
Optimalisasi Manajemen Bantuan Sosial Dengan Implementasi Data Mining Menggunakan Algoritma K-Means Clustering

Melan Astriani Mburu Hamu¹⁾, Alfrian Carmen Talakua²⁾

¹Prodi Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Kristen Wira Wacana Sumba

Artikel Info

Genesis Artikel :

Diterima, 15 April 2024

Direvisi, 15 Mei 2024

Diterbitkan, 30 Juni 2024

Kata Kunci :

Manajemen Bantuan Sosial

Data Mining

K-Means Clustering

ABSTRAK

Penelitian ini mengkaji optimalisasi manajemen bantuan sosial di Indonesia menggunakan teknik data mining, khususnya algoritma k-means clustering. Bantuan sosial merupakan instrumen penting dalam mengurangi kemiskinan dan ketimpangan, namun pelaksanaannya sering menghadapi tantangan administratif dan logistik. Dengan mengelompokkan penerima bantuan berdasarkan data demografis dan kondisi ekonomi mereka, penelitian ini bertujuan untuk meningkatkan efisiensi dan efektivitas distribusi bantuan. Dataset dari Kelurahan Prailiu yang berisi 350 data penerima bantuan dianalisis menggunakan perangkat lunak RapidMiner. Hasil clustering menunjukkan bahwa pembagian data menjadi 4 cluster memberikan hasil terbaik dengan nilai Davies Bouldin sebesar -1.085, menunjukkan jarak antar item yang relatif dekat. Implementasi teknologi data mining memungkinkan pemerintah untuk mengidentifikasi kebutuhan spesifik kelompok penerima, mengurangi risiko pemborosan dan penyalahgunaan dana, serta meningkatkan dampak positif program bantuan sosial. Saran diberikan untuk terus mengembangkan sistem pengelolaan data dan melatih pengelola bantuan sosial dalam penggunaan teknologi ini untuk memastikan keberlanjutan dan peningkatan efektivitas program bantuan sosial.

ABSTRACT

This study examines the optimization of social aid management in Indonesia using data mining techniques, specifically the k-means clustering algorithm. Social aid is a crucial instrument in reducing poverty and inequality, but its implementation often faces administrative and logistical challenges. By grouping aid recipients based on their demographic data and economic conditions, this study aims to improve the efficiency and effectiveness of aid distribution. A dataset from Kelurahan Prailiu, containing 350 aid recipient records, was analyzed using RapidMiner software. The clustering results show that dividing the data into 4 clusters yields the best outcome with a Davies Bouldin index of -1.085, indicating relatively close distances between items. Implementing data mining technology enables the government to identify the specific needs of recipient groups, reduce the risk of resource wastage and fund misuse, and enhance the positive impact of social aid programs. Recommendations are made to continuously develop the data management system and train social aid administrators in the use of this technology to ensure the sustainability and increased effectiveness of social aid programs.

Penulis Korespondensi :

Melan Astriani Mburu Hamu

Prodi Teknik Informatika

Universitas Kristen Wira Wacana Sumba

1. PENDAHULUAN

Bantuan sosial merupakan salah satu instrumen utama dalam upaya pemerintah untuk mengurangi kemiskinan, meningkatkan kesejahteraan sosial, dan mengatasi ketimpangan dalam masyarakat. Di Indonesia, seperti di banyak negara lainnya, bantuan sosial menjadi tulang punggung bagi jutaan individu dan keluarga yang menghadapi kesulitan ekonomi. Program bantuan sosial mencakup berbagai bentuk, mulai dari bantuan tunai, paket sembako, subsidi pendidikan dan kesehatan, hingga program-program pelatihan dan pengembangan keterampilan. Bantuan sosial memberikan akses yang penting bagi warga yang kurang mampu untuk memenuhi kebutuhan dasar mereka, seperti makanan, perawatan kesehatan, pendidikan, dan tempat tinggal yang layak.[1]

Meskipun bantuan sosial memberikan manfaat yang signifikan bagi penerima, pelaksanaannya sering kali dihambat oleh berbagai tantangan administratif dan logistik. Volume besar penerima, alokasi sumber daya yang terbatas, serta kebijakan yang kompleks menjadi beberapa masalah yang sering dihadapi.[2] Oleh karena itu, optimalisasi manajemen bantuan sosial sangat penting untuk memastikan bahwa bantuan tersebut disalurkan secara efisien, tepat waktu, dan tepat sasaran. Tanpa manajemen yang baik, risiko pemborosan sumber daya, penyalahgunaan, atau penyimpangan dana bantuan sosial menjadi tinggi. Optimalisasi manajemen juga memungkinkan pemerintah untuk mengukur dan mengevaluasi dampak program bantuan sosial secara lebih efektif, sehingga memungkinkan adanya perbaikan berkelanjutan.

