

Data Mining using K-means method for feasibility selection of Non-cash food Assistance recipients in the Era of Covid-19

Rusdiansyah^{1)*}, Hendra Supendar²⁾, Tuslaela³⁾

¹⁾²⁾Universitas Bina Sarana Informatika, ³⁾ Universitas Nusa Mandiri

¹⁾Rusdiansyah.rds@bsi.ac.id, hendra.hds@bsi.ac.id, tuslaela.tll@nusamandiri.ac.id,

Submitted : Aug 28, 2021 | **Accepted** : Oct 2, 2021 | **Published** : Oct 15, 2021

Abstract: All countries in the world are currently experiencing a severe economic crisis following the outbreak of the COVID-19 outbreak. In Indonesia, the Large-Scale Social Restriction (PSBB) policy is reported to have increased the number of poor people. Social assistance is a government program to improve the social welfare of the lower economic community. In carrying out the program, the central government and local governments coordinate with each other so that the program is right on target without any element of fraud. In the neighbourhood of Rukun Warga 001, Kelapa Dua Village, there are still obstacles in selecting the eligibility for social assistance recipients, namely Non-Cash Food Aid. The data on the poor are not in accordance with the actual conditions. In this study, to implementing data mining with the K-Means Algorithm. The K-Means Clustering algorithm is used to classify people who are classified as eligible to receive social assistance and those who are not entitled to receive social assistance. The data sample used is the data of Rukun Warga 001, Kelapa Dua Village. The results of this study indicate that cluster 1 with the appropriate category of receiving social assistance according to government programs in the Rukun Warga 001 neighbourhood of Kelapa Dua sub-district amounted to 13 families. And cluster 2 in the category of not eligible to receive social assistance amounted to 97 heads of families out of a total of 110 heads of families in RW 001.

Keywords: Social Assistance, COVID 19, K-Means Clustering, Data Mining

INTRODUCTION

Indonesia is a developing country consisting of 34 provinces with different populations and percentages of poverty levels in each district or city (Sinukun, 2018). Based on the results of a district or city Central Statistics Agency (BPS) survey from 2015 to 2019, it shows the ups and downs of the population and the percentage of poverty levels (Probosiwi, 2016). The achievement of this poverty level is important and becomes the main focus of the city government together with local governments to consistently reduce the percentage of the poverty rate by creating a Non-Cash Food Assistance (BPNT) program which aims to reduce the burden of spending on poor and vulnerable families in meeting their food needs, increasing social-economic conditions in the community, as well as increasing human resources, especially in the poor group (Triana, 2018). In terms of efforts to overcome the problem of poverty, accurate poverty data is needed so that it can implement social assistance programs in the form of Non-Cash Food Assistance (BPNT) so that it can create an effective distribution of social assistance programs that are right on target, encourage financial inclusion, and are equitable (Anggleni, 2018). Determining whether or not recipients are eligible to receive social assistance, namely Non-Cash Food Assistance (BPNT) will have an impact on decreasing the percentage of poverty. Data mining is the automatic analysis of large or complex data with the aim of finding important patterns or trends that are usually not aware of their existence. In the Rukun Warga 001 Kelapa Dua, there are still problems in the selection of eligibility for social assistance recipients, namely Non-Cash Food Assistance. The data of the poor are not following the actual conditions. People who were in poor condition may increase to the middle group, while on the other hand there are poor people who have not been previously registered. As a result, residents who are considered in top condition or can receive this Non-Cash Food Aid, while residents in poor condition do not receive it. This research uses a data mining approach using the K-Means Clustering method to group social assistance recipients on target and evenly. There are 2 clusters or

*name of corresponding author



groups that are eligible to receive social assistance as the first cluster and those who are not eligible to receive social service as the second cluster. Starting with the first-level local government or commonly called the Rukun Tetangga who directly meet face-to-face with the citizens, who know the economic conditions, and environmental conditions, it is hoped that the government's program to reduce the percentage of the poverty rate will run smoothly and in accordance with applicable regulations.

LITERATURE REVIEW

Data mining is part of the Knowledge Discovery in Databases process which consists of several stages such as data selection, pre-processing, transformation, data mining, and evaluation of results. (Widaningsih, 2019). The utilization of data mining is helpful for determining groups of students' educational values and their parents' salaries (Nugraha, 2021), while the algorithm used for the grouping process uses the k-means algorithm (Marlina, 2018). The K-means algorithm is a form of non-hierarchical clustering that tries to partition the data into one or more clusters. With the k-means algorithm, the data variables are grouped based on the midpoint of a cluster (Ade Bastian, 2018).

