

Application of the K-Means Algorithm for Clustering Social Assistance Recipients (Case Study: Binjai City Social Agency)

Meliyani^{1*}, Husnul Khair², Magdalena Simanjuntak³

^{1,2,3}STMIK Kaputama Binjai

meliyani.business@gmail.com^{1*}, Husnul.khair@gmail.com², magdalena.simanjuntak84@gmail.com³

Abstract

Social assistance distribution requires accurate data management to ensure that assistance reaches eligible beneficiaries. The Binjai City Social Agency manages the National Socio-Economic Single Data (DTSEN), which contains socio-economic information but is primarily used for administrative purposes. This study aims to apply the K-Means clustering algorithm to identify patterns among social assistance recipients based on the National Decile, Family Hope Program (PKH) recipient status, and Basic Food Assistance recipient status. The research employed the Knowledge Discovery in Databases (KDD) methodology, consisting of Data Selection, Data Preprocessing, Transformation, Data Mining, and Evaluation. A total of 847 DTSEN records were analyzed using MATLAB, while the optimal number of clusters was determined using the Elbow Method. The results indicated that the optimal number of clusters was three, representing recipient groups with different socio-economic characteristics based on National Decile and social assistance status. The clustering results provide useful information for understanding recipient characteristics and can support the Binjai City Social Agency in managing social assistance data more effectively.

Keywords: *K-Means Algorithm, Clustering, DTSEN, Social Assistance, MATLAB*

1. Introduction

Social assistance is a government program aimed at helping low-income communities meet their basic needs. In Indonesia, the National Socio-Economic Single Data (DTSEN) serves as the primary database for implementing social assistance programs. However, the data are mainly used for administrative purposes, while their potential to identify patterns among social assistance recipients has not been fully utilized. Data mining techniques, particularly clustering, can be applied to discover hidden patterns within large datasets. The K-Means algorithm is widely used because of its simplicity and effectiveness in grouping data with similar characteristics. Previous studies have shown that K-Means performs well in clustering socio-economic and social assistance data. However, the application of K-Means using DTSEN data with the variables of National Decile, Family Hope Program (PKH) recipient status, and Basic Food Assistance recipient status remains limited. Therefore, this study applies the K-Means algorithm to cluster social assistance recipients using DTSEN data from the Binjai City Social Agency. The clustering process follows the Knowledge Discovery in Databases (KDD) methodology, while the Elbow Method is used to determine the optimal number of clusters. The results are expected to provide useful information for understanding recipient characteristics and to support more effective social assistance data management.

2. Research Methods

This study employs the Knowledge Discovery in Databases (KDD) methodology to analyze social assistance recipient data from the Binjai City Social Agency. The KDD process consists of five stages: Data Selection, Data Preprocessing, Transformation, Data Mining, and Evaluation.

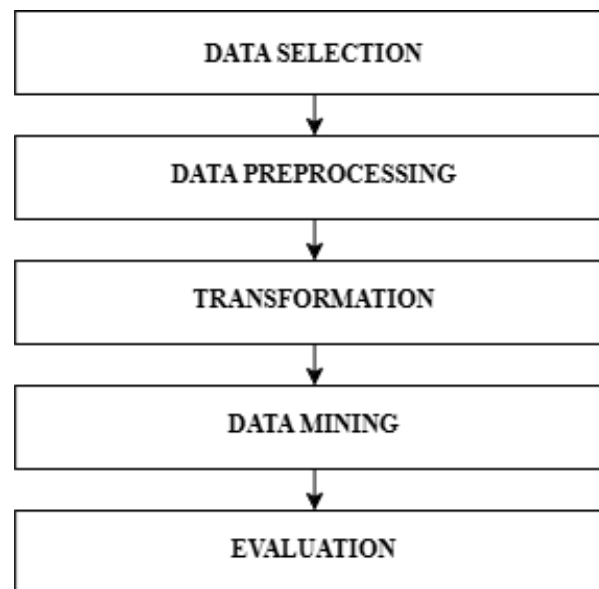


Fig.1 Research Methodology Structure

2.1. Systems Analysis

The Binjai City Social Agency manages the National Socio-Economic Single Data (DTSEN), which contains socio-economic information used for social assistance administration. However, the data have not been optimally utilized to identify patterns among social assistance recipients. Therefore, this study applies the K-Means clustering algorithm to group recipients based on the National Decile, Family Hope Program (PKH) recipient status, and Basic Food Assistance recipient status. The clustering results are expected to support more effective social assistance data management and decision-making.

