

Prediksi Kelayakan Penerima Bantuan Sosial Menggunakan Metode K-Prototype Pada Desa Air Hitam

Saul Michael Marthin Silitonga¹, Humuntal Rumapea², Indra Kelana Jaya³
^{1,2,3}Fakultas Ilmu Komputer, Universitas Methodist Indonesia

Info Artikel	ABSTRAK
Histori Artikel: Received, Des 9 , 2026 Revised, Jan 20, 2026 Accepted, Feb 15, 2026	Penentuan kelayakan penerima bantuan sosial merupakan permasalahan yang sering dihadapi oleh pemerintah desa, khususnya dalam memastikan bantuan tepat sasaran. Proses penilaian yang masih dilakukan secara manual seringkali menimbulkan subjektivitas, ketidaktepatan, serta ketidakkonsistenan dalam pengambilan keputusan. Oleh karena itu, diperlukan suatu metode yang mampu mengolah data secara objektif dan sistematis. Penelitian ini bertujuan untuk memprediksi kelayakan penerima bantuan sosial pada masyarakat Desa Air Hitam, Kecamatan Gebang, Kabupaten Langkat, Sumatera Utara dengan menggunakan metode K-Prototypes. Metode ini dipilih karena mampu mengolah data campuran yang terdiri dari data numerik dan data kategorikal secara bersamaan. Data yang digunakan dalam penelitian ini meliputi beberapa atribut, yaitu jumlah pendapatan, jumlah kendaraan roda dua, jumlah kendaraan roda empat, jumlah ternak ayam, jumlah ternak kambing, jumlah ternak sapi, serta kepemilikan rumah. Tahapan penelitian meliputi proses pengumpulan data, preprocessing data, transformasi dan encoding data kategorikal, penerapan algoritma K-Prototypes untuk proses clustering, serta analisis hasil pengelompokan. Hasil dari penelitian ini menunjukkan bahwa data masyarakat dapat dikelompokkan menjadi dua cluster utama, yaitu kelompok masyarakat yang layak menerima bantuan sosial dan kelompok masyarakat yang tidak layak menerima bantuan sosial. Selain itu, sistem yang dibangun berbasis aplikasi Streamlit mampu mempermudah proses pengolahan data dan visualisasi hasil clustering secara lebih sistematis dan mudah dipahami oleh pengguna. Dengan demikian, penerapan metode K-Prototypes dalam penelitian ini dapat menjadi solusi alternatif yang efektif dalam membantu pemerintah desa dalam menentukan kelayakan penerima bantuan sosial secara lebih objektif, akurat, dan tepat sasaran.
Keywords: Data Mining, Clustering, K-Prototype, Bantuan Sosial, Streamlit.	<i>This is an open access article under the CC BY-SA license.</i>



Saul Michael Marthin Silitonga,
Fakultas Ilmu Komputer, Universitas Methodist Indonesia, Medan,
Jl. Hang Tuah No.8, Medan - Sumatera Utara.
Email: saulsilitonga3@gmail.com

1. PENDAHULUAN

Bantuan sosial merupakan salah satu instrumen penting pemerintah dalam upaya meningkatkan kesejahteraan masyarakat dan mengurangi tingkat kemiskinan. Ketepatan dalam penentuan penerima bantuan menjadi faktor krusial agar program yang dijalankan dapat memberikan dampak yang optimal. Namun, dalam praktiknya, proses seleksi penerima bantuan sosial masih sering dilakukan secara manual dan subjektif, sehingga berpotensi menimbulkan ketidaktepatan

sasaran serta ketidakadilan dalam distribusi bantuan[1].

Permasalahan tersebut juga ditemukan pada tingkat desa, termasuk di Desa Air Hitam, Kecamatan Gebang, Kabupaten Langkat. Banyaknya data masyarakat dengan berbagai atribut seperti pendapatan, kepemilikan aset, dan jumlah ternak menyebabkan proses analisis menjadi kompleks apabila tidak didukung oleh sistem yang terkomputerisasi. Oleh karena itu, diperlukan pendekatan berbasis data yang mampu mengelompokkan masyarakat secara objektif berdasarkan karakteristik yang dimiliki[2].