Below you can see the research that has been done by other researchers using the K-means method:

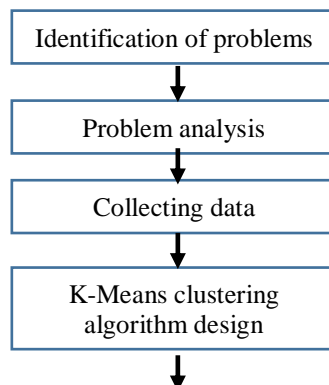
1. This study developed a data mining concept to help predict students who get scholarship programs using the K-Means Clustering algorithm. This algorithm partitions data into clusters, so that data with the same characteristics are grouped into the same cluster and data with different characteristics are grouped into other clusters. This data mining application displays the results in the form of each data grouped based on the value and salary of parents as a consideration for getting a scholarship (Wicaksono, 2016).
2. This study aims to find beneficiaries who deserve cash assistance, social assistance, non-cash assistance, food assistance, and families of hope with the following criteria: poor people, not working, and not receiving more than one aid. The results of this study are the criteria do not work with a value of C1 0.250 and C2 0.969, which has the highest value of the feasibility level compared to the other two criteria (Sahril, 2021).
3. Poverty classification can help prevent the mistargeting of social assistance receipts. The researcher applied the K-Means method made with 4 clusters. From the calculation of the Davies Bouldin Index in the Performance Vector in Rapidminer software, it produces a value of 0.07 where if the value is closer to 0 (zero), then the data will be better. From the calculation results obtained 101 data in the first cluster which will later receive Rastra assistance, 165 data in the second cluster and 46 data in the third cluster that will receive PKH (Hope Family Program) assistance which will be selected again later with a total of 85 data in the fourth cluster that will receive Non-Cash Food Assistance (Sunia, 2019).
4. Researchers conducted clustering to help the Jambi City Social Service in poor grouping families so that assistance could be channelled appropriately. By observing the problem, data mining is needed to classify aid recipients using the K-Means method in clustering the poor. The calculation results show that from 512 samples of poor population data in South Jambi District, with a total of 5 clusters, with 3 iterations, 13 residents are obtained as members of cluster 1, 153 residents are in cluster 2, 129 residents are in cluster 3, 138 residents are in cluster 3 cluster 4, and 79 residents are in cluster 5 (Khoir, 2019).

According to the four works of literature, which became the number two research guideline, with added variables including the variables Age, Gender, home, status, education, Work.

METHOD

Research Design

The researcher's plan will be use as a guide for the progress of the research (Widiani, 2018).



*name of corresponding author



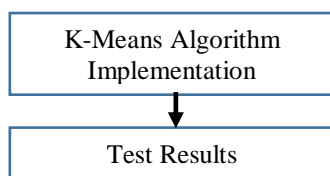


Figure 1. Research Design

Explanation from Figure 1, the steps in the research method to find solutions to the analyzed discussion themes.

1. Identification of problems
Problem identification is the first step in knowing the problem to be studied so that the right solution can be found. Identification of the situation in this study is that there are still beneficiaries who do not meet the requirements of recipients of assistance. There are no criteria for determining beneficiaries who are considered truly worthy of receiving aid.
2. Problem analysis
Problem analysis is a step to identify problems that have been determining in the scope or boundaries. The analysis of the situation required is how to classify and provide the basis for beneficiaries who meet the requirements of beneficiaries.
3. Collecting data
In collecting data, observations, interviews, and questionnaires were distributed to find out clearly the existing problems. The data taken is data related to the conditions of existing beneficiaries.
4. K-Means clustering algorithm design
The process design from the system model to the k-means algorithm is used by determining the cluster centre randomly from the beneficiary data and then calculating the closest distance to the cluster centre using the Euclidian Distance formula.
5. K-Means Algorithm Implementation
Implementation is the application of the k-means algorithm stages where the entire K-means Clustering process is applied—starting from determining the centre of the cluster then looking for the closest distance to the center of the cluster with the Euclidian Distance formula until the iterative process looks for a new cluster value from the previous stage.
6. Test Results
This stage is the process of testing and analyzing the results obtained from the Rapidminer software, and the system is tested with stages and modelling.

Object of research

Researchers collect data on residents of the area of RW 001 Kelapa Dua Village. The data used is only taken some data in accordance with the provisions given by the Chairperson of RW 001 Kelapa Dua Village because according to government regulations regarding raw data, a maximum of 10% may only be used. The initial data used in this study were 110 heads of families from the Rukun Warga 001 neighbourhood of Kelurahan Kelapa Dua.

