

Klasifikasi Penentuan Prioritas Bantuan PKH Dengan Clustering K-Means dan C4.5 Pada Desa Talang Belido

Siti Amaliyah¹, Jasmir², Sharipuddin³

¹²³ Fakultas Ilmu Komputer, Program Studi Magister Sistem Informasi, Universitas Dinamika Bangsa, Kota Jambi, Indonesia
Email: ¹sitiomalayah3296@gmail.com, ²ijay_jasmir@yahoo.com, ³sharifbuhaira@gmail.com

Email Penulis Korespondensi: sitiomalayah3296@gmail.com

Artikel Info :

Artikel History :

Submitted : 19-04-2025

Accepted : 09-06-2025

Published : 30-04-2026

Kata Kunci:

PKH; Data_Mining,
Clustering; K-Means;
C4.5;
Davies_Bouldin_Index;;
Rapidminer

Abstrak– Program Keluarga Harapan (PKH) adalah bantuan sosial bersyarat yang diberikan kepada Keluarga Miskin (KM) yang telah ditetapkan sebagai penerima manfaat. Namun, distribusi bantuan ini masih menghadapi tantangan dalam hal pemerataan dan ketepatan sasaran. Pengelolaan data yang besar, efisiensi waktu, dan akurasi data menjadi tantangan tersendiri dalam menentukan penerima yang paling berhak di Desa Talang Belido. Untuk mengatasi hal tersebut, penelitian ini menerapkan teknik data mining dengan menggabungkan metode *Clustering K-Means* dan algoritma *C4.5*, yang bertujuan untuk mempermudah proses pengelompokan serta memberikan gambaran pohon keputusan dari data yang dianalisis. Penelitian ini menggunakan data tahun 2023-2024 sebanyak 1.357 sampel dengan atribut meliputi: nama kepala keluarga, pekerjaan, jumlah tanggungan, pendapatan, kondisi rumah, status kepemilikan rumah, dan status kesejahteraan penduduk. Data tersebut diolah menggunakan tools *Rapidminer* dan menghasilkan tiga *cluster*: *Cluster 1* sebagai prioritas utama (layak) dengan 377 anggota, *Cluster 2* sebagai bahan pertimbangan dengan 482 anggota, dan *Cluster 3* sebagai kategori tidak layak dengan 498 anggota. Evaluasi menunjukkan nilai akurasi metode *Clustering* berdasarkan *Davies-Bouldin Index (DBI)* sebesar 0,187 dengan *silhouette score* 0,75, sementara akurasi algoritma *C4.5* mencapai 99,7%.

Keywords:

PKH; Data_Mining; K-Means_Clustering; C4.5;
Davies_Bouldin_Index;
Rapidminer

Abstract– The Family Hope Program (PKH) is a conditional social assistance program provided to Poor Families (KM) who have been designated as beneficiaries. However, the distribution of this assistance still faces challenges in terms of equity and accuracy. The management of large data volumes, time efficiency, and data accuracy are significant challenges in determining the most eligible recipients in Talang Belido Village. To address these issues, this study applies data mining techniques by combining the K-Means Clustering method and the C4.5 algorithm, aiming to simplify the grouping process and provide a decision tree representation of the analyzed data. This study utilizes data from the years 2023-2024, consisting of 1,357 samples with attributes including the head of the family's name, occupation, number of dependents, income, housing condition, homeownership status, and residents' welfare status. The data was processed using RapidMiner tools and resulted in three clusters: Cluster 1 as the top priority (eligible) with 377 members, Cluster 2 as a consideration with 482 members, and Cluster 3 as the ineligible category with 498 members. The evaluation indicated that the accuracy of the Clustering method, based on the Davies-Bouldin Index (DBI), was 0.187, and silhouette score 0.75 while the accuracy of the C4.5 algorithm reached 99.7%.

