

## CLUSTERING PENERIMA BANTUAN SOSIAL MENGGUNAKAN ALGORITMA K-MEANS DI DESA JATIPANCUR

Muhammad Ma'arif Syaefullah<sup>1</sup>, Martanto<sup>2</sup>, Umi Hayati<sup>3</sup>

<sup>1,3</sup>Teknik Informatika, <sup>2</sup>Manajemen Informatika, STMIK IKMI Cirebon  
Jalan Perjuangan No. 10 B Majasem, Kota Cirebon, Indonesia  
25muhammad.arif@gmail.com

### ABSTRAK

Program bantuan sosial ialah upaya pemerintah dalam membagikan sokongan kepada warga yang memerlukan, khususnya di wilayah pedesaan. Namun tidak jarang program bantuan sosial terjadi masalah diantaranya tidak tepat sasaran, salah pendataan dan kesalahan lainnya. Permasalahan yang sering terjadi pada saat penyaluran bantuan sosial di desa jatipancur ialah tidak tepat sasaran. Penelitian ini bertujuan guna mempraktikkan metode Informasi Mining dengan memakai algoritma *K-Means* dalam konteks clustering penerima dorongan sosial dengan metode *Knowledge Discovery in Databases (KDD)*. Sumber data diambil dari data puskesmas amanah desa jatipancur. Langkah awal mengaitkan pra- pemrosesan informasi guna mensterilkan serta menanggulangi kemampuan permasalahan, semacam informasi yang tidak lengkap ataupun informasi yang tidak relevan. Berikutnya, algoritma *K-Means* diimplementasikan untuk mengelompokkan informasi jadi sebagian kelompok bersumber pada kesamaan atribut. Dari proses penerapan teknik data mining dengan algoritma *K-Means Clustering* dihasilkan kelompok penerima bantuan sosial dengan jumlah cluster 10 terdiri dari *cluster\_0* berjumlah 36 kepala keluarga, *cluster\_1* berjumlah 2 kepala keluarga, *cluster\_2* berjumlah 23 kepala keluarga, *cluster\_3* berjumlah 54 keluarga, *cluster\_4* berjumlah 16 kepala keluarga, *cluster\_5* berjumlah 156 kepala keluarga, *cluster\_6* berjumlah 5 kepala keluarga, *cluster\_7* berjumlah 20 keluarga, *cluster\_8* berjumlah 40 kepala keluarga, *cluster\_9* berjumlah 2 kepala keluarga dan nilai *Davies Bouldin Index* yang optimal yaitu 0,057.

**Kata kunci :** Data Mining, Clustering, Penerima Bantuan Sosial, Algoritma K-Means, Knowledge Discovery in Databases (KDD).

### 1. PENDAHULUAN

Perhatian pemerintahan terhadap kemiskinan terkait dengan kesulitan memenuhi kebutuhan dasar seperti sandang, pangan, dan tempat tinggal. Kompleksitas masalah ini berdampak pada ketidakstabilan ekonomi, terbatasnya akses publik, kualitas pendidikan rendah, kesehatan yang buruk, dan kurangnya kesempatan ekonomi. Meskipun program bantuan sosial membantu masyarakat yang memerlukan, terutama yang kurang mampu secara finansial, seringkali terjadi kendala seperti ketidakakuratan penyaluran, kesalahan data, dan hambatan lainnya.[1]

Isu utama dalam penyaluran bantuan sosial di Desa Jatipancur adalah kurangnya akurasi dalam menentukan penerima manfaat sehingga sebagian besar penerima dianggap mampu secara ekonomi. Hal ini disebabkan oleh metode pengolahan data yang masih bergantung pada keputusan langsung melalui musyawarah oleh pemerintah desa untuk menentukan penerima bantuan. Pendekatan ini kurang transparan dan dapat menimbulkan ketidakseimbangan sosial serta kecurigaan terhadap pemerintah desa dan lembaga yang mengelola bantuan sosial. Kemajuan teknologi yang semakin pesat saat ini menawarkan solusi untuk mengatasi permasalahan penyaluran bantuan sosial ini[2]. Oleh karena itu, diperlukan suatu metode yang dapat secara jelas mengidentifikasi penerima bantuan yang layak dan tidak layak di Desa Jatipancur.