Population and Sample

The population is the whole of a collection of elements that have a number of common characteristics, which consist of fields to be studied (Samsuri, 2018). The population of the research conducted by the author is the Rukun Warga Area 001, Kelurahan Kelapa Dua. The sample is a sub-group of the people selected for use in research. The sample from the research conducted by the author is 10 Heads of Families in one RT per RW, so the total example in this study is 110 Heads of Families.

Research variable

The research variables to be studied are based on the general conditions for assisting the government:(Sahril, 2021)

1. Age
2. Gender
3. home status
4. Education
5. Work

*name of corresponding author



Data Analysis Model

Data analysis is an effort made by working with data, organizing data, sorting it into manageable units, looking for and finding patterns, finding out what is important and what is learned and deciding what to tell others.(Yefni, 2019). In this study, using quantitative data analysis methods, where quantitative research is research carried out to carry out hypotheses that have been prepared at the beginning of the study and are based on the theory built(ZALUCHU, 2020).

RESULT

Calculation of the K-Means Clustering Algorithm

The calculation is based on a sample of 110 beneficiary data which is divided into 2 (two) clusters and refers to the 5 (five) variables studied. The calculation starts by determining the centroid value randomly in the first iteration and then calculating the closest distance to get the cluster value because the k-means method allocates data into clusters from data that has the closest distance to the centre point of each cluster by calculating the distance of each data using the Euclidean formula. Distance. The next step is to do the second iteration by recalculating the closest distance to the centroid value that has been determined from the average value in the first iteration to get the cluster value in the second iteration. This study stops at the second iteration because no more data is moving from each cluster.

Table 1. Comparison table of iteration 1 and iteration 2 . clusters

1 iteration			2 iteration		
Number	C1	C2	Number	C1	C2
1		1	1		1
2	1		2	1	
3		1	3		1
4		1	4		1
5		1	5		1
6		1	6		1
7	1		7	1	
8		1	8		1
9		1	9		1
10		1	10		1
11	1		11	1	
12		1	12		1
13		1	13		1
14		1	14		1
15		1	15		1
16	1		16	1	
17		1	17		1
18		1	18		1
19		1	19		1
20	1		20	1	
21		1	21		1
22		1	22		1
23		1	23		1
24		1	24		1
25		1	25		1

*name of corresponding author



26		1	26		1
27		1	27		1
28		1	28		1
29		1	29		1
30		1	30		1
31	1		31	1	
32		1	32		1
33		1	33		1
34		1	34		1
35		1	35		1
36		1	36		1
37		1	37		1
38		1	38		1
39		1	39		1
40		1	40		1
41	1		41	1	
42		1	42		1
43		1	43		1
44		1	44		1
45		1	45		1
46		1	46		1
47		1	47		1
48		1	48		1
49		1	49		1
50		1	50		1
51		1	51		1
52		1	52		1
53		1	53		1
54		1	54		1
55		1	55		1
56	1		56	1	
57		1	57		1
58		1	58		1
59		1	59		1
60		1	60		1
61		1	61		1
62	1		62	1	
63		1	63		1
64		1	64		1
65		1	65		1
66		1	66		1

*name of corresponding author



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

67	1		67	1	
68		1	68		1
69		1	69		1
70		1	70		1
71		1	71		1
72		1	72		1
73		1	73		1
74		1	74		1
75		1	75		1
76		1	76		1
77		1	77		1
78	1		78	1	
79		1	79		1
80		1	80		1
81		1	81		1
82		1	82		1
83		1	83		1
84		1	84		1
85		1	85		1
86	1		86	1	
87		1	87		1
88		1	88		1
89		1	89		1
90		1	90		1
91		1	91		1
92		1	92		1
93		1	93		1
94		1	94		1
95		1	95		1
96		1	96		1
97		1	97		1
98		1	98		1
99		1	99		1
100		1	100		1
101		1	101		1
102	1		102	1	
103		1	103		1
104		1	104		1
105		1	105		1
106		1	106		1
107		1	107		1

*name of corresponding author



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

108	1	108	1
109	1	109	1
110	1	110	1

In table 1, the calculation is declared complete if the comparison between the clusters in the two iterations is the same value and the position of the cluster does not change. Then iteration only until the second iteration only.

Rapidminer Software Testing

The data that has been processed manually in the form of iteration data will be proven whether the data is in accordance with the desired decision from the relationship between the data distance and the cluster value by the k-means algorithm. Therefore, it is necessary to test using Rapidminer Studio Software.

When the selection of the K-Means Clustering Method is complete, then connect the Read Excel operator with the K-Means operator by drawing a line. The next step is to select Cluster Distance Performance, then connect as shown below.