1. PENDAHULUAN

Kemiskinan merupakan salah satu masalah mendasar yang menjadi fokus perhatian pemerintah di setiap negara, termasuk Indonesia [1]. Ada berbagai faktor yang menyebabkan kemiskinan di Indonesia, seperti tingginya pertumbuhan penduduk, rendahnya tingkat pendidikan, kurangnya motivasi untuk bekerja, terbatasnya peluang kerja yang tersedia, dan lain sebagainya [2]. Kombinasi dari berbagai faktor ini membuat kemiskinan menjadi isu yang kompleks untuk diatasi. Menurut Badan Pusat Statistik [3] pada bulan maret 2024, tingkat kemiskinan di Indonesia mencapai 9,03%. Menteri Koordinator bidang Pembangunan Manusia dan Kebudayaan RI, Muhadjir Effendy [4] mengatakan anggotanya terus berupaya mencapai target yang telah ditetapkan oleh Presiden Joko Widodo, yaitu mengurangi tingkat kemiskinan hingga mendekati 7,5%.

Salah satu program yang menjadi andalan pemerintah dalam mengurangi kemiskinan adalah Program Keluarga Harapan (PKH). PKH adalah program untuk mengurangi kemiskinan dan mengembangkan sistem perlindungan sosial bersyarat terhadap masyarakat miskin, yang bertujuan mempercepat pencapaian *Millennium Development Goals (MDGs)* [5]. Meskipun tujuan program ini sangat baik, dalam praktiknya penyaluran bantuan PKH ini sering menghadapi berbagai masalah terkait pengolahan data dan ketidakmerataan dalam penyaluran bantuan [6]. Permasalahan ini juga dialami oleh Desa Talang Belido. Warga mengeluhkan bahwa penyaluran bantuan PKH dianggap tidak merata. Hal ini dikarenakan warga yang merasa berhak mendapatkan

bantuan justru tidak terpilih, sementara warga yang dianggap tidak berhak mendapatkan bantuan justru menerima bantuan tersebut.

Permasalahan lainnya adalah jumlah data penduduk yang besar di Desa Talang Belido menyebabkan pengelola data membutuhkan waktu lama untuk menentukan penerima bantuan PKH. Hal ini dikarenakan verifikasi data masih dilakukan secara manual dengan memeriksa data satu per satu. Selain itu, terdapat data yang belum diperbarui atau tidak lengkap. Terkadang, pihak desa hanya menerima usulan dari setiap RT terkait warga yang direkomendasikan untuk mendapatkan bantuan PKH. Untuk mengatasi masalah pengolahan data tersebut dan penentuan prioritas warga yang berhak mendapat bantuan Program Keluarga Harapan (PKH), bisa menggunakan teknik *data mining*, salah satunya dengan metode *Clustering K-Means* dan algoritma C4.5.

Hal ini bisa dibuktikan dengan melihat penelitian sebelumnya yang berjudul “Penerapan Data Mining Untuk Menentukan Penerima Bantuan BLT Menggunakan Metode *Clustering K-Means* Pada Desa Pamulihan” [7], menghasilkan 9 *cluster* dengan status penerima bantuan, dalam pertimbangan, dan bukan penerima Dengan nilai DBI yang paling optimal sebesar 0,75. “Analisis Data Mining Menggunakan Metode Algoritma C4.5 Menentukan Penerima Bantuan Langsung Tunai” [8], menghasilkan model pohon keputusan yang mengelompokkan data warga dengan kategori menerima bantuan dan tidak menerima bantuan. “Implementasi Algoritma K-Means dan C4.5 Dalam Menentukan Tingkat Penyebaran COVID-19 Di Indonesia” [9] menghasilkan informasi mengenai pemetaan jumlah persebaran kasus Covid-19 di Indonesia berdasarkan zona hitam, zona merah, zona kuning, dan zona hijau. “Perbandingan Algoritma C4.5 Dan Naïve Bayes Untuk Menentukan Kelayakan Penerima Bantuan Program Keluarga Harapan” [10] menunjukkan bahwa algoritma C4.5 menghasilkan akurasi 91,25% sedangkan algoritma Naïve Bayes 87,11%. Oleh karena itu, algoritma C4.5 dianggap lebih unggul untuk diterapkan dalam menentukan kelayakan PKH. “Cluster Analysis and Classification of Emotions in K-Pop Music with K-Means and C 4.5 Algorithm” [11] membuktikan bahwa kombinasi metode K-Means dan C4.5 menghasilkan nilai akurasi tertinggi sebesar 0,95 dibandingkan metode clustering lainnya.

