Comparative Analysis Of K Means And K Medoids Algorithms In Determining Social Assistance In Padang Sidimpuan City, North Sumatra

Fadel Muhammad Siregar [1], Ujang Juhardi[2], Muntahanah[3], Agung Kharisma Hidayah [4].

Universitas Muhammadiyah Bengkulu

Abstract: The government's efforts to tackle poverty are by issuing several programs that can help meet people's needs. However, in practice, it is not uncommon for the distribution of aid or other government programs to not be on target, because there are no clear procedures and calculations in determining which people are entitled to receive aid. Therefore, it is necessary to have a calculation mechanism that involves data on the demographic characteristics of the community using clustering. In this clustering process, there are 2 algorithms that are often used, namely K-medoid and K-means. The research aims to classify which communities are a priority for receiving assistance and which are not a priority. In order to get more accurate results, this research also tested the two clustering algorithms to get the best algorithm and the best number of clusters based on the dataset owned by looking at the Davies Bouldin Index (DBI). This research concluded that with the data set the best algorithm was K-medoids and the number of clusters was 2 with a DBI value of 1.332. Then, in the results of the clustering carried out from a total of 1031 data analyzed, it was found that 396 residents were eligible or made priority for receiving assistance and 635 residents who were not yet included in the priority list for receiving assistance.

Keywords: Assistance, Government, Clustering, K-means, K-medoids, DBI

Introduction

Economic development is a series of efforts made to develop economic activities to improve people's lives. Efforts to reduce poverty have long been carried out by the government by issuing several programs that can help meet the needs of the community, both those that are directly consumptive, in the field of education, as well as those issued by the government in the form of capital with the aim of supporting the development of Micro, Small and Medium Enterprises (MSMEs) so that the community has a business to improve the economic level of the community itself. These programs include the Smart Indonesia Program (PIP) provided through the Smart Indonesia Card (KIP), the Healthy Indonesia Program (PIS) provided through the Healthy Indonesia Card (KIS), the Family Hope Program (PKH), and other programs aimed at helping meet the needs of the poor (Pratiwi, Noorsyarifa and Apsari, 2022).

The distribution of aid or government programs at the regional level is supervised and monitored by the Social Service which is part of a government organization that is the

implementer of policies made by the government at a higher level. However, in practice it is not uncommon for the distribution of assistance or other government programs to be off target, in fact, people who should be entitled to assistance do not get it or vice versa. This happens because there are no clear procedures and calculations in determining which people are entitled to assistance, all are still based on observations and considerations that are potentially subjective in determining which people are more entitled to get and which should not be entitled. Therefore, it is necessary to have a calculation mechanism that involves data on demographic characteristics of the community such as asset ownership status, latest education, etc.

Therefore, it is necessary to group or cluster residents based on their economic level by taking into account demographic characteristics so that it can facilitate the local government in this case the Padang Sidimpuan City Social Service in selecting residents who are entitled to government assistance or programs. In response to this, there is a technique that can be used in carrying out the clustering process or grouping quickly based on mathematical calculations, namely Data Mining. Data mining is the process of finding interesting patterns or information in selected data using certain techniques or methods. The techniques, methods, or algorithms in data mining vary greatly (Yuli Mardi, 2019).

Clustering is one of the techniques in data mining, a method used to group data with the same characteristics into the same region and data with different characteristics into other regions. In this clustering process there are 2 algorithms that are often used, namely K-medoid and K-means.

Based on the background that has been described above, the researcher plans to conduct research with the title Comparative analysis of k-means and k-medoids algorithms in determining social assistance at the Padang Sidimpuan City Social Service, North Sumatra. In this study, researchers will conduct economic clustering using the two algorithms, as well as examine which of the two algorithms is the best in performance based on the Davies Bouldin Index (DBI) value.