Dalam riset ini, peneliti menggunakan perangkat lunak RapidMiner untuk melakukan pemrosesan data penerima bantuan sosial. Pendekatan yang digunakan adalah teknik data mining menggunakan metode K-Means untuk mengelompokkan penerima bantuan sosial. Penerapan metode ini memiliki kepentingan yang besar dalam menghindari tingkat kesalahan yang tinggi yang biasa terjadi dalam pengelolaan data di lingkungan masyarakat. Hal ini juga menekankan pentingnya pendataan yang akurat untuk mencegah kesalahan yang mungkin terjadi.

Tujuan utama penelitian ini adalah untuk mengelompokkan penerima bantuan sosial di Desa Jatipancur agar dapat disalurkan kepada individu atau kelompok yang membutuhkan, sekaligus mencari nilai K yang paling optimal berdasarkan evaluasi dari Davies Bouldin Index[3]. Melalui pendekatan yang lebih akurat dan tepat sasaran dalam penyaluran bantuan sosial, diharapkan alokasi dana yang disalurkan dapat memberikan dampak yang lebih signifikan dan nyata bagi kesejahteraan masyarakat Desa Jatipancur. Upaya ini diharapkan dapat mengurangi keraguan atau ketidakpercayaan masyarakat terhadap pemerintah desa dan lembaga yang bertanggung jawab dalam pengelolaan bantuan sosial di Desa Jatipancur.

### 2. TINJAUAN PUSTAKA

Penelitian yang dilakukan dalam lima tahun terakhir oleh para peneliti sebelumnya memiliki relevansi signifikan dengan artikel yang sedang diteliti

saat ini. Hasil studi-studi tersebut menggarisbawahi beberapa temuan yang terkait langsung dengan topik yang sama yang sedang diselidiki. Kajian-kajian terdahulu tersebut memberikan kontribusi penting dalam pemahaman yang lebih mendalam terkait isu yang sedang dibahas dalam penelitian ini. Seperti pada penelitian yang dilakukan oleh Siti Amaliyah[4], membahas mengenai analisis kelayakan menerima bantuan sosial menggunakan algoritma K-Means dengan enam atribut untuk mengevaluasi kelayakan penerima bantuan sosial. Bertujuan memperbaiki efisiensi penyaluran bantuan agar lebih tepat sasaran. Analisis menunjukkan bahwa Desa Kuala Dendang mencatat 74 Kepala Keluarga menerima Bantuan PKH. Namun, analisis K-Means mengidentifikasi 114 kepala keluarga (dalam Cluster 1) layak mendapat bantuan, di mana 40 di antaranya tidak tercatat dalam data Desa Kuala Dendang. Pada penelitian yang dilakukan oleh Asep Surahman[5] membahas prediksi penerima bantuan sosial menggunakan algoritma Naïve Bayes untuk menentukan calon penerima bantuan yang layak. Algoritma Naïve Bayes merupakan metode statistik yang memprediksi probabilitas keanggotaan suatu kelas. Penelitian ini menggunakan algoritma tersebut karena memiliki akurasi dan kecepatan yang tinggi, cocok untuk database besar. Hasil pengujian menunjukkan bahwa penggunaan algoritma Naïve Bayes dalam clustering memberikan akurasi sebesar 91.85%. Dalam penelitian[6] ditemukan tiga cluster dengan karakteristik yang berbeda terkait tingkat kemiskinan. Cluster(0) memiliki tingkat kemiskinan sedang dengan jumlah penduduk miskin yang sedang, pengeluaran per kapita sedang, dan rata-rata lama waktu sekolah juga sedang. Cluster(1) menunjukkan tingkat kemiskinan tertinggi dengan jumlah penduduk miskin terbanyak, pendidikan rendah, dan pengeluaran per kapita yang paling sedikit. Sementara Cluster(2) menunjukkan tingkat kemiskinan terendah dengan jumlah penduduk miskin yang sedikit, pendidikan tinggi, dan pengeluaran per kapita tertinggi. Berdasarkan hasil ini, pemerintah disarankan untuk memprioritaskan bantuan di Kabupaten Pandeglang, Kabupaten Lebak, dan Kabupaten Serang, terutama dalam hal beasiswa pendidikan, dana sosial, dan perbaikan infrastruktur untuk meningkatkan kesejahteraan penduduk Banten.