2.2. Social Assistance

Social assistance is a government program aimed at helping low-income households meet their basic needs and improve their welfare. In this study, two social assistance programs are used as clustering variables, namely the Family Hope Program (PKH) and Basic Food Assistance. These variables, together with the National Decile, represent the socio-economic conditions of households and are used to identify recipient groups with similar characteristics.

2.3. Problem Solving Methods

The research problem is addressed using the K-Means clustering algorithm based on the Knowledge Discovery in Databases (KDD) methodology. The KDD process consists of Data Selection, Data Preprocessing, Transformation, Data Mining, and Evaluation, which systematically transform raw data into meaningful information.

2.3.1 Data Mining

Data mining is the process of extracting meaningful information from large datasets. In this study, data mining is applied to identify patterns among social assistance recipients using the K-Means clustering algorithm.

2.3.2 Clustering

Clustering is a data mining technique that groups data with similar characteristics into the same cluster. This study employs the K-Means algorithm because of its simplicity, computational efficiency, and effectiveness in clustering numerical data based on the minimum distance to cluster centroids.

2.3.3 MATLAB

MATLAB was used to normalize the data, determine the optimal number of clusters using the Elbow Method, perform K-Means clustering, and visualize the clustering results. The software also facilitated the analysis of cluster centroids and member distribution.

2.3.4 Research Supporting Data for 2025

The dataset used in this study was obtained from the National Socio-Economic Single Data (DTSEN) managed by the Binjai City Social Agency in 2025. A total of 847 records were analyzed using three variables: National Decile, Family Hope Program (PKH) recipient status, and Basic Food Assistance recipient status. These variables were selected because they represent the socio-economic characteristics of social assistance recipients. Table 1 presents a sample of the dataset used in this study.

Table 1. Sample Data For 2025

No	District	National Decile	Name	PKH Social Assistance Recipient Status	Sembako Social Assistance Recipient Status
1	West Binjai	2	A	Administrator (PKH Jul - Sep 2025)	Administrator (Basic Food Aid Jul - Sep 2025)
2	West Binjai	3	B	No	Family (Basic Food Aid Jul - Sep 2025)
3	West Binjai	3	C	No	Administrator (Basic Food Aid Jul - Sep 2025)
4	East Binjai	2	D	Family (PKH Jul - Sep 2025)	Family (Basic Food Aid Jul - Sep 2025)
5	Binjai City	3	E	Family (PKH Jul - Sep 2025)	Family (Basic Food Aid Jul - Sep 2025)
6	Binjai City	4	F	Administrator (PKH Jul - Sep 2025)	Administrator (Basic Food Aid Jul - Sep 2025)
7	North Binjai	2	G	No	Administrator (Basic Food Aid Jul - Sep 2025)
8	West Binjai	1	H	Administrator (PKH Jul - Sep 2025)	Administrator (Basic Food Aid Jul - Sep 2025)
9	West Binjai	4	I	Family (PKH Jul - Sep 2025)	Family (Basic Food Aid Jul - Sep 2025)
10	West Binjai	2	J	Administrator (PKH Jul - Sep 2025)	Administrator (Basic Food Aid Jul - Sep 2025)

3. Result and Discussion

3.1. Algoritma K-Means

K-Means is a non-hierarchical clustering algorithm used to partition data into several clusters based on the similarity of characteristics. The algorithm groups data objects by calculating the distance between each data point and the cluster centroid using the Euclidean distance. In this study, the K-Means algorithm is applied to group social assistance recipients based on three variables: National Decile, Family Hope Program (PKH) recipient status, and Basic Food Assistance recipient status. The distance calculation between data points and centroids is performed using the Euclidean distance equation:

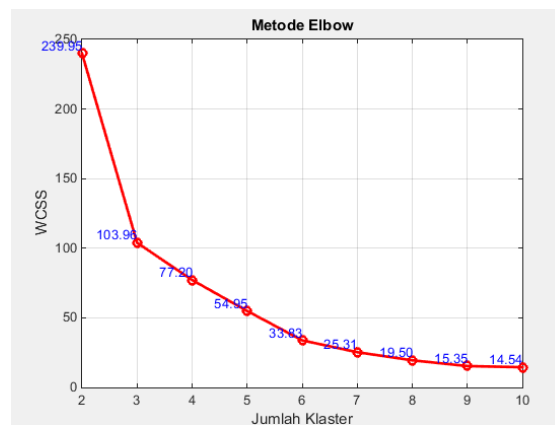
$$D_{ij} = \sqrt{(x_{1i} - x_{1j})^2 + (x_{2i} - x_{2j})^2 + \dots + (x_{ki} - x_{kj})^2}$$

The centroid value is updated by calculating the average value of all data points within each cluster. The process continues iteratively until the centroid values no longer change significantly.

$$C(i) = \frac{x_1 + x_2 + x_3 + \dots + x_n}{\sum x}$$

3.2. Interface Discussion

The interface discussion presents the implementation results of the K-Means clustering process using MATLAB. The system displays the Elbow Method graph to determine the optimal number of clusters. Based on the graph analysis, the optimal value was obtained at K = 3, indicated by the formation of an elbow point and a significant decrease in SSE values before reaching the third cluster. The MATLAB interface also displays the clustering results, centroid values, and visualization of the three clusters. These results are used to analyze the characteristics of each social assistance recipient group based on the selected variables.

**Fig 2.** Optimal Elbow Cluster Results

3.3. Program Analysis

Program analysis was conducted to evaluate the clustering process performed using MATLAB. The K-Means algorithm completed the clustering process after three iterations, resulting in three optimal clusters. The clustering results consisted of 273 recipients in Cluster 1, 370 recipients in Cluster 2, and 204 recipients in Cluster 3. Each cluster represents different characteristics of social assistance recipients based on the National Decile, PKH recipient status, and Basic Food Assistance recipient status. The cluster results were then analyzed to determine the characteristics of each group. Cluster interpretation was performed based on centroid values, where clusters with lower National Decile values and higher assistance recipient status indicate lower welfare characteristics, while higher National Decile values represent higher welfare characteristics.