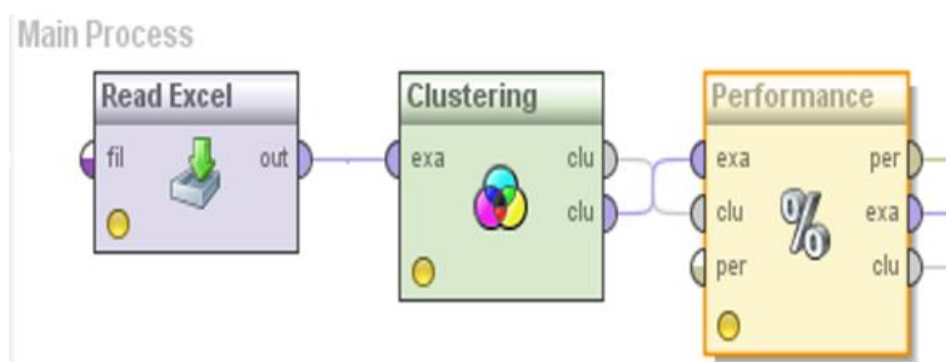


Figure 2. Display Connectivity

In figure 2, this stage reconnects the K-Means operator with the Cluster Distance Performance operator by crossing the first line with the second line and vice versa. Then reconnect with the output port so that it can be processed. Then click Run to generate clusters. After the data is successfully imported, the data is dragged to the process panel then the data will be linked to the k-means previously searched for in the operator panel and after that the project can be run.

Cluster Model

Cluster 0 : 13 Items
Cluster 1 : 97 Items
Total Number of items : 110 items

Figure 3. Implementation Results

In Figure 3 using K-Means Clustering modeling as shown above, with the initialization of the number of clusters as many as 2 processes, the results obtained with clusters formed by grouping cluster 0 totaling 13 data and cluster 1 as many as 97 data.

DISCUSSIONS

The classification process requires the existence of a rule to determine the eligibility for social assistance in the form of Non-Cash Food Assistance. In this study, clusters are grouped into two clusters, namely:

1. Cluster 1 = eligible to receive Non-Cash Food Assistance
2. Cluster 2 = those who are not eligible to receive Non-Cash Food Assistance

The results in the first cluster are groups on data with No. 2, No. 7, No. 11, No. 16, No. 20, No. 31, No. 41, No. 56, No. 62, No. 67, No. 78, No.86, and No.102. While the rest are included in the second cluster or group.

*name of corresponding author



In the Cluster menu, the Rapid Miner model shows the results of grouping cluster 0 with 13 data and cluster 1 with 697 data. Regarding calculations using the K-means algorithm, there are the same results as testing with the Rapidminer software.

CONCLUSION

Based on the results of the research and discussion in the previous chapter, it can be concluded that the results of the analysis carried out by the author, namely the application of manual calculation data mining in Excel Software using the K-Means Clustering method, resulted in two types of clustering in the form of C1, namely groups that were eligible to receive Social Assistance. in the form of Non-Cash Food Aid amounting to 13 Heads of Families and C2, namely the group that is not eligible for the selection of Social Assistance in the form of Non-Cash Food Aid totalling 97 Heads of Families. The test results on the RapidMiner Application obtained to provide the same results as manual calculations in Excel Software so that it can be used for processing type grouping and distribution in a targeted, effective and efficient manner to Beneficiary Families who are entitled to receive Social Assistance in the form of Non-Cash Food Assistance which is useful for reducing the percentage of poverty in Indonesia, especially in the Rukun Warga 001 Kelurahan Kelapa Dua.