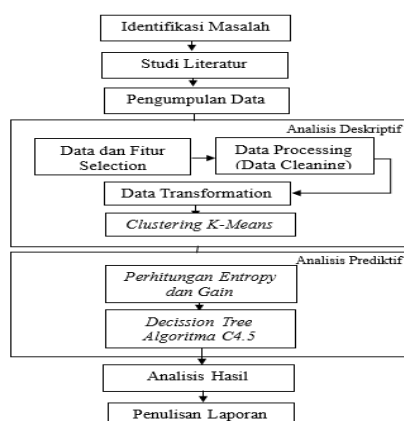
Dari beberapa penelitian di atas, penulis memilih *Data Mining* karena dapat menggali atau menghasilkan informasi dari data yang besar dan mengidentifikasi pola dari data yang tidak beraturan [12], sehingga bisa memberikan prediksi dalam pengambilan keputusan. *Clustering K-Means* dipilih karena dapat mengelompokkan data berdasarkan *cluster* dengan karakteristik yang sama [13]. Serta algoritma C4.5 dipilih karena bisa mengklasifikasikan data dan menggambarkannya dalam bentuk pohon keputusan [14] sehingga bentuk klasifikasi data mudah dipahami.

Beberapa studi telah memanfaatkan algoritma data mining untuk membantu klasifikasi data penerima bantuan, namun umumnya hanya menggunakan satu algoritma saja [15], tanpa adanya penggabungan pendekatan *clustering* dan pohon keputusan untuk memberikan gambaran secara detail kriteria apa yang paling berpengaruh dalam perhitungan prioritas penerima bantuan PKH.

Penelitian ini bertujuan untuk *mengcluster* data calon penerima bantuan PKH di Desa Talang Belido berdasarkan karakteristik yang telah ditentukan menggunakan metode *clustering k-means* serta membuat pohon keputusan (*Decision Tree*) menggunakan metode C4.5 berdasarkan hasil *clustering* yang telah dilakukan. Penelitian ini menggunakan enam atribut meliputi nama kepala keluarga, pekerjaan, jumlah anak atau tanggungan, pendapatan atau gaji per bulan, kondisi dan status kepemilikan rumah, serta status kesejahteraan penduduk untuk menentukan prioritas bantuan PKH pada Desa Talang Belido. Dimana penentuan atribut ini ditentukan berdasarkan kriteria penduduk miskin [16] yang memenuhi syarat untuk menerima bantuan PKH [17]. Dengan adanya penelitian ini diharapkan hasilnya dapat digunakan sebagai alat bantu untuk pengambilan keputusan terhadap penyaluran bantuan PKH secara adil dan lebih tepat sasaran.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian



Gambar 2. 1 Alur Kerja Penelitian

Alur penelitian di atas dirancang untuk memudahkan pencapaian hasil penelitian, memastikan penelitian selesai tepat waktu dan menjalankannya sesuai harapan. berikut penjelasan setiap langkah-langkah dalam penelitian tersebut:

1. Identifikasi Masalah

Mengidentifikasi serta merumuskan permasalahan pada penelitian untuk mengetahui masalah yang dialami Desa Talang Belido dalam memberikan bantuan Program Keluarga Harapan (PKH) kepada masyarakat agar bantuan tersebut bisa tepat sasaran.

2. Studi Literatur

Mempelajari buku referensi, artikel, dan hasil penelitian terkait yang relevan dengan masalah yang sedang diteliti. Studi literatur ini bertujuan untuk memperoleh landasan teoritis mengenai isu yang diteliti agar penulis dapat memahami permasalahan yang diteliti secara mendalam dan berdasarkan teori yang kuat

3. Pengumpulan Data

Peneliti mengumpulkan data penduduk Desa Talang Belido sesuai atribut yang dibutuhkan dalam penelitian. Proses pengumpulan data ini dilakukan dengan *observasi* dan wawancara kepada pihak desa dan pendamping PKH untuk mengetahui informasi terkait penerimaan bantuan PKH pada desa tersebut.