Data mining merupakan pendekatan yang kuat dalam menggali pengetahuan yang tersembunyi dari data yang besar dan kompleks. Dalam konteks manajemen bantuan sosial, terdapat banyak data yang terkait dengan penerima bantuan, seperti data demografis, data ekonomi, riwayat penerimaan bantuan, dan lain sebagainya. Dengan menerapkan teknik data mining, pemerintah dapat mengidentifikasi pola-pola atau informasi berharga dari data ini, seperti tren kebutuhan, profil penerima yang rentan, atau kecenderungan penyalahgunaan. Dengan pemahaman yang lebih mendalam tentang data ini, pemerintah dapat membuat keputusan yang

lebih tepat dalam merancang dan melaksanakan program bantuan sosial. [3]

Algoritma k-means clustering dipilih karena kemampuannya dalam mengelompokkan data menjadi kelompok-kelompok yang serupa berdasarkan kesamaan fitur atau karakteristik. Dalam konteks manajemen bantuan sosial, algoritma ini dapat digunakan untuk mengelompokkan penerima bantuan sosial berdasarkan profil atau kondisi ekonomi mereka. Misalnya, algoritma k-means dapat membantu pemerintah dalam mengidentifikasi kelompok-kelompok yang memiliki kebutuhan serupa, seperti keluarga dengan anak-anak balita, lansia yang tinggal sendiri, atau keluarga dengan anggota yang memiliki disabilitas. [4] Dengan demikian, pemerintah dapat menyesuaikan strategi bantuan sosial secara lebih spesifik dan efektif sesuai dengan kebutuhan masing-masing kelompok, sehingga meningkatkan dampak positif dari program bantuan sosial tersebut.

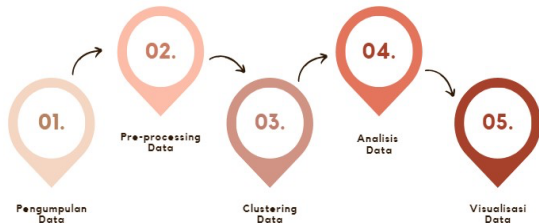
Dengan mengintegrasikan teknologi data mining dan algoritma k-means clustering dalam manajemen bantuan sosial, pemerintah dapat mencapai solusi yang lebih efektif dan efisien dalam menyediakan bantuan kepada warga yang membutuhkan. Hal ini tidak hanya akan mengurangi risiko pemborosan dan penyalahgunaan dana bantuan sosial, tetapi juga meningkatkan dampak positif dari program-program tersebut. Dengan demikian, implementasi data mining dan algoritma k-means clustering dalam manajemen bantuan sosial merupakan langkah strategis untuk meningkatkan kesejahteraan sosial dan mengurangi ketimpangan ekonomi di masyarakat.[5]

2. STATE OF THE ART

Dari beberapa penelitian terdahulu, dapat disimpulkan bahwa terdapat persamaan objek antara penelitian terdahulu dengan penelitian yang akan dilakukan peneliti yaitu Optimalisasi Manajemen Bantuan Sosial Dengan Implementasi Data Mining Menggunakan Algoritma K-Means Clustering. Optimalisasi yang dilakukan adalah optimalisasi manajemen bantuan sosial. Sedangkan perbedaan penelitian terdahulu dengan penelitian yang akan dilakukan peneliti antara lain adalah metode, lokasi dan waktu.

3. METODE PENELITIAN

Penelitian ini menggunakan studi literatur. Studi literatur ini dilakukan untuk memahami secara mendalam konsep-konsep dasar tentang bantuan sosial, teknologi data mining, algoritma k-



means clustering, serta praktik-praktik terbaik dalam manajemen bantuan sosial. Adapun hasil dari kajian pustaka digunakan untuk menyusun tahapan pengolahan data yang dijelaskan sebagai berikut:

Gambar 1. Tahapan Analisis Data

3.1 Pengumpulan Data Masyarakat

Pada tahapan ini data yang dikumpulkan berupa data mengenai penerima bantuan sosial dari sumber yang tersedia, seperti basis data pemerintah, lembaga sosial, atau organisasi non-pemerintah yang terlibat dalam program bantuan sosial. Data yang di kumpulkan berupa data demografis, sosial-ekonomi, data geografis dan kebutuhan dari masyarakat Kelurahan Prailiu. Jumlah data dari Kelurahan Prailiu yaitu sebanyak 350 data.