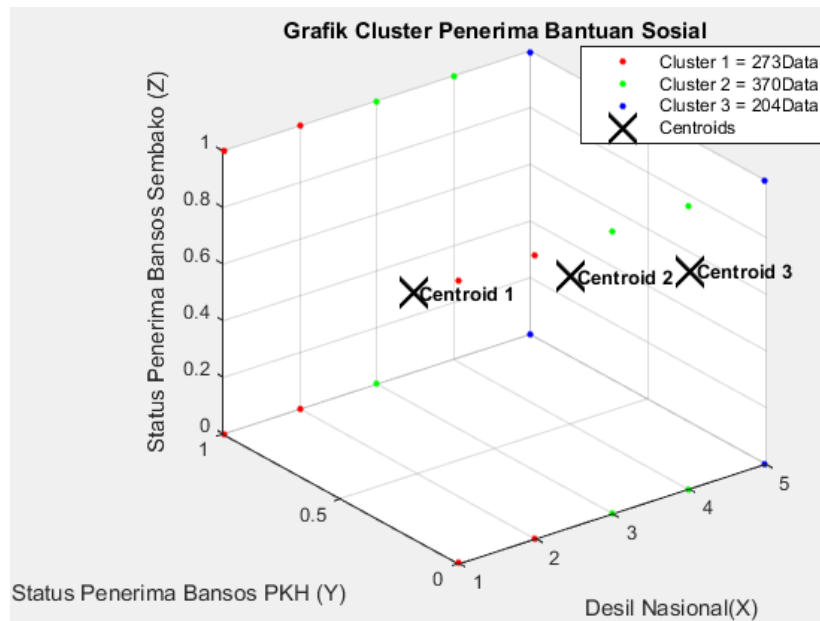


Fig 3. 3 Cluster Test Results

3.4. Cluster Interpretation

Based on the centroid analysis, Cluster 3 represents recipients with lower welfare characteristics because it has lower National Decile values and a higher proportion of PKH and Basic Food Assistance recipients. Cluster 1 represents recipients with medium welfare characteristics, while Cluster 2 represents recipients with higher welfare characteristics due to higher National Decile values and lower dependence on social assistance programs.

4. Results

1. Cluster 1 comprises 273 data points. The centroid values are 1.61 for National Decile (X), 0.39 for PKH Social Assistance Recipient Status (Y), and 0.72 for Basic Food Assistance Recipient Status (Z). These values indicate that Cluster 1 is dominated by individuals with a low national decile; a portion still receives PKH social assistance, while the majority receives basic food assistance.
2. Cluster 2 comprises 370 data points. The centroid values are 3.48 for the National Decile (X), 0.33 for PKH social assistance recipient status (Y), and 0.64 for *Sembako* social assistance recipient status (Z). These values indicate that members of Cluster 2 fall within the middle national decile; some receive PKH social assistance, while the majority receive *Sembako* social assistance.
3. Cluster 3 comprises 204 data points. The centroid values are 5.00 for the National Decile (X), 0.31 for PKH social assistance recipient status (Y), and 0.53 for *Sembako* social assistance recipient status (Z). These values indicate that Cluster 3 is dominated by individuals with a higher national decile compared to other clusters; a small proportion receive PKH social assistance, while approximately half of the cluster members receive *Sembako* social assistance.

5. Conclusion

This study successfully applied the K-Means algorithm to cluster social assistance recipients using DTSEN data from the Binjai City Social Agency. Based on the Elbow Method, the optimal number of clusters was three. The resulting clusters represent different socio-economic characteristics based on National Decile, PKH recipient status, and Basic Food Assistance recipient status. The clustering results provide useful information to support social assistance data management and decision-making.