### 3. METODE PENELITIAN

Data untuk penelitian ini berasal dari Pusat Kesejahteraan Sosial Amanah di Desa Jatipancur pada bulan Desember 2022. Sumber data ini terdiri dari data primer dan data sekunder. Data primer diperoleh melalui observasi dan wawancara langsung dengan staf Pusat Kesejahteraan Sosial Amanah serta anggota masyarakat yang merupakan penerima manfaat dari program bantuan sosial. Sedangkan data sekunder diperoleh dari kantor sekretariat Pusat Kesejahteraan Sosial Amanah di Desa Jatipancur dengan izin dari

kepala desa dan koordinator Pusat Kesejahteraan Sosial Amanah di desa tersebut.

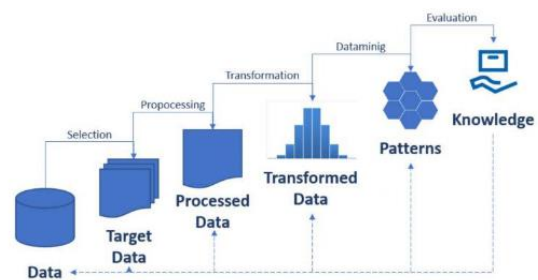
Dalam penelitian ini, pengumpulan data dilakukan melalui dua teknik, yaitu observasi dan wawancara.

#### 1) Teknik Observasi

Penulis melakukan observasi dengan melakukan pengamatan langsung di Desa Jatipancur. Observasi dilakukan pada minggu pertama bulan Oktober di kantor desa Jatipancur.

#### 2) Teknik Wawancara

Dalam teknik wawancara, penulis melakukan wawancara dengan Bapak Ahmad Jaelani, seorang pegawai Pusat Kesejahteraan Sosial (Puskesmas) Amanah di Desa Jatipancur.



Gambar 1. Proses Metode Penelitian

Penelitian ini menggunakan tahapan *Knowledge Discovery in Database (KDD)* seperti pada gambar 1. yang bertujuan untuk mengeksplorasi serta menganalisis dataset besar agar menghasilkan informasi yang bermanfaat di bidang pendidikan. Algoritma yang dipakai dalam mengklasifikasikan penerima bantuan sosial ialah algoritma k-means[7]. Secara umum, tahapan Knowledge Discovery in Databases (KDD) dapat diuraikan sebagai berikut:

#### 1) Data Selection

Sebelum melangkah ke tahap data mining, langkah awal adalah seleksi data dengan menampilkan data aslinya. Data yang telah dipilih akan disimpan secara terpisah dari basis data operasional.

#### 2) Data Preprocessing

Langkah selanjutnya adalah pembersihan atau preprocessing data. Hal ini mencakup seleksi data yang relevan, menghapus duplikasi, serta memeriksa dan memperbaiki kesalahan data guna memastikan konsistensi data.

#### 3) Data Transformation

Transformasi adalah proses ketiga di mana informasi yang telah dipilih diubah ke dalam bentuk yang sesuai dengan prosedur penambangan data agar dapat digunakan secara efektif sesuai dengan tujuan pencarian pola atau informasi dalam basis data.

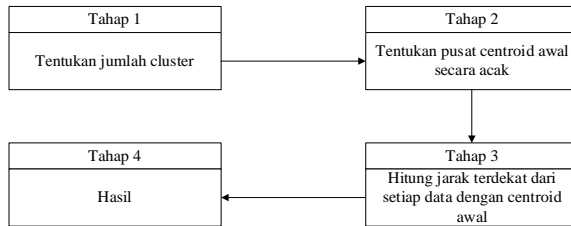
#### 4) Data Mining

Data mining, sebagai inti dari proses KDD, melibatkan penggunaan berbagai metode untuk

576

#### 4.4. Data Mining

Pada tahap ini penulis menggunakan metode data mining dengan algoritma K-Means. K-Means merupakan metode pengelompokan data numerik berdasarkan jarak antar data, yang membagi data ke dalam kelompok dengan jumlah tertentu. Penulis melakukan percobaan K2 sampai dengan K10 dan maksimal iterasi 10 agar mendapatkan nilai K yang optimal untuk proses *K-Means Clustering* dan iterasi dilakukan 10 kali agar tidak ada perubahan yang signifikan terhadap posisi centroid.