Salah satu pendekatan yang dapat digunakan adalah teknik data mining dengan metode clustering. Metode K-Prototypes merupakan algoritma clustering yang dirancang untuk mengolah data campuran, yaitu data numerik dan kategorikal secara bersamaan. Hal ini menjadikan K-Prototypes relevan untuk diterapkan dalam kasus penentuan kelayakan penerima bantuan sosial yang memiliki karakteristik data heterogen.

Penelitian ini bertujuan untuk menerapkan metode K-Prototypes dalam mengelompokkan data masyarakat ke dalam beberapa cluster berdasarkan tingkat kelayakan menerima bantuan sosial. Hasil pengelompokan diharapkan dapat menjadi dasar dalam pengambilan keputusan yang lebih objektif, akurat, dan transparan. Selain itu, penelitian ini juga mengembangkan sistem berbasis aplikasi untuk mempermudah proses analisis data serta visualisasi hasil clustering. Dengan demikian, penelitian ini diharapkan dapat memberikan kontribusi dalam meningkatkan efektivitas penyaluran bantuan sosial melalui pemanfaatan teknologi data mining[3]

2. METODE PENELITIAN

Kerangka penelitian ini diawali dengan identifikasi masalah, yaitudanya ketidaktepatan sasaran, subjektivitas penilaian, dan keterbatasan metode manual dalam penentuan penerima bantuan sosial di Desa Air Hitam. Permasalahan ini kemudian melahirkan kebutuhan akan pendekatan berbasis data dan analitik untuk meningkatkan objektivitas dan akurasi pengambilan keputusan.

Kerangka penelitian ini dapat dilihat pada Gambar 1. sebagai panduan alur penelitian secara keseluruhan.



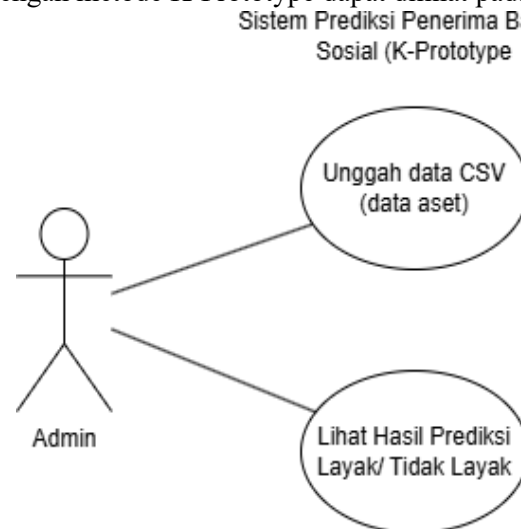
Gambar 1. Kerangka Penelitian

2.1. Desain Sistem

Perancangan sistem merupakan tahap untuk menyusun alur dan mekanisme sistem yang akan digunakan dalam penelitian[4].

1. Use Case Diagram

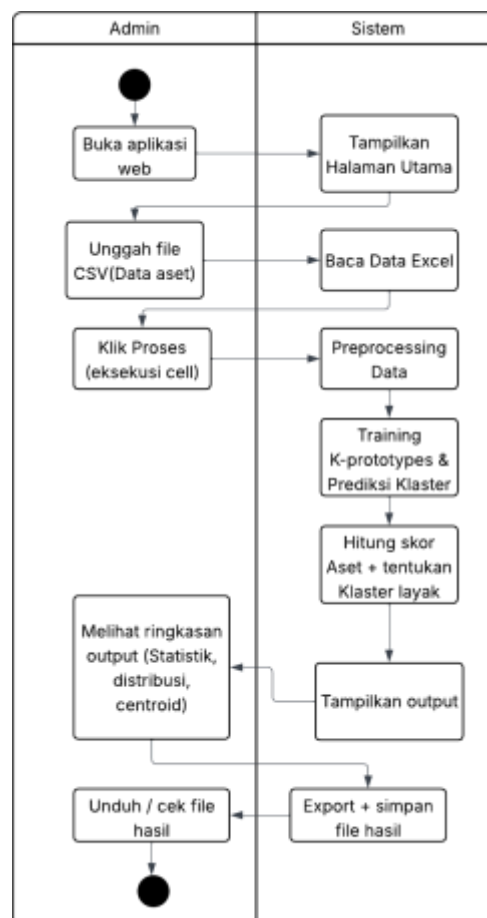
Use Case Diagram menggambarkan alur kerja atau aktivitas sistem dari sudut pandang admin. Rancangan *use case* sistem dengan metode K-Prototype dapat dilihat pada Gambar 2.