3.2 Pre-processing Data

Pada tahapan ini dilakukan proses persiapan dan penyiapan data sebelum dilakukan analisis lebih lanjut. Ini melibatkan serangkaian langkah untuk membersihkan, mengubah, dan mempersiapkan data agar sesuai dengan kebutuhan analisis.

Tabel 1. Inisialisasi Nilai Data

Jenis Kelamin	Laki-Laki	1
	Perempuan	2
Usia (Tahun)	20 - 35	1
	36 - 50	2
	51 - 65	3
	66 - 80	4
	≥ 81	5

Pekerjaan	Petani	1
	Pedagang	2
	Nelayan	3
	Buruh	4
	Tidak Bekerja	5
Penghasilan (Rp)	0	1
	100.000 - 500.000	2
	600.000 - 1.000.000	3
	1.100.000 - 1.500.000	4
	1.600.000 - 2.000.000	5
Jumlah Tanggungan	Tidak ada	1
	1 orang	2
	2 orang	3
	3 orang	4
	≥ 4 orang	5
Kondisi Rumah	Layak Huni	1
	Tidak Layak Huni	2

Tabel 2. Dataset Masyarakat

No.	Nama	Jenis Kelamin	Usia (Tahun)	Pekerjaan	Penghasilan (Rp)	Jumlah Tanggungan	Kondisi Rumah
1.	Amo*	1	1	1	2	5	1
2.	Abd**	1	3	2	2	5	1
3.	Abn**	1	1	1	3	4	2
4.	Abu****	1	2	2	3	5	1
5.	Abu***	2	1	4	4	4	1
6.	Ach****	1	3	1	2	5	1
7.	Bab***	2	2	1	2	5	1
8.	Bal**	1	3	1	2	3	1
9.	Bar*	1	3	1	2	2	2
10.	Bar****	1	2	3	3	3	1
...
350	Yon****	1	2	1	2	2	1

3.3 Clustering Data

Pada tahapan ini dilakukan clustering data menggunakan algoritma k-means dengan langkah-langkah sebagai berikut:

- Pilih Jumlah Cluster (k): Tentukan jumlah cluster yang diinginkan untuk pembagian data.
- Inisialisasi Pusat Cluster: Acak pilih k titik awal sebagai pusat cluster.
- Alokasikan Data ke Cluster Terdekat: Setiap data dalam dataset akan dialokasikan ke cluster terdekat berdasarkan jarak Euclidean ke

pusat-pusat cluster.

$$D(i, j) = \sqrt{(X1i - X1j)^2 + (X2i - X2j)^2 + \dots + (Xki - Xkj)^2}$$

Dimana :

- D(i,j) = Jarak data ke i ke pusat cluster j
- Xki = Data ke i pada atribut data ke k
- Xkj = Titik pusat ke j pada atribut ke k

- Hitung Ulang Pusat Cluster: Setelah semua data dialokasikan ke cluster, hitung ulang pusat cluster dengan menggunakan rata-rata dari semua data dalam setiap cluster.
- Tugaskan Ulang Data ke Pusat Cluster Baru: Setiap data akan diberikan label cluster baru berdasarkan pusat cluster yang baru dihitung.
- Evaluasi Konvergensi: Ulangi langkah-langkah 3 hingga 5 sampai tidak ada perubahan pada posisi pusat cluster atau hingga kriteria berhenti tertentu terpenuhi.

Berdasarkan tahapan yang telah dijelaskan, nilai K yang digunakan untuk clustering akan bervariasi antara 2, 3, 4, dan 5. Dalam beberapa percobaan pengujian, data akan dibagi ke dalam jumlah maksimal 5 cluster. Selanjutnya, kualitas setiap cluster akan dievaluasi menggunakan nilai Davis-Bouldin Index, di mana nilai yang mendekati 0 menunjukkan cluster dengan jarak yang baik. Proses pengolahan data ini akan dilakukan menggunakan perangkat lunak RapidMiner untuk mempermudah pengolahan dan visualisasi data.

3.4 Analisis Data

Pada tahapan ini dilakukan proses analisis data yang bertujuan untuk melihat jumlah data pada setiap cluster dan kemudian mengurutkannya berdasarkan peringkat terbaik. Proses ini juga akan membantu memberikan rekomendasi bagi pengelola bantuan sosial Kelurahan Prailiu untuk melihat potensi penerima bantuan secara minimal dan maksimal dari setiap kelompok yang terbentuk.