Gambar 7. Langkah *clustering k-means*

Adapun langkah-langkah algoritma *K-Means* sebagai berikut[8]:

- 1) Tentukan jumlah cluster
- 2) Tentukan pusat centeroid awal, pada dasarnya tahap ini dipilih dari data secara random

$$V_{ij} = \frac{1}{N_i} \sum_{k=0}^{N_i} X_{kj} \quad (1)$$

Keterangan:

$V_{ij}$  = Centroid rata-rata cluster ke-I untuk variabel ke-j

$N_i$  = Jumlah anggota cluster ke-i

i, k = Indeks dari cluster

j = Indeks dari variabel

$X_{kj}$  = nilai data ke-k variabel ke-j untuk cluster tersebut

- 3) Hitung jarak terdekat dari setiap data dengan centeroid awal yang sudah ditentukan menggunakan rumus Euclidean Distance.

$$De = \sqrt{(X_i - S_i)^2 + (y_i - t_i)^2} \quad (2)$$

Keterangan:

De = Euclidean Distance

i = Banyaknya objek

(x, y) = Koordinat objek

(s, t) = Koordinat centroid

Rumus dalam menentukan nilai *DBI* sebagai berikut[9]:

$$DBI = \frac{1}{K} \sum_{i=1}^K \text{MAX}_{i \neq j} (R_{ij}) \quad (3)$$

Keterangan:

K = jumlah cluster yang dihasilkan oleh *K-Means*

$R_{ij}$  = rata-rata jarak antara semua pasangan titik data dalam cluster I ke pusat cluster j

Setelah dilakukan percobaan dengan K=2 sampai dengan K=10, dihasilkan nilai DBI yang optimal menggunakan Operator *Cluster Distance Performance* adalah K=10 dengan nilai 0,057.

Proses performance pada RapidMiner dihasilkan nilai DBI K=2 pada gambar 8. berikut.

#### Davies Bouldin

Davies Bouldin: 0.137

Gambar 8. Nilai DBI pada k=2

Proses performance pada RapidMiner dihasilkan nilai DBI K=3 pada gambar 9. berikut.

#### Davies Bouldin

Davies Bouldin: 0.115

Gambar 9. Nilai DBI pada k=3

Proses performance pada RapidMiner dihasilkan nilai DBI K=4 pada gambar 10. berikut.

#### Davies Bouldin

Davies Bouldin: 0.092

Gambar 10. Nilai DBI pada k=4

Proses performance pada RapidMiner dihasilkan nilai DBI K=5 pada gambar 11. berikut.

#### Davies Bouldin

Davies Bouldin: 0.096

Gambar 11. Nilai DBI pada k=5

Proses performance pada RapidMiner dihasilkan nilai DBI K=6 pada gambar 12. berikut.

**Davies Bouldin**

Davies Bouldin: 0.079

Gambar 12. Nilai DBI pada k=6  
Proses performance pada RapidMiner dihasilkan nilai DBI K=7 pada gambar 13. berikut.

**Davies Bouldin**

Davies Bouldin: 0.075

Gambar 13. Nilai DBI pada k=7

Proses performance pada RapidMiner dihasilkan nilai DBI K=8 pada gambar 14. berikut.

**Davies Bouldin**

Davies Bouldin: 0.077

Gambar 14. Nilai DBI pada k=8

Proses performance pada RapidMiner dihasilkan nilai DBI K=9 pada gambar 15. berikut.

**Davies Bouldin**

Davies Bouldin: 0.080

Gambar 15. Nilai DBI pada k=9

Proses performance pada RapidMiner dihasilkan nilai DBI K=10 pada gambar 16. berikut.