Gambar 2. Use Case Diagram

2. Activity Diagram

Activity Diagram menggambarkan alur kerja atau aktivitas sistem secara dinamis. Rancangan *activity diagram* sistem ini dapat dilihat pada Gambar 3.



Gambar 3. Activity Diagram

2.2. K-Prototype

K-Prototype adalah metode *clustering* (pengelompokan) yang digunakan untuk data campuran, yaitu data yang memiliki atribut numerik dan atribut kategorikal. Metode ini merupakan gabungan dari K-Means (untuk data numerik) dan K-Modes (untuk data kategorikal)[3], [5]. Pada penelitian ini K-prototype digunakan untuk mengelompokkan warga berdasarkan kepemilikan rumah, jumlah kendaraan, dan jumlah ternak. Hasil pengelompokan ini kemudian digunakan untuk menentukan kelompok warga yang layak atau tidak layak menerima bantuan sosial[6]. Untuk mengelompokkan data bertipe campuran yang terdiri dari atribut numerik dan kategorikal, penelitian ini menggunakan algoritma k-prototype yang bekerja dengan meminimalkan fungsi objektif berupa total ketidaksamaan antara data dan pusat kluster. Fungsi ini menemukan pembagian kluster yang menghasilkan jarak total terkecil, sehingga data dalam satu kluster memiliki karakteristik yang paling mirip[7].

Algoritma K-Prototypes menggabungkan means (rata-rata) dari K-Means dan modes (modus) dari K-Modes untuk membuat titik pusat hibrida dengan nama “prototype”[6]. Algoritma ini membuat formula/rumus Dissimilarity Coefficient dan Cost Function yang dapat digunakan pada data mixed-type. Dissimilarity Coefficient untuk data mixed-type dibagi menjadi dua perhitungan terpisah. Bagian kategori menggunakan Hamming distance dan bagian numerik menggunakan Euclidean distance.

Beberapa tahapan dalam algoritma k-Prototype adalah sebagai berikut[7]

- Menentukan banyak kelompok (k) yang akan dibentuk.
- Inisialisasi awal prototype sebagai pusat kelompok sejumlah k sesuai dengan jumlah kelompok yang dibuat.
- Melakukan perhitungan jarak dengan ukuran jarak campuran tipe data terhadap pusat kelompok yang telah ditentukan.
- Alokasi objek ke dalam kelompok yang memiliki nilai jarak terdekat.
- Melakukan perhitungan nilai centroid baru menggunakan nilai rata-rata untuk variabel numerik nilai modus untuk variabel kategorik.
- Realokasi objek ke dalam masing-masing kelompok berdasarkan perhitungan jarak terdekat dengan nilai pusat kelompok yang baru.
- Tahap c diulang sampai iterasi maksimum tercapai atau tidak ada lagi perubahan objek dalam kelompok.