Methodology

A. Data collection

Data collection in this study was carried out using the following methods:

a. Interview

Interviews will be conducted with one employee of the Padang Sidimpuan City Social Service, North Sumatra to get the data needed.

b. Documentation

The documentation method is carried out to obtain data and information in the form of books, archives, documents, numbers, and images in the form of reports and information that can support research. In this research, researchers will submit requests for archives or citizen data documents to the Social Service to be analyzed with data mining.

B. Research Method

KDD Process

In the Knowedge Discovery in Database (KDD) process, the stages that will be carried out in the KDD process in this research:

a. Data Section

At the data section stage, data selection is carried out that does not need to be included in the dataset so that the data is not too much and the remaining data will actually be a reference in the data mining process.

b. Data Pre-Processing

The data pre-processing stage is carried out to clean up noise data such as completing empty data or deciding to exclude empty data in the data mining process stage. Pre-processing processes include daltal cleaning to clean data from nois, data integraltion process of merging and equalizing data that may come from different sources or files, data transformation is the stage of transforming data into a format that can be calculated, for example, from non-numbers to numbers.

c. Data Mining

This stage is the core stage in this research, calculations will be carried out using the Rapidminer data mining application with the k-means and k-medoids algorithms and also do the manual version. There are several stages carried out at this stage, namely, with the data set used, it will be tested according to the two algorithms how many clusters are ideal to be formed and used as the number of k or clusters, then the clustering process will be carried out with the number of k or clusters based on the number of clusters and the best algorithm according to the previous test.

Result and Discussion

At this stage, it discusses recognizing the discussion of the data mining process, namely clustering the data that has been obtained during the research using the k-means and k-medoids methods using the RapidMiner data mining application.

Research has been conducted at the social service of Padang Sidimpuan City, North Sumatra. The data that was successfully obtained in the research or which will become a dataset in the data mining process is in the form of 1057 citizen data. The data must go through selection before calculations are carried out in accordance with the steps in the data mining process.

Figure 1 Dataset
Data Selection

At this stage, data is selected that does not need to be included in the dataset to be discarded so that there is not too much data to be processed and the remaining data is truly data that will be used as a reference in the data mining process.

In the existing dataset, data disposal was carried out on column data that was deemed unnecessary, such as employment status of the head of the family, roof quality, floor quality, etc. The reason this data was discarded was because most residents did not answer so that valid information was not obtained in this column. The following displays the data from the selection results that have been carried out:

Figure 2 Selection Data Results

Data Pre-Processing

On data sets that are already owned, this stage is done by cleaning noise, filling or cleaning empty data rows, data transformation, namely changing data from uncountable data to data that can be calculated. Meanwhile, data transformation is carried out by initiating information in the form of sentences into numbers so that all the data contained in the dataset are numbers or numeric which can be calculated using the k-means and k-medoids algorithms. There are no provisions or formulas for the data transformation process, so in this research data transformation refers to the quality or value of goods used by residents. So for the lowest value, for example the type of floor for residents of land, it will be given a value of 1 and vice versa, if the type of floor used by residents is marble or the like, it will also be given a high value. Likewise with other types. One of the references used by researchers in terms of data transformation is research from (Fakhrul Gunawan, Rakhmat Umbara and Kasyidi, 2022) with the title "Grouping the Economic Status of Tanjungsari Village Families using the K-Means Clustering Method", in his research data transformation was carried out regarding the type of house used. lived in, where he gave a score of 6 to "own", 5 to "contract/rent", 3 to "borrowed" to 1 to "other". After this process, a total of 1031 data were obtained that were ready for analysis.

Figure 3 Data Noise Cleaning Results

Conclusion

Based on the research that has been done, the conclusions of the results of this study are as follows: (a) Based on the clustering results of a total of 1,031 citizen data, it was successfully grouped into 2, namely underprivileged citizens who deserve assistance in cluster 0 as many as 396 walrgal data and citizens who are categorized as malmpu so that they have not or have not become a priority in providing assistance in cluster 1 as many as 635 datal citizens.(b) The results of calculations on Rapidminer and manually using Microsoft Excel show the same results, namely cluster 0 as many as 396 and cluster 1 as many as 635 citizen data. (c) Based on the data set obtained and used in this study, and based on the tests carried out, it is found that the k-medoids algorithm with the number of clusters 2 is still more behind than k-means. (d) The results of clustering citizen data only provide a reference to the authorities to get accurate data related to the distribution of assistance so that the assistance is right on target.:

References

Alghifari, F. and Juardi, D. (2021) 'Penerapan Data Mining Pada Penjualan Makanan Dan Minuman Menggunakan Metode Algoritma Naïve Bayes', *Jurnal Ilmiah Informatika*, 9(02), pp. 75–81. Available at: https://doi.org/10.33884/jif.v9i02.3755.

Amalina, T. *et al.* (2022) 'Metode K-Means Clustering Dalam Pengelompokan Penjualan Produk Frozen Food', *Jurnal Ilmiah Wahana Pendidikan*, 8(15), pp. 574–583. Available at: https://doi.org/10.5281/zenodo.7052276.

Arifandi, M. *et al.* (2021) 'Implementasi Algoritma K-Medoids Untuk Clustering Wilayah Terinfeksi Kasus Covid-19 Di Dki Jakarta', *JTT (Jurnal Teknologi Terapan)*, 7(2), p. 120. Available at: https://doi.org/10.31884/jtt.v7i2.353.

Candra, M.A.A. and Wulandari, I.A. (2021) 'SISTEM INFORMASI BERPRESTASI BERBASIS WEB PADA SMP NEGERI 7 KOTA METRO', *Jurnal Mahasiswa Ilmu Komputer (JMIK)*, 66(8), pp. 465–471. Available at: https://doi.org/10.51620/0869-2084-2021-66-8-465-471.

Dinata, R.K. *et al.* (2020) 'Reduksi Atribut Menggunakan Information Gain untuk Optimasi Cluster Algoritma K-Means', *Jurnal Edukasi dan Penelitian Informatika (JEPIN)*, 6(1), p. 48. Available at: https://doi.org/10.26418/jp.v6i1.37606.

Fakhrul Gunawan, F., Rakhmat Umbara, F. and Kasyidi, F. (2022) 'MIND (Multimedia Artificial Intelligent Networking Database Pengelompokan Status Ekonomi Keluarga Desa Tanjungsari menggunakan Metode K-Means Clustering', *Journal MIND Journal | ISSN*, 7(2), pp. 204–217. Available at: https://doi.org/10.26760/mindjournal.v7i2.204-217.

Hermanto, B., Yusman, M. and Nagara, N. (2019) 'Sistem Informasi Manajemen Keuangan pada PT. Hulu Balang Mandiri Menggunakan Framework Laravel', *Jurnal Komputasi*, 7(1), pp. 17–26. Available at: https://doi.org/10.23960/komputasi.v7i1.2051.

Kennardi and Budi, A. (2019) 'PENERAPAN DATA MINING UNTUK PREDIKSI NILAI TUKAR PETANI TANAMAN PANGAN DI INDONESIA DENGAN METODE LINEAR REGRESSION DAN SUPPORT VECTOR MACHINE', *Jurnal Informatika dan Bisnis PENERAPAN*, p. 32.

Manullang, N. *et al.* (2021) 'Implementasi Teknik Data Mining untuk Prediksi Peminatan Jurusan Siswa Menggunakan Algoritma C4.5', *Jurnal Ilmu Komputer dan Teknologi*, 2(2), pp. 1–5. Available at: https://doi.org/10.35960/ikomti.v2i2.700.

Miftahuddin, Y., Umaroh, S. and Karim, F.R. (2020) 'Perbandingan Metode Perhitungan Jarak Euclidean, Haversine, Dan Manhattan Dalam Penentuan Posisi Karyawan', *Jurnal Tekno Insentif*, 14(2), pp. 69–77. Available at: https://doi.org/10.36787/jti.v14i2.270.