4. Analisis Deskriptif

proses yang digunakan untuk mendeskripsikan data, mengidentifikasi pola, tren, atau karakteristik data tanpa membuat kesimpulan atau prediksi di luar data yang dianalisis.

a. Data dan Fitur Selection

Menyeleksi fitur data dari kumpulan data yang diperoleh sebelum memulai tahap penggalian informasi dalam *data mining*.

b. Data Processing

Menyeleksi atribut yang berpengaruh dalam *Clustering* untuk menentukan *cluster* penduduk miskin pada Desa Talang Belido.

c. Data Transformation

Proses transformasi pada data yang telah dipilih dengan format yang sesuai, sehingga data tersebut sesuai untuk proses *data mining*.

d. Clustering K-Means

Mencari hasil dari perhitungan metode *Clustering K-Means*. dimana proses perhitungan ini dilakukan dengan perhitungan secara manual dan nantinya dibandingkan dengan hasil perhitungan menggunakan *tools rapidminer*.

5. Analisis Prediktif

Analisis prediktif adalah suatu proses menggunakan data untuk membuat model yang dapat memprediksi hasil di masa depan atau pada data baru.

a. Menentukan nilai *entropy* dan *gain*

Setelah proses *clustering* data dilakukan, maka kita harus menghitung nilai *entropy*. Semakin besar nilai *entropy* menunjukkan ketidakpastian yang besar [18]. Dengan nilai ini kita dapat mengevaluasi seberapa baik atribut memisahkan data ke dalam kelas yang berbeda. Selanjutnya dengan perhitungan *gain* untuk memilih atribut terbaik yang digunakan pada tiap *node* di pohon keputusan.

b. Decision Tree algoritma C4.5

Menggambarkan hasil perhitungan data dari nilai *entropy* dan *gain* menjadi suatu model pohon keputusan, sehingga mudah dimengerti karena ada tampilan visualnya untuk memprediksi data selanjutnya.

6. Analisis Hasil

Hasil dari analisis akan dinilai keakuratan data model yang sudah didapatkan dengan perhitungan dari gabungan dua metode dalam *data mining* yaitu algoritma *K-Means* untuk pengelompokan dan pengelolaan data penduduk miskin dan algoritma *C4.5* untuk prediksi hasil yang akan didapat dari proses *clustering* yang telah dilakukan.

7. Penulisan Laporan

Pada tahap penulisan laporan akhir penelitian, dilakukan berdasarkan kerangka yang telah dirancang yang terdiri dari pendahuluan, landasan teori, metodologi penelitian, analisis hasil, penutup, serta lampiran bukti hasil penelitian.

3. HASIL DAN PEMBAHASAN

3.1 Analisis Data Mining Metode *Clustering K-Means*

Penulis melakukan perhitungan menggunakan 4.775 data *training* yang kemudian diseleksi menjadi 1.357 data penduduk dengan menggunakan atribut meliputi nama kepala keluarga, pekerjaan, jumlah anak atau tanggungan, pendapatan atau gaji per bulan, kondisi dan status kepemilikan rumah, serta status kesejahteraan penduduk. Perhitungan ini akan menghasilkan 3 *cluster*, dimana *cluster* 1 akan dijadikan sebagai prioritas 1 (layak menerima bantuan), *cluster* 2 akan dijadikan prioritas 2 (dalam pertimbangan), dan *cluster* 3 akan dijadikan prioritas 3 (tidak layak menerima bantuan)