Algoritma k-Prototype memerlukan metode validasi agar jumlah kelompok yang terbentuk optimal. Salah satu metode validasi pada analisis cluster yaitu Average Silhouette Width. Nilai Average Silhouette Width ada di rentang -1 sampai +1. Jika nilai Average Silhouette Width mendekati +1, maka jumlah kelompok yang terbentuk semakin optimal. Persamaan Average Silhouette Width dituliskan seperti pada Persamaan (7)

$$S(i) = \frac{1}{n} \sum_{i=1}^n \frac{b(x_i) - a(x_i)}{\max(a(x_i); b(x_i))} \quad (7)$$

$$a(x_i) = \frac{1}{n_y - 1} \sum_{jy=1}^{n_y} d(x_{iy}, x_{jy}) \text{ dan } b(x_i) = \min_{y \neq z} \frac{1}{n_z} \sum_{jz=1}^{n_z} d(x_{iy}, x_{jz})$$

Dengan

$d(x_{iy}, x_{jy})$ = jarak objek ke- i pada kelompok y dengan objek ke- j pada kelompok y

$d(x_{iy}, x_{jz})$ = jarak objek ke- i pada kelompok y dengan objek ke- j pada kelompok z

2.3. Menentukan Cluster

Menentukan *cluster* dilakukan untuk memilih jumlah kelompok terbaik berdasarkan karakteristik data menggunakan metode Elbow dan Silhouette Score sebelum proses clustering dengan K-Prototypes dilakukan. Pada tahap ini, data yang telah melalui proses preprocessing dianalisis untuk mengetahui jumlah *cluster* yang paling sesuai dan mampu menggambarkan perbedaan kondisi sosial ekonomi masyarakat[8]. Penentuan *cluster* dilakukan menggunakan metode Elbow dan Silhouette Score, di mana metode Elbow digunakan untuk melihat titik terbaik saat penambahan jumlah kluster tidak lagi memberikan perubahan signifikan, sedangkan Silhouette Score digunakan untuk mengukur kualitas pemisahan antar kluster.

Silhouette score adalah salah satu metrik untuk mengukur performa dari algoritma algoritma clustering. Silhouette score bekerja dengan cara mengukur rata-rata jarak dari setiap titik data satu cluster pada setiap titik data cluster lain (intercluster) yang terdekat (b_i). Hasil ini akan dikurangi dengan rata-rata jarak dari setiap titik data satu cluster pada setiap titik data cluster itu sendiri (intracluster) (a_i). Setelah itu hasil tersebut akan dibagi dengan nilai tertinggi antara a_i dengan b_i .

Berikut adalah rumus Silhouette score

$$s_i = \frac{b_i - a_i}{\max(b_i, a_i)}$$

Jumlah kelompok yang dipilih merupakan jumlah kluster optimal yang akan digunakan pada proses pengelompokan menggunakan metode K-Prototypes.

2.4. Prediksi

Menentukan hasil prediksi dilakukan dengan menganalisis kluster yang terbentuk dan menetapkan status kelayakan bantuan sosial bagi setiap calon penerima berdasarkan hasil *clustering*. Pada tahap ini, setiap kluster yang terbentuk dianalisis berdasarkan karakteristik dan rata-rata skor aset yang dimilikinya. *Cluster* dengan nilai rata-rata skor aset paling rendah ditetapkan sebagai *cluster* yang paling layak menerima bantuan sosial, sedangkan kluster dengan nilai yang lebih tinggi dikategorikan sebagai tidak layak[9]. Berdasarkan kluster tempat data tersebut tergabung, setiap calon penerima bantuan sosial kemudian diberikan status kelayakan, yaitu layak atau tidak layak menerima bantuan sosial. Hasil prediksi ini digunakan sebagai dasar rekomendasi bagi pihak desa dalam proses penentuan penerima bantuan sosial.

2.5 Data Mining

Data mining adalah proses analitis yang dirancang untuk mengeksplorasi sejumlah besar data untuk pengetahuan yang berharga, konsisten, dan tersembunyi. Salah satu metode yang digunakan dalam data mining adalah klasifikasi[8]. Data mining merupakan sebuah disiplin ilmu yang mempelajari metode untuk mengekstraksi informasi atau menemukan pola yang menarik dari data dalam jumlah besar. Data mining adalah operasi pencarian pola atau hubungan dalam sekumpulan data berukuran besar yang membantu pengguna untuk mengubah data mentah menjadi pengetahuan yang berguna. Dalam praktiknya, teknik ini melibatkan integrasi dari berbagai bidang ilmu seperti statistika, kecerdasan buatan, dan sistem basis data.