**Davies Bouldin**

Davies Bouldin: 0.057

Gambar 16. Nilai DBI pada k=10

**4.5. Evaluation**

Dalam tahap ini, dilakukan evaluasi terhadap data untuk mengukur tingkat akurasi dalam proses pengelompokan data (clustering). Evaluasi clustering bertujuan untuk menilai kualitas dari hasil pengelompokan yang telah dilakukan. Penelitian ini menggunakan Davies Bouldin Index (DBI) sebagai metode evaluasi internal clustering guna menentukan jumlah cluster yang optimal. DBI digunakan untuk mengukur seberapa baik hasil clustering, dimana nilai DBI yang lebih rendah menandakan kualitas clustering yang lebih baik serta menunjukkan jumlah cluster yang optimal[10][9]. Evaluasi dilakukan di platform RapidMiner dengan menggunakan operator Cluster Distance Performance dalam serangkaian eksperimen yang mencakup pengujian dengan 2, 3, 4, 5, 6, 7, 8, 9, dan 10 cluster untuk menentukan konfigurasi cluster yang paling tepat.

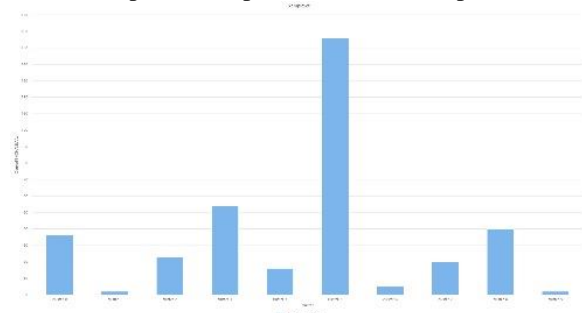
Tabel 1. Tabel Perbandingan Nilai DBI Tiap Cluster

Percobaan	Jumlah Cluster	Nilai DBI
1	K2	0,137
2	K3	0,115
3	K4	0,092
4	K5	0,096
5	K6	0,079
6	K7	0,075
7	K8	0,077
8	K9	0,080
9	K10	0,057

Pada tabel 1. dapat dilihat setelah dilakukan proses evaluasi didapatkan tingkat performance yang paling baik terdapat pada K(10) dengan nilai DBI = 0.057. sehingga dipilih K yang akan digunakan yaitu K=10.

**4.6. Knowledge**

Kemudian hasil pengelompokan cluster penulis visualisasikan kedalam bentuk chart Bar(Column) untuk mempermudah pemahaman terhadap data.



Gambar 17. Hasil visualisasi bentuk bar(column)

Dari gambar 17. Merupakan visualisasi data cluster yang berada pada K=10 yaitu terdapat 9 cluster/items yang dikelompokkan berdasarkan penghasilannya. Dapat dilihat bahwa penghasilan terbanyak terdapat pada cluter\_5.

## 5. KESIMPULAN DAN SARAN

Setelah melaksanakan teknik data mining menggunakan algoritma K-Means Clustering pada data penerima bantuan sosial di Desa Jatipancur, dihasilkan penerapan teknik data mining untuk mengelompokkan penerima bantuan sosial telah dilakukan menggunakan pendekatan Knowledge Discovery in Databases (KDD) serta algoritma K-Means Clustering, menggunakan data yang telah terkumpul. Proses ini melibatkan penggunaan software RapidMiner dengan menerapkan 6 operator, seperti Read Excel, Select Attributes, Set Role, Nominal to Numerical, K-Means, dan Performance. Evaluasi dilakukan terhadap nilai K yang optimal dari 10 percobaan pada algoritma K-Means Clustering, dihasilkan nilai k=2, dengan nilai Davies Bouldin Index (DBI) sebesar 0,137, nilai k=3, dengan nilai Davies Bouldin Index (DBI) sebesar 0,115, nilai k=4, dengan nilai Davies Bouldin Index (DBI) sebesar 0,092, nilai k=5, dengan nilai Davies Bouldin Index (DBI) sebesar 0,096, nilai k=6, dengan nilai Davies Bouldin Index (DBI) sebesar 0,079, nilai k=7, dengan nilai Davies Bouldin Index (DBI) sebesar 0,075, nilai k=8, dengan nilai Davies Bouldin Index (DBI) sebesar 0,077, nilai k=9, dengan nilai Davies Bouldin Index (DBI) sebesar 0,080, nilai k=10, dengan nilai Davies Bouldin Index (DBI) sebesar 0,057. Sehingga nilai K yang optimal terdapat pada K=10. Adapun saran untuk pemerintah Desa Jatipancur adalah meningkatkan selektivitas dalam pemilihan penerima bantuan sosial dengan memberikan prioritas pada individu yang telah keluar dari masa produktif bekerja, sehingga dapat lebih efektif mengatasi tingkat kemiskinan di wilayah tersebut. Selain itu, untuk penelitian berikutnya, direkomendasikan untuk memperluas variabel yang digunakan dalam analisis dan meningkatkan jumlah iterasi eksperimen guna menentukan jumlah klaster dengan nilai Davies Bouldin Index (DBI) yang optimal untuk pengelompokan yang lebih baik.