References

- [1] Aini, Rini Nurul. 2025. "Digitalisasi Program PKH: Studi DTSN Di Kecamatan Sinjai Utara." 10(2). doi:10.14710/jiip.v10i2.28816.
- [2] Fadhl, Khotim, & Laila Rohmatun Nazila. 2023. "Pengaruh Implementasi Bantuan Sosial Bpnt Dan Pkh Terhadap Efektivitas Penanggulangan Kemiskinan." *Jurnal Education and Development* 11(2): 196–202. doi:10.37081/ed.v11i2.4654.
- [3] Fatwa, Muhamad, Ristu Rizki, Pandiangan Sriwinarty, & Edi Supriyadi. 2022. "Pengaplikasian Matlab Pada Perhitungan Matriks." *Papanda Journal of Mathematics and Science Research* 1(2): 81–93. doi:10.56916/pjmsr.v1i2.260.
- [4] Jannata, S. 2021. "Aplikasi Pengarsipan Surat Masuk Dan Surat Keluar Pada Dinas Penanaman Modal Dan Pelayanan Terpadu Satu Pintu Kota Samarinda Berbasis Web." <https://repository.wicida.ac.id/4025/>.
- [5] Jelita, Tiara, Relita Buaton, & Magdalena Simajuntak. 2023. "Pengelompokan Bidang Usaha Terhadap Bantuan Produktif Usaha Mikro (BPUM) Berdasarkan Wilayah Deli Serdang Menggunakan Metode Clustering K-Means (Studi Kasus: Dinas Koperasi Dan UMKM Kabupaten Deli Serdang)." *Journal of Computer Science and Information Technology E-ISSN* 3(2): 50.
- [6] Lentiana Dakhi, Yustina, & Besse Arnawisuda Ningsi. 2024. "Pengelompokan Kabupaten Dan Kota Provinsi Sumatera Utara Berdasarkan Indikator Kesejahteraan Rakyat Menggunakan Algoritma K-Means." 4(3): 993–1003.
- [7] Oktarian, Wawan, & Siti Alisa. 2022. "Kesejahteraan Masyarakat Di Desa Campakasari Kecamatan Campaka Kabupaten Purwakarta." 3(1): 1–14.
- [8] Putri, Wildani, & Muhammad Afdal. 2023. "Penerapan Algoritma K-Means Untuk Pengelompokan Data Penyandang Disabilitas Di Kabupaten Rokan Hilir." *Indonesian Journal of Informatic Research and Software Engineering (IJIRSE)* 3(1): 30–38. doi:10.57152/ijirse.v3i1.526.
- [9] Rady Putra, Lalu Ganda, & Anthony Anggrawan. 2021. "Pengelompokan Penerima Bantuan Sosial Masyarakat Dengan Metode K-Means." *MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer* 21(1): 205–14. doi:10.30812/matrik.v21i1.1554.
- [10] Rahmah, Siti, Ramona Handayani, & Sugianor. 2024. "Efektivitas Program Keluarga Harapan (PKH) Di Desa Hakurung Kecamatan Daha Utara Kabupaten Hulu Sungai Selatan." *Jurnal Kebijakan Publik* 1(4): 962–72.
- [11] Ramadani, & R. Sally Marisa Sihombing. 2024. "Evaluasi Program Bantuan Pangan Non Tunai (BPNT) Pada Masyarakat Kurang Mampu Di Desa Tebing Tanjung Kabupaten Langkat." *Jurnal Ilmiah Ilmu Administrasi Negara* 11(1): 164–74.
- [12] Riani, Adella Putri, Apriade Voutama, & Taufik Ridwan. 2023. "Penerapan K-Means Clustering Dalam Pengelompokan Hasil Belajar Peserta Didik Dengan Metode Elbow." *Jurnal Teknologi Sistem Informasi dan Sistem Komputer TGD* 6(1): 164–72.
- [13] Rika Sari, & Muhammad Yasin. 2025. "Penerapan Data Mining Untuk Clustering Kondisi Sosial Ekonomi Berdasarkan Kepemilikan Jaminan Kesehatan Menggunakan Algoritma K-Means." *Jurnal Teknik Informatika dan Teknologi Informasi* 5(3): 44–55. doi:10.55606/jutiti.v5i3.6082.
- [14] Rindi Asti Ananda, Yani Maulita, & Husnul Khair. 2024. "Clustering Menggunakan Algoritma K-Means Untuk Mengelompokkan Data Perjudian Berdasarkan Wilayah Di Kota Binjai (Studi Kasus : Pengadilan Negeri Binjai)." *Switch : Jurnal Sains dan Teknologi Informasi* 2(4): 167–80. doi:10.62951/switch.v2i4.226.
- [15] Salsabila, Nanda, Nilna Muna, Vikko Hilmi Pradana, & Wirawan Firman Nurcahya. 2024. "Analisis Efektivitas Bantuan Sosial (Bansos) Dalam Mengatasi Kemiskinan Di Indonesia." *Journal of Macroeconomics and Social Development* 1(4): 1–13. doi:10.47134/jmsd.v1i4.317.
- [16] Septiani, Nuraeni, & Ruli Herdiana. 2023. "Penerapan Algoritma K-Means Clustering Untuk Harga Rumah Di Jakarta Selatan." *Trending: Jurnal Ekonomi, Akuntansi dan Manajemen* 1(2): 35–47.
- [17] Sulistiyawati, Ari, & Eko Supriyanto. 2021. "Implementasi Algoritma K-Means Clustering Dalam Penentuan Siswa Kelas Unggulan." *Jurnal Tekno Kompak* 15(2): 25. doi:10.33365/jtk.v15i2.1162.