3. HASIL DAN PEMBAHASAN

Penelitian ini menerapkan metode K-Prototypes untuk mengelompokkan data masyarakat berdasarkan karakteristik sosial ekonomi yang dimiliki. Dataset yang digunakan terdiri dari atribut numerik dan kategorikal, yaitu jumlah pendapatan, jumlah kendaraan roda dua, jumlah kendaraan roda empat, jumlah ternak ayam, jumlah ternak kambing, jumlah ternak sapi, serta kepemilikan rumah. Sebelum proses clustering dilakukan, data terlebih dahulu melalui tahap *preprocessing* yang meliputi pembersihan data, transformasi, serta encoding pada variabel kategorikal.

3.1. Implementasi Sistem

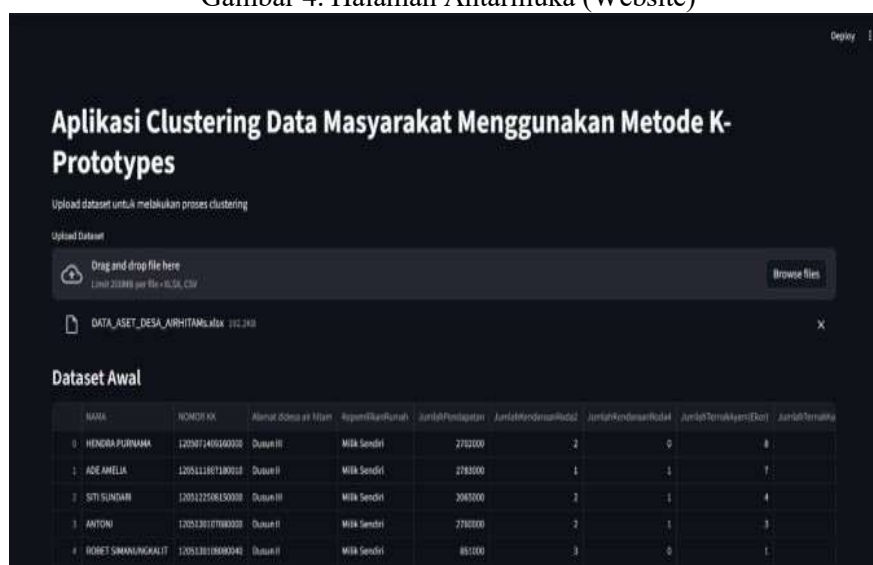
Implementasi sistem merupakan tahap penerapan rancangan sistem yang telah dibuat pada tahap perancangan yang dapat digunakan untuk melakukan proses pengolahan data dan *clustering* secara otomatis.

1. Halaman Utama Sistem

Halaman utama merupakan tampilan awal aplikasi ketika sistem dijalankan pada halaman ini pengguna dapat melihat judul aplikasi serta fitur untuk mengunggah dataset yang akan digunakan dalam proses *clustering*. Tampilan Halaman utama dapat dilihat pada Gambar 4 dan Gambar 5.



Gambar 4. Halaman Antarmuka (Website)