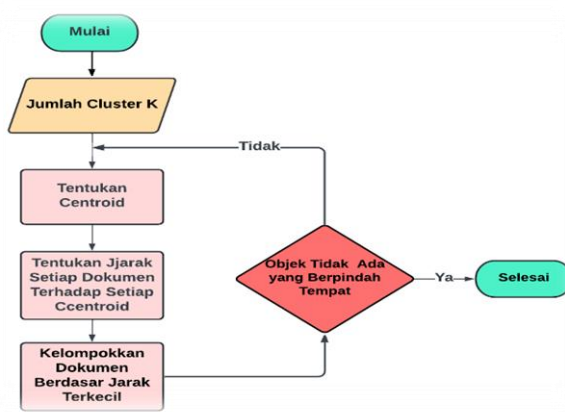
3.1.1 Data Transformation

Pada tahap ini format data akan disesuaikan untuk melakukan perhitungan, Penyesuaian format data diperlukan untuk mempermudah pengolahan nilai. Oleh karena itu, data yang masih berbentuk "karakter" harus diubah ke dalam format "numerik atau angka" agar dapat diolah.

Tabel 3. 1 Inisialisasi Nilai Data

1	JUMLAH TANGGUNGAN	
	0	TIDAK ADA ANAK
	1	1 ANAK
	2	2 ANAK
	3	3 ANAK
	4	>3 ANAK
2	KONDISI RUMAH	
	1	BAIK
	2	CUKUP
	3	KURANG BAIK
3	STATUS KEPEMILIKAN	
	1	HAK MILIK
	2	NUMPANG
	3	SEWA
4	JUMLAH PENGHASILAN	
	1	>4000000
	2	3100000-4000000
	3	2100000-3000000
	4	<2000000
5	STATUS KESEJAHTERAAN	
	1	BAIK
	2	CUKUP
	3	KURANG

3.1.2 Flowchart *Clustering K-Means*



Gambar 3. 1 Flowchart Clustering K-Means

3.1.3 Nilai Centroid Awal

Tabel 3. 2 Nilai Centroid Awal

Centroid 1	2	2	1	3	2
Centroid 2	1	2	1	3	1
Centroid 3	2	1	1	2	1

Penentuan nilai *centroid* awal ini ditentukan secara acak (*random*) oleh tools *Rapidminer*. Nilai pada *centroid 1* akan digunakan sebagai nilai *acuan* untuk *cluster 1*, Nilai pada *centroid 2* akan digunakan sebagai nilai *acuan* untuk *cluster 2*, dan Nilai pada *centroid 3* akan digunakan sebagai nilai *acuan* untuk *cluster 3*.

3.1.4 Perhitungan Euclidian Distance, BCV, WCV, Dan Ratio

a. Euclidian Distance

$$\text{Euclidian Distance} = d(x, y) = |x - y| = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \dots \dots \dots (3.1)$$

b. BCV (Between Cluster Variation)

$$D(M1, M2) = \sqrt{(2 - 1)^2 + (2 - 2)^2 + (1 - 1)^2 + (3 - 3)^2 + (2 - 1)^2} = 1,4142$$

$$D(M1, M3) = \sqrt{(2 - 2)^2 + (2 - 1)^2 + (1 - 1)^2 + (3 - 2)^2 + (2 - 1)^2} = 1,7321$$

$$D(M2, M2) = \sqrt{(1 - 2)^2 + (2 - 1)^2 + (1 - 1)^2 + (3 - 2)^2 + (1 - 1)^2} = 1,7321$$

$$\begin{aligned} \text{Jadi BCV} &= D(M1, M2) + D(M1, M3) + D(M2, M3) \\ &= 1,4142 + 1,7321 + 1,7321 = 4,8783 \end{aligned}$$

c. WCV (Within Cluster Variation)