3.5 Visualisasi Data

Tahapan visualisasi data ini membantu untuk menggambarkan informasi yang terdapat dalam data dengan cara yang lebih mudah dipahami dan menarik. Ini memungkinkan untuk mengeksplorasi data secara visual dan mendapatkan wawasan yang lebih dalam dari hasil analisis. Visualisasi data ini

dapat berupa tabel dan diagram sehingga kita dapat melihat jumlah dan jarak dari setiap data masyarakat pada masing-masing cluster yang terbentuk.

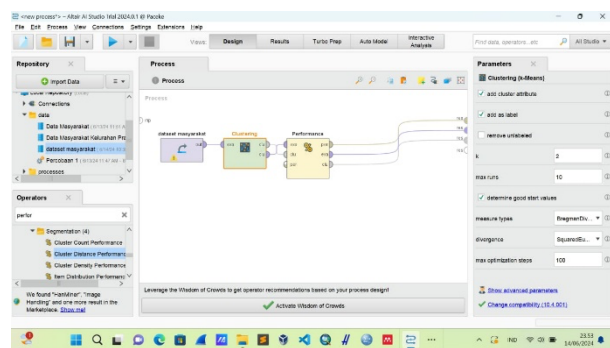
4. HASIL DAN PEMBAHASAN

Proses pengolahan data dimulai dengan mengimpor dataset masyarakat ke dalam software RapidMiner. Dataset ini diatur dengan menjadikan Nama sebagai Label, sedangkan Jenis Kelamin, Usia, Pekerjaan, Penghasilan, Jumlah Tanggungan, dan Kondisi Rumah diatur dengan tipe data integer.

Row No.	Nama	Jenis Keta.	Usia (Tahun)	Pekerjaan	Penghasilan	Jumlah Tan.	Kondisi Ru..
1	AM O SUPU..	1	1	1	2	5	1
2	ABDULAZIZ	1	3	2	2	5	1
3	ABDUL K..	1	1	1	4	3	3
4	AKHILAHAM S.	1	7	2	3	6	1
5	AKULANDAR..	2	1	4	4	4	1
6	ADHANI UM..	1	2	1	2	0	1
7	ABIRAND IOTI.	2	2	1	2	0	1
8	BELLA NURUL..	1	2	1	2	3	1
9	SARA KUSIA..	1	3	1	2	2	2
10	HARTAWATI..	1	7	3	3	3	1
11	SURYOSUD..	1	4	2	3	4	1
12	MENDILIN..	2	1	1	2	0	1
13	CHERLES TAN..	1	2	1	2	0	2

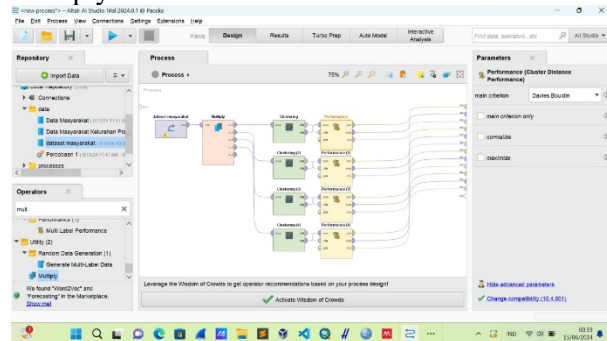
Gambar 2. Dataset Masyarakat

Setelah proses input dataset, langkah berikutnya adalah memilih modul dengan menetapkan nilai awal K sebesar 2 dan menggunakan jenis pengukuran Bregman divergences karena semua data yang diolah adalah data numerik. Kemudian, menggunakan perpindahan maksimum titik centroid hingga 100 kali. Sebelum melakukan clustering, perlu ditambahkan pengukuran performa untuk melihat nilai pembagian cluster, yang akan terbagi menjadi 2 dengan Davies Bouldin sebesar -1.365.



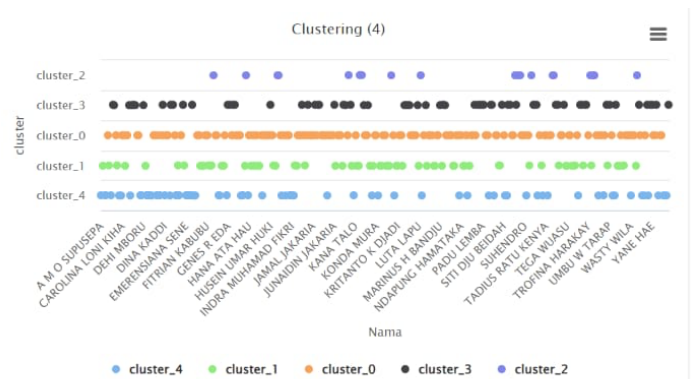
Gambar 3. Pengaturan Cluster

Pada tahap pertama, penggunaan nilai $K = 2$ menghasilkan 198 data di Cluster 0 dan 152 data di Cluster 1 dengan nilai Davies Bouldin sebesar -1.365. Nilai ini menunjukkan jarak yang relatif dekat antara setiap item. Namun, untuk menentukan apakah analisis data dengan nilai $K = 2$ cukup baik, kita perlu menguji beberapa nilai K lainnya. Pada proses berikutnya, kita akan mencoba nilai $K = 3$, $K = 4$, dan $K = 5$ dengan menambahkan modul tambahan yaitu modul multiply data.