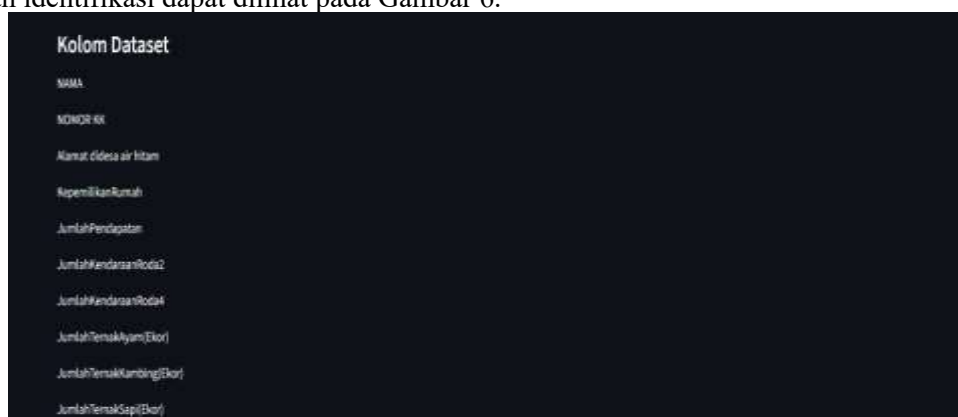


Gambar 5. Tampilan Setelah digunakan

Berdasarkan gambar 5 sistem menampilkan hasil berupa daftar calon penerima bantuan sosial yang dihasilkan dari file excel.

2. Identifikasi Atribut Data

Sistem ini melakukan identifikasi atribut data yang digunakan dalam proses klastering. Tampilan identifikasi dapat dilihat pada Gambar 6.



Gambar 6. Identifikasi Atribut Data

Penentuan atribut ini bertujuan agar algoritma K-Prototype dapat memproses data numerik dan kategorikal secara bersamaan.

3. Hasil Clustering

Hasil *clustering* kemudian ditambahkan kedalam dataset sebagai kolom baru dengan nama *cluster*. Kolom ini menunjukkan kelompok data yang dihasilkan oleh algoritma K-Prototype. Hasil *cluster* dapat dilihat pada Gambar 7 dan Gambar 8.

Kategori Benda di Rumah	Keperluan Jumlah	JumlahPendapatan	JumlahKendaraanRoda2	JumlahKendaraanRoda4	JumlahTernakKamflKor	JumlahTernakKamflKor	JumlahTernakKamflKor	JumlahTernakKamflKor	Cluster
0. Jajanan II	Milik Sendiri	2702000	2	0	0	0	2	1	1
1. Jajanan II	Milik Sendiri	2783000	1	1	1	0	1	1	1
2. Jajanan II	Milik Sendiri	2065000	2	1	4	0	3	1	1
3. Jajanan II	Milik Sendiri	2740000	2	1	3	0	5	1	1
4. Jajanan II	Milik Sendiri	851000	3	0	1	0	4	0	0
5. Jajanan V	Milik Sendiri	1541000	1	1	1	0	0	0	0
6. Jajanan III	Milik Sendiri	2046000	2	1	2	0	1	1	1
7. Jajanan VII	Milik Sendiri	2387000	2	1	7	0	2	1	1
8. Jajanan VI	Milik Sendiri	2260000	1	1	8	0	1	1	1
9. Jajanan X	Milik Sendiri	4354000	1	1	3	0	1	0	0

Gambar 7 Dataset setelah Clustering

Cluster	count
1	154
0	146

Gambar 8 Jumlah Data Tiap Cluster

Gambar 7 digunakan untuk menganalisis karakter masing-masing *cluster* sehingga dapat digunakan sebagai dasar dalam menentukan kelompok masyarakat yang berpotensi layak menerima bantuan sosial. Gambar 8 dan digunakan untuk melihat jumlah anggota data pada setiap *cluster* setelah proses pengelompokan dilakukan.

3.2. Pengujian Metode K-Prototype

Metode ini diterapkan untuk mengelompokkan data masyarakat berdasarkan beberapa variabel seperti pendapatan perbulan, jumlah kendaraan, jumlah ternak, sebagai data numerik dan status kepemilikan rumah sebagai data kategorikal.

1. Clustering Metode K-Prototype

Algoritma K-Prototype digunakan karena dataset penelitian ini memiliki kombinasi atribut numerik dan kategorikal. K-Prototype merupakan pengembangan dari metode k-means dan k-modes sehingga mampu menangani kedua jenis data tersebut. Cluster dapat dilihat pada tabel 1.