Muningsih, E., Maryani, I. and Handayani, V.R. (2021) 'Penerapan Metode K-Means dan Optimasi Jumlah Cluster dengan Index Davies Bouldin untuk Clustering Propinsi Berdasarkan Potensi Desa', *Jurnal Sains dan Manajemen*, 9(1), p. 96. Available at: www.bps.go.id.

Noerkaisar, N. (2021) 'Efektivitas Penyaluran Bantuan Sosial Pemerintah untuk Mengatasi Dampak Covid-19 di Indonesia', *Jurnal Manajemen Perbendaharaan*, 2(1), pp. 83–104. Available at: https://doi.org/10.33105/jmp.v2i1.363.

Pratiwi, S.A., Noorsyarifa, G.C. and Apsari, N.C. (2022) 'Upaya Penanggulangan Kemiskinan Ekonomi di Indonesia Melalui Perspektif Pekerja Sosial', *Focus : Jurnal Pekerjaan Sosial*, 5(1), p. 72. Available at: https://doi.org/10.24198/focus.v5i1.39965.

Rafi Nahjan, M., Nono Heryana and Apriade Voutama (2023) 'Implementasi Rapidminer Dengan Metode Clustering K-Means Untuk Analisa Penjualan Pada Toko Oj Cell', *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(1), pp. 101–104. Available at: https://doi.org/10.36040/jati.v7i1.6094.

Rahmah, E. *et al.* (2022) 'Penerapan Algoritma K-Medoids Clustering Untuk Menentukan Srategi Promosi Pada Data Mahasiswa (Studi Kasus : Stikes Perintis Padang)', *Jurnal Nasional Komputasi dan Teknologi Informasi (JNKTI)*, 5(3), pp. 556–564. Available at: https://doi.org/10.32672/jnkti.v5i3.4355.

Rahmansyah, W. *et al.* (2020) 'Pemetaan Permasalahan Penyaluran Bantuan Sosial Untuk Penanganan Covid-19 Di Indonesia', *Jurnal Pajak dan Keuangan Negara (PKN)*, 2(1), pp. 90–102. Available at: https://doi.org/10.31092/jpkn.v2i1.995.

Sholeh, M. and Aeni, K. (2023) 'Perbandingan Evaluasi Metode Davies Bouldin, Elbow dan Silhouette pada Model Clustering dengan Menggunakan Algoritma K-Means', *STRING (Satuan Tulisan Riset dan Inovasi Teknologi)*, 8(1), p. 56. Available at: https://doi.org/10.30998/string.v8i1.16388.

Sindi, S. *et al.* (2020) 'Analisis Algoritma K-Medoids Clustering Dalam Pengelompokan Penyebaran Covid-19 Di Indonesia', *Jurnal Teknologi Informasi*, 4(1), pp. 166–173. Available at: https://doi.org/10.36294/jurti.v4i1.1296.

Sudrajat, W., Cholid, I. and Petrus, J. (2022) 'Penerapan Algoritma K-Means Clustering untuk Pengelompokan UMKM Menggunakan Rapidminer', *Jurnal JUPITER*, 14(1), pp. 27–36.

Wahyuni, D. (2021) 'Permasalahan Penyaluran Bantuan Sosial dalam Masa Pemberlakuan Pembatasan Kegiatan Masyarakat', *Info Singkat*, XIII(15/I/Puslit/Agustus/2021), pp. 13–18.

Wira, B., Budianto, A.E. and Wiguna, A.S. (2019) 'Implementasi Metode K-Medoids Clustering Untuk Mengetahui Pola Pemilihan Program Studi Mahasiwa Baru Tahun 2018 Di Universitas Kanjuruhan Malang', *RAINSTEK : Jurnal Terapan Sains & Teknologi*, 1(3), pp. 53–68. Available at: https://doi.org/10.21067/jtst.v1i3.3046.

Yuli Mardi (2019) 'Data Mining : Klasifikasi Menggunakan Algoritma C4 . 5 Data mining merupakan bagian dari tahapan proses Knowledge Discovery in Database (KDD) . Jurnal Edik Informatika', *Jurnal Edik Informatika*, 2(2), pp. 213–219.

.