$$\text{WCV} = \text{Total min } \{d(X_i, Y_i)^2\}$$

d. Ratio

$$\text{Ratio} = \text{BCV/WCV}$$

3.1.5 Nilai Centroid Final

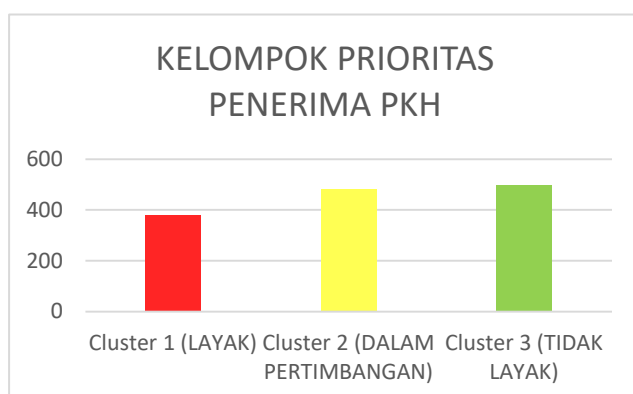
Tabel 3. 3 Nilai Centroid Final

Centroid 1	2,337	2,008	1,101	2,997	2,406
Centroid 2	0,707	1,988	1,075	2,992	1,309
Centroid 3	1,649	1,114	1,000	1,637	1,026

3.1.6 Perbandingan Nilai BCV, WCV, dan Ratio Iterasi

Nilai	Iterasi 1	Iterasi 2	Iterasi 3	Iterasi 4	Iterasi 5
BCV	4,8783	5,8722	6,0004	6,1000	6.0996
WCV	2943	1597,8955	1543,8117	1523,7298	1523.0203
Ratio	0,0017	0,0037	0,0039	0,0040	0.0040

Proses perhitungan *Clustering K-Means* ini berhenti pada iterasi ke-5 dikarenakan nilai *ratio* yang diperoleh telah lebih kecil atau sama dengan iterasi sebelumnya, serta tidak ada lagi data yang berpindah tempat.



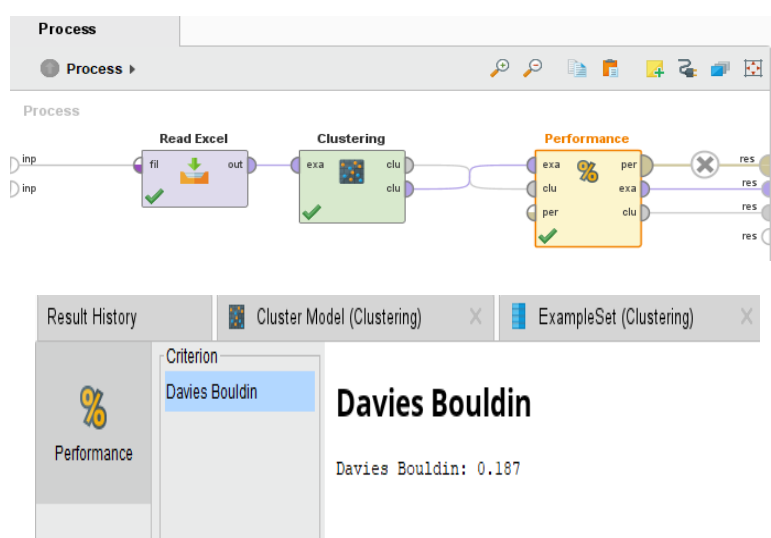
3.1.7 Hasil Akhir Perhitungan *Clustering K-Means*

Keterangan :

- Cluster 1 sebagai prioritas yang layak mendapat bantuan PKH sebanyak 377 orang.
- Cluster 2 sebagai bahan pertimbangan dalam menerima bantuan PKH sebanyak 482 orang
- Cluster 3 sebagai yang tidak layak menerima bantuan PKH sebanyak 498 orang.

3.1.8 Hasil Akurasi *Clustering K-Means*

Pada metode *Clustering K-Means*, untuk melihat hasil nilai akurasi atau hasil kualitas data yang dihasilkan bisa dilakukan dengan menambahkan *operator performance cluster distance* pada *column process*.



Gambar 3. 2 Nilai DBI

Pada *tools rapidminer* hasil akurasi dapat dilihat berdasarkan nilai *DBI (Davies Bouldin Index)*. Berdasarkan ketentuan *DBI* yang memiliki rentang nilai (0-1), semakin nilai yang dihasilkan mendekati 0, maka kualitas data yang dihasilkan semakin baik [19].

