

Table 3.2 is the result of the performance vector operation where the evaluation of the results of the average within centroid distance close to 0 means that each member in the cluster is in close proximity. Evaluation using the davies bouldin index has an internal cluster scheme seen from the quantity and proximity between cluster results.

The smaller the davies bouldin index value obtained (non-negative)>=0), the better the cluster obtained from grouping using the clustering method. The calculation results using the K-Means algorithm show a value of 0.738. This number means that each object in the cluster has a fairly good similarity because it is close to the number 0.

CONCLUSION

Based on the results of research conducted by the author, the following conclusions can be drawn:

1. The application of the K-Means algorithm divides the dataset into three groups: deserving of assistance, can be considered receiving assistance and not deserving of assistance.
2. The test results get a devies bouldin index value of 0.738 which means that the similarity between cluster members is quite good.
3. The results of this test are very worthy to be used as recommendations for beneficiaries in the Sukoharjo III pekon

ACKNOWLEDGMENTS

Pekon Sukoharjo III who has provided opportunities as a place for research, Aisyah Pringsewu University which has provided a forum for researchers

Reference

[1] Anwar Pauji, Aisyah, S., Surip, A., Saputra, R., &; Ali, I. (2022). Implementation of the K-nearest neighbor algorithm in determining recipients of direct cash assistance . KOPERTIP : Scientific Journal of Information and Computer Management, 4(1), 21–27. https://doi.org/10.32485/kopertip.v4i1.114

[2] Damuri, A., Riyanto, U., Rusdianto, H., &; Aminudin, M. (2021). Implementation of Data Mining with Naïve Bayes Algorithm for Eligibility Classification of Food Aid Beneficiaries . JURIKOM (Journal of Computer Research), 8(6), 219. https://doi.org/10.30865/jurikom.v8i6.3655

[3] Dian Anggriani, S., Syahril, M., &; Mariami, I. (2020). Data Mining Algorithm C4.5 to analyze the recipient population of the Family Hope Program (PKH). Journal of Informatics and Computer Management Science (Jurnal SAINTIKOM), x.

[4] Filki, Y. (2022). K-Means Clustering Algorithm in Predicting Village Fund Direct Cash Transfer (BLT) Recipients. Journal of Business Economic Informatics, 4, 166–171. https://doi.org/10.37034/infeb.v4i4.166

[5] Hasanah, R. L., Hasan, M., Pangesti, W. E., Wati, F. F., &; Gata, W. (2019). Classification of recipients of village aid funds using the KNN (K-nearest neighbor) method. Journal of Techno Nusa Mandiri, 16(1), 1–6. https://doi.org/10.33480/techno.v16i1.25

[6] Nursia, A. N., Ramdhan, W., &; Kifti, W. M. (2022). Analysis of the Eligibility of Covid-19 Beneficiaries Using the K–Means Method. Building of Informatics, Technology and Science (BITS), 3(4), 574–583. https://doi.org/10.47065/bits.v3i4.1399

[7] Introduction, I. (2023). Data Mining Determines Cluster Recipients of Assistance Programs with the K-Means Method. 7(1), 291–300.RPL 2022_2023 GNP Schedule.pdf. (n.d.).

[8] Rusdiansyah, R., Supendar, H., &; Tuslaela, T. (2021). Data Mining using K-means method for feasibility selection of Non-cash food Assistance recipients in the Era of Covid-19. SinkrOn, 6(1), 25–33. https://doi.org/10.33395/sinkron.v6i1.11101

[9] Shah, S. (2021). Data mining to predict student achievement of kosgoro institute of business and informatics 1957 based on socioeconomic parents and past achievements. 2(2), 60–71.

[10] Trisaputra, Y. (2016). Profile Classification of High School / Vocational Students Entering PTN (State Universities) with k-Nearest Neighbor Profile Classification of High School / Vocational Students Entering PTN (State Universities) with k-Nearest Neighbor Yuandri Trisaputra, Indriyani, Shellafuri Mar. September 2015, 0–15