## DAFTAR PUSTAKA

- [1] L. G. R. Putra and A. Anggrawan, "Pengelompokan Penerima Bantuan Sosial Masyarakat dengan Metode K-Means," *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 21, no. 1, pp. 205–214, 2021, [Online]. Available: <https://journal.universitasbumigora.ac.id/index.php/matrik/article/view/1554>
- [2] S. G. Pratama, P. Mahmudi, Ali, B.eng, and M. S. Dr.Ir. Achmadi, Sentot, *KLASIFIKASI PENENTUAN PENERIMA BANTUAN PANGAN NON TUNAI MENGGUNAKAN METODE K-MEANS CLUSTERING*. eprints.itn.ac.id, 2020. [Online]. Available: <http://eprints.itn.ac.id/4547/>
- [3] B. Baskoro, A. Gunaryati, and A. Rubhasy, "KLASIFIKASI PENDUDUK KURANG MAMPU DENGAN METODE K-MEANS UNTUK OPTIMALISASI PROGRAM BANTUAN SOSIAL," *INFOMATEK J. Inform. Manaj. dan Teknol.*, vol. 25, no. 2, pp. 41–48, 2023, [Online]. Available: <http://repository.unas.ac.id/6521/>
- [4] S. Amaliyah and S. R. Agustini, "Penerapan Data Mining Untuk Menentukan Kelompok Prioritas Penerima Bantuan PKH Menggunakan Metode Clustering K-Means Pada Desa Kuala Dendang," *J. Inform. dan rekayasa Komput.*, vol. 3, no. 1, pp. 453–458, 2023.
- [5] A. Surahman and U. Hayati, "IMPLEMENTASI ALGORITMA NAÏVE BAYES UNTUK PREDIKSI PENERIMA BANTUAN SOSIAL," *Jati (Jurnal Mhs. Tek. Inform.)*, vol. 7, no. 1, pp. 1–6, 2023.
- [6] Y. R. Sari, A. Sudewa, D. A. Lestari, and T. I. Jaya, "Penerapan Algoritma K-Means Untuk Clustering Data Kemiskinan Provinsi Banten Menggunakan Rapidminer," *CESS (Journal Comput. Eng. Syst. Sci.)*, vol. 5, no. 2, pp. 192–198, 2020, [Online]. Available: <https://jurnal.unimed.ac.id/2012/index.php/cess/article/view/18519>
- [7] M. K. Mulyadien and U. Enri, "Algoritma K-Means Untuk Pengelompokan Bantuan Langsung Tunai (BLT)," *J. Ilm. Wahana Pendidik.*, vol. 8, no. 12, pp. 198–210, 2022, [Online]. Available: <http://jurnal.peneliti.net/index.php/JIWP/article/view/2025>
- [8] Z. Nabila, A. R. Isnain, P. Permata, and Z. Abidin, "Analisis Data Mining Untuk Clustering Kasus Covid-19 Di Provinsi Lampung Dengan Algoritma K-Means," *J. Teknol. Dan Sist. Inf.*, vol. 2, no. 2, pp. 100–108, 2021, [Online]. Available: <http://jim.teknokrat.ac.id/index.php/sisteminformasi/article/view/868>
- [9] M. Orisa, "Optimasi Cluster pada Algoritma K-Means," *Pros. SENIATI*, vol. 6, no. 2, pp. 430–437, 2022, doi: 10.36040/seniati.v6i2.5034.
- [10] E. Muningsih, I. Maryani, and V. R. Handayani, "Penerapan Metode K-Means dan Optimasi Jumlah Cluster dengan Index Davies Bouldin untuk Clustering Propinsi Berdasarkan Potensi Desa," *J. Sains dan Manaj.*, vol. 9, no. 1, p. 96, 2021, [Online]. Available: [www.bps.go.id](http://www.bps.go.id)