REFERENCES

- Anggleni, A. (2018). IMPLEMENTASI KEBIJAKAN PROGRAM KARTU KELUARGA SEJAHTERA (KKS) DALAM MENINGKATKAN KESEJAHTERAAN MASYARAKATMISKIN DI KELURAHAN SEKIP JAYA KECAMATAN KEMUNING KOTA PALEMBANG. *Journal PPS UNISTI*, 1(1), 24–39. <https://doi.org/10.48093/jiask.v1i1.3>
- Bastian, A. (2018). PENERAPAN ALGORITMA K-MEANS CLUSTERING ANALYSIS PADA PENYAKIT MENULAR MANUSIA (STUDI KASUS KABUPATEN MAJALENGKA). *Sistem Informasi*, (1), 26–32.
- Khoir, D. (2019). Penerapan Metode K-Means Pengelompokan Calon Penerima Bantuan Sosial di Desa Lemberang Endratul. 6(1), 1–9.
- Marlina, D. (2018). Implementasi Algoritma K-Medoids dan K-Means untuk Pengelompokan Wilayah Sebaran Cacat pada Anak. *Jurnal CoreIT: Jurnal Hasil Penelitian Ilmu Komputer Dan Teknologi Informasi*, 4(2), 64. <https://doi.org/10.24014/coreit.v4i2.4498>
- Nugraha, Y. R. A. (2021). IMPLEMENTASI ALGORITME SUPPORT VECTOR MACHINE UNTUK KLASIFIKASI PENYESUAIAN UANG KULIAH TUNGGAL TERDAMPAK PANDEMI COVID-19 (Studi Kasus: Universitas Siliwangi). *Teknik Informatika Kaputama (JTIK)*, 5(2), 210–218.
- Probosiwi, R. (2016). Pengangguran dan Pengaruhnya terhadap Tingkat Kemiskinan Unemployment and Its Influence on Poverty Level. *Jurnal Penelitian Kesejahteraan Sosial*, 15(02), 89–100. Retrieved from file:///D:/Data Taufik/Artikel 2021/Proposal Penelitian/Artikel SMK Banten/1349-3931-1-SM (1).pdf
- Sahril, S. (2021). ANALISIS KELAYAKAN PENERIMA BANTUAN COVID-19 MENGGUNAKAN METODE K-MEANS PADA KECAMATAN SAGULUNG KOTA BATAM. *Comasie*, 05(01).
- Samsuri, D. (2018). PENGARUH SIKAP DAN MOTIVASI TERHADAP KINERJA KARYAWAN PADA MTsN TAKERAN KABUPATEN MAGETAN. *Capital: Jurnal Ekonomi Dan Manajemen*, 1(1), 49. <https://doi.org/10.25273/capital.v1i1.2137>
- Sinukun, S. M. and R. S. (2018). Prediksi Tingkat Kemiskinan Provinsi Gorontalo Menggunakan Metode Gabungan K-Means dan Generalized Regression Neural Network. *J. Energy*, 10(2), 29–34.
- Sunia. (2019). PENERAPAN DATA MINING UNTUK CLUSTERING DATA PENDUDUK MISKIN MENGGUNAKAN ALGORITMA K-MEANS Dina. *Jurnal Ilmiah Mahasiswa Teknik Informatika, Vol 1 No 2*(2016), 121–134.
- Triana, L. (2018). PEMENUHAN GIZI KELUARGA MISKIN PENERIMA BANTAUAN PANGAN NON TUNA. In *Puslit.Kemsos.Go.Id.* Retrieved from <https://puslit.kemsos.go.id/upload/post/files/09985082a75f0beda3a66388151acd84.pdf>
- Wicaksono, A. E. (2016). IMPLEMENTASI DATA MINING DALAM PENGELOMPOKAN DATA PESERTA DIDIK DI SEKOLAH UNTUK MEMPREDIKSI CALON PENERIMA BEASISWA DENGAN MENGGUNAKAN ALGORITMA K- MEANS (STUDI KASUS SMAN 16 BEKASI). *Jurnal Teknologi Rekayasa*, 21(3), 206–216. Retrieved from <http://www.sman16bekasi.sch.id>
- Widaningsih, S. (2019). PERBANDINGAN METODE DATA MINING UNTUK PREDIKSI NILAI DAN WAKTU KELULUSAN MAHASISWA PRODI TEKNIK INFORMATIKA DENGAN ALGORITMA C4.5, NAÏVE BAYES, KNN, DAN SVM. *Jurnal Tekno Insentif*, 13(1), 16–25. <https://doi.org/10.36787/jti.v13i1.78>
- Widiani, L. S. (2018). PENERAPAN MEDIA FILM SEBAGAI SUMBER BELAJAR UNTUK MENINGKATKAN KEMAMPUAN MENGOLAH INFORMASI SISWA DALAM PEMBELAJARAN

*name of corresponding author



SEJARAH. *Factum: Jurnal Sejarah Dan Pendidikan Sejarah*, 7(1), 123–132.

Yefni, D. (2019). PEMBERDAYAAN KAWASAN KUMUH MELALUI PROGRAM KOTA TANPA KUMUH (KOTAKU) KOTA PEKANBARU. *Masyarakat Madani*, 4(2), 61.
<https://doi.org/10.24014/jmm.v4i2.8414>

Zaluchu, S. E. (2020). STRATEGI PENELITIAN KUALITATIF DAN KUANTITATIF DI DALAM PENELITIAN AGAMA. *JTeologi Injili Dan Pembinaan Warga Jemaat*, 4, 28–38.

*name of corresponding author



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.