Tabel 1. Cluster Penerima bansos

Cluster	keterangan
Cluster 0	Layak Bansos
Cluster 1	Tidak Layak Bansos

2. Centroid Awal

Centroid awal dapat dilihat pada Tabel 2.

Tabel 2. Centroid Awal

Cluster	Pendapatan	Roda 2	Roda 4	Ayam	Kambing	Sapi
1	2.399.709	2.02	0.50	5.02	2.51	2.47
0	1.123.0171	1.99	0.49	5.02	2.50	2.52

Cluster 0 menunjukkan masyarakat layak menerima bantuan sosial, sedangkan *cluster* 1 menunjukkan masyarakat yang tidak layak menerima bantuan sosial.

3. Perhitungan K-Prototype Data 1 (data ke-1).

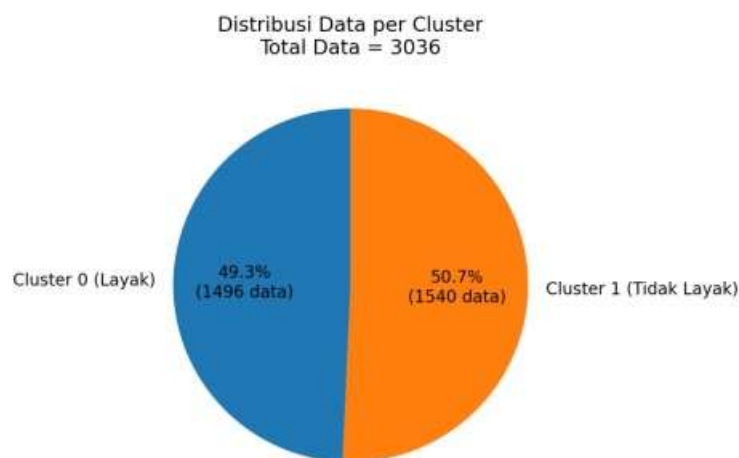
Jarak data ke *cluster* 1 :

$$\begin{aligned}
 & (2.702.000 - 2.399.702)^2 + (2 - 2,02)^2 + (0 - 0,50)^2 + (8 - 5,02)^2 + (2 - 2,51)^2 + (1 - 2,47)^2 \\
 & = (302.298)^2 + (-0,02)^2 + (-0,5)^2 + (2,98)^2 + (-0,51)^2 + (-1,47)^2 \\
 & = 91.384.080.804 + 0,0004 + 0,25 + 8,8804 + 0,2601 + 2,1609 = 91.384.080.815,5518
 \end{aligned}$$

Jarak data ke *cluster* 0 :

$$\begin{aligned}
 & (2.702.000 - 1.123.017)^2 + (2 - 1,99)^2 + (0 - 0,49)^2 + (8 - 5,02)^2 + (2 - 2,50)^2 + (1 - 2,52)^2 \\
 & = (1.578.983)^2 + (0,01)^2 + (-0,51)^2 + (2,98)^2 + (-0,5)^2 + (-1,52)^2 \\
 & = 2.493.092.576.209 + 0,0001 + 0,2601 + 8,8804 + 0,25 + 2,3014 \\
 & = 2.493.092.576.220,701
 \end{aligned}$$

Penentuan klaster, karena : $91.384.080.815,5518 < 2.493.092.576.220,701$ Maka data masuk *cluster* 1 yaitu tidak layak menerima bantuan sosial.



Gambar 9. Grafik Distribusi

Gambar 9 digunakan untuk melihat jumlah anggota data pada setiap *cluster* setelah proses pengelompokan dilakukan, dari grafik jumlah data masyarakat yang masuk kelompok 0 adalah 1496 data (49,3%) kelompok yang layak mendapatkan bantuan sosial dan masyarakat yang masuk kelompok 1 adalah 1540 data (50,7%) kelompok yang tidak layak.

4. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa metode K-Prototypes mampu digunakan secara efektif untuk mengelompokkan data masyarakat berdasarkan karakteristik sosial ekonomi yang terdiri dari data numerik dan kategorikal. Penerapan metode ini menghasilkan dua *cluster* utama yang merepresentasikan kelompok masyarakat yang layak dan tidak layak menerima bantuan sosial. Hasil clustering menunjukkan bahwa masyarakat dengan tingkat pendapatan rendah serta kepemilikan aset yang terbatas cenderung masuk ke dalam kategori layak menerima bantuan sosial, sedangkan masyarakat dengan kondisi ekonomi yang lebih baik tergolong tidak layak menerima bantuan. Dengan demikian, metode K-Prototypes dapat membantu dalam menghasilkan pengelompokan data yang lebih objektif dan konsisten dibandingkan dengan metode penilaian manual.

REFERENSI

- [1] F. M. Sarimole and L. Hakim, "Klasifikasi Barang Menggunakan Metode Clustering K-Means Dalam Penentuan Prediksi Stok Barang," *J. Sains dan Teknol.*, vol. 5, no. 3, pp. 846–854, 2024, doi: 10.55338/saintek.v5i3.2709.
- [2] Tahyani, A. S. Sunge, and M. Wangsadanureja, "Penerapan Data Mining Untuk Mempermudah Produksi Diapers Dengan Menggunakan Algoritma Regresi Linier," *ISSN 2962-3545 Pros. SAINTEK Sains dan Teknol. Vol.1 No.1 Tahun 2022 Call Pap. dan Semin. Nas. Sains dan Teknol. Ke-1 2022 Fak. Tek. Univ. Pelita Bangsa, Juli 2022*, vol. 1, no. 1, pp. 176–179, 2022, [Online]. Available: <https://www.jurnal.pelitaBangsa.ac.id/index.php/SAINTEK/article/view/1165/757>
- [3] H. G. Simanullang, A. P. Silalahi, and G. P. Pangaribuan, "PENGELOMPOKAN KOMENTAR PENGGUNA JKN MOBILE MENGGUNAKAN METODE K-MEDOIDS," 2025.
- [4] H. G. Simanullang, A. P. Silalahi, and D. R. Manalu, "Sistem Informasi Pendaftaran Mahasiswa Baru Menggunakan Framework Codeigniter dan Application Programming Interface," *Ultim. InfoSys J. Ilmu Sist. Inf.*, vol. 12, no. 1, pp. 67–73, 2021, doi: 10.31937/si.v12i1.1803.
- [5] T. R. Baruara, H. G. Manullang, and A. P. Silalahi, "ANALISA ALGORITMA K-MEDOIDS DALAM," vol. 4, no. 2, pp. 1–10, 2024.
- [6] P. Ariani Fitri, Mustafid, and P. Kartikasari, "PENERAPAN ALGORITMA k-PROTOTYPE UNTUK PENGELOMPOKAN DESA DI KABUPATEN BEKASI BERDASARKAN INFRASTRUKTUR DIGITAL," *J. GAUSSIAN*, vol. 13, pp. 479–489, 2025, doi: 10.14710/j.gauss.13.2.479-489.
- [7] N. Saniyah, N. Suarna, and W. Prihartono, "CLUSTERING PELANGGARAN LALU LINTAS PADA KENDARAAN BERMOTOR MENGGUNAKAN ALGORITMA K-PROTOTYPE," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 8, no. 1, pp. 772–778, 2024.
- [8] M. Müller, L. Longard, and J. Metternich, "Comparison of preprocessing approaches for text data in digital shop floor management systems," *Procedia CIRP*, vol. 107, no. March, pp. 179–184, 2022, doi: 10.1016/j.procir.2022.04.030.
- [9] H. G. Simanullang, A. P. Silalahi, and D. Sartika, "PREDIKSI JUMLAH PASIEN COVID-19 DI INDONESIA MENGGUNAKAN LEAST SQUARE METHOD BERBASIS ANDROID," *INFORMATIKA*, vol. 14, no. 1, pp. 86–93, 2022.