Gambar 4. Multiply Data

Setelah melakukan cluster dengan nilai k yang maka ditemukan bahwa Cluster dengan Nilai $K = 3$ terbagi menjadi Cluster 0: 180 items, Cluster 1: 26 items, Cluster 2: 144 items, dengan total number of items: 350 dengan nilai Davies Bouldin: -1.171, sedangkan untuk Nilai $K = 4$ terbagi menjadi Cluster 0: 134 items, Cluster 1: 120 items, Cluster 2: 19 items, Cluster 3: 77 items, dengan total number of items: 350 dengan nilai Davies Bouldin: -1.089, kemudian nilai $K = 5$ terbagi menjadi Cluster 0: 129 items, Cluster 1: 68 items, Cluster 2: 19 items, Cluster 3: 68 items, Cluster 4: 66 items, dengan total number of items: 350 dengan nilai Davies Bouldin: -1.085, Sehingga dapat disimpulkan pada kasus ini data lebih baik dibagi menjadi 4 cluster karena memiliki jarak yang cukup dekat dengan nilai Davies Bouldin = -1.085. Dari pembagian data cluster data dapat divisualisasikan agar dapat terlihat jelas jarak antar items dan pembagian cluster.



Gambar 5. Visualisasi Data Cluster

5. SIMPULAN

Penelitian ini menunjukkan bahwa penggunaan teknik data mining, khususnya algoritma k-means clustering, dapat mengoptimalkan manajemen bantuan sosial. Dengan mengelompokkan penerima bantuan berdasarkan profil dan kondisi ekonomi, pemerintah dapat menyesuaikan strategi bantuan sosial secara lebih efektif. Hasil analisis menunjukkan bahwa pembagian data menjadi 4 cluster memberikan hasil terbaik dengan nilai Davies Bouldin sebesar -1.085, menunjukkan jarak antar item yang relatif dekat. Implementasi teknologi ini memungkinkan alokasi sumber daya yang lebih efisien, mengurangi risiko pemborosan, dan meningkatkan dampak positif dari program bantuan sosial. Visualisasi data lebih lanjut membantu menggambarkan pembagian dan jarak antar item dalam masing-masing cluster, mendukung pengambilan keputusan yang lebih tepat.

Disarankan agar pemerintah terus mengembangkan dan memperbarui sistem pengelolaan data dengan teknologi data mining untuk memastikan akurasi dan relevansi data penerima bantuan. Selain itu, pelatihan bagi pengelola bantuan sosial tentang penggunaan teknologi ini akan meningkatkan efektivitas implementasi. Evaluasi berkala dan penyesuaian strategi berdasarkan hasil analisis data juga penting untuk memaksimalkan manfaat program bantuan sosial.

DAFTAR PUSTAKA

- [1] A. C. Talakua, “Penerapan Data Mining untuk memprediksi Nilai Akhir Matakuliah Menggunakan Metode K-Means Clustering Application of Data Mining to predict Final Course Grades Using the K-Means Clustering Method,” vol. 2, no. Desember, pp. 2962–5998, 2023.
- [2] S. E. O. Utline, “data,” vol. 5, no. 2, pp. 1612–1614, 2014.
- [3] R. Buatun and S. Solikhun, “The Application of Numerical Measure Variations in K-Means Clustering for Grouping Data,” *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 23, no. 1, pp. 103–112, 2023, doi: 10.30812/matrik.v23i1.3269.
- [4] S. K. Dini and A. Fauzan, “Clustering Provinces in Indonesia based on Community Welfare Indicators,” *EKSAKTA J. Sci. Data Anal.*, vol. 1, no. 1, pp. 56–63, 2020, doi: 10.20885/eksakta.vol1.iss1.art9.
- [5] W. Sarjono and M. Samiaji, “Crisp-DM Implementation for Elderly Social Protection,” *Int. J. Soc. Serv. Res.*, vol. 3, no. 11, pp. 2842–2850, 2023, doi: 10.46799/ijssr.v3i11.591.