Dari proses yang telah penulis lakukan berdasarkan **Gambar 3.2**, didapatkanlah nilai *DBI* dari data yang telah diolah sebesar 0,187 atau jika dalam persen setara 81,3%. Untuk nilai *silhouette score*

menggunakan *python* mendapatkan nilai akurasi sebesar 0,75 Hal ini membuktikan bahwa hasil *cluster* data yang dilakukan bernilai baik.

3.1.9 Analisis Data Mining Metode C4.5

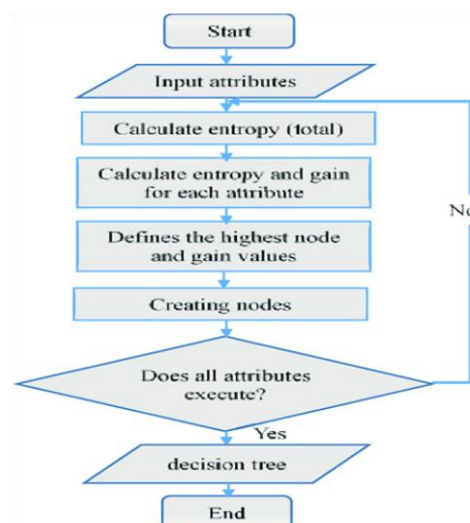
Setelah proses *Clustering K-Means* telah selesai, penulis akan melanjutkan perhitungan data menggunakan metode *C4.5* untuk memperoleh model pohon Keputusan (*Decision Tree*). Data yang digunakan dalam perhitungan ini adalah hasil data dari proses *clustering K-Means* yang telah dilakukan.

JUMLAH TANGGUNG	KONDISI RUMAH	STATUS PEMILIK	JUMLAH PENGHASILAN	STATUS KESEJAHTERAAN	KRITERIA
2	1	1	1	1	TIDAK LAYAK
0	2	1	3	3	DALAM PERTIMBANGAN
2	2	1	3	2	LAYAK
2	1	1	1	1	TIDAK LAYAK
0	1	1	1	1	TIDAK LAYAK
4	1	1	2	1	TIDAK LAYAK
4	1	1	1	1	TIDAK LAYAK
2	1	1	2	1	TIDAK LAYAK
1	2	1	3	2	DALAM PERTIMBANGAN
1	1	1	2	1	TIDAK LAYAK
1	2	1	3	1	DALAM PERTIMBANGAN
2	2	1	3	2	LAYAK
4	2	1	3	3	LAYAK
2	2	1	3	2	LAYAK
0	2	1	3	2	DALAM PERTIMBANGAN
2	2	1	3	2	LAYAK
1	1	1	1	1	TIDAK LAYAK
4	2	1	2	2	LAYAK
1	3	1	3	3	LAYAK
0	2	1	3	2	DALAM PERTIMBANGAN
3	2	1	3	2	LAYAK

Gambar 3. 3 Data Training C4.5

Beberapa data yang ditampilkan di atas merupakan bagian dari total 1.357 data yang akan digunakan dalam analisis menggunakan metode *C4.5*. Data tersebut akan menjadi acuan dalam pembuatan pohon keputusan (*Decision Tree*).

3.1.10 Flowchart Algoritma C4.5



Gambar 3. 4 Flowchart Algoritma C4.5

3.2 Perhitungan Entropy dan Gain

Tahap pertama yang perlu dilakukan adalah menghitung nilai *entropy* total dari data *training* yang digunakan. setelah itu, dilanjutkan dengan perhitungan nilai *entropy* dan *gain* dari masing-masing atribut yang

digunakan. Adapun rumus yang digunakan untuk menghitung nilai *entropy* dan *gain* dapat dilihat sebagai berikut:

a. **Entropy Total**

Jumlah Kasus (S) = 1357

Jumlah Kasus Layak (S1) = 377

Jumlah Kasus Dalam Pertimbangan (S2) = 482

Jumlah Kasus Tidak Layak (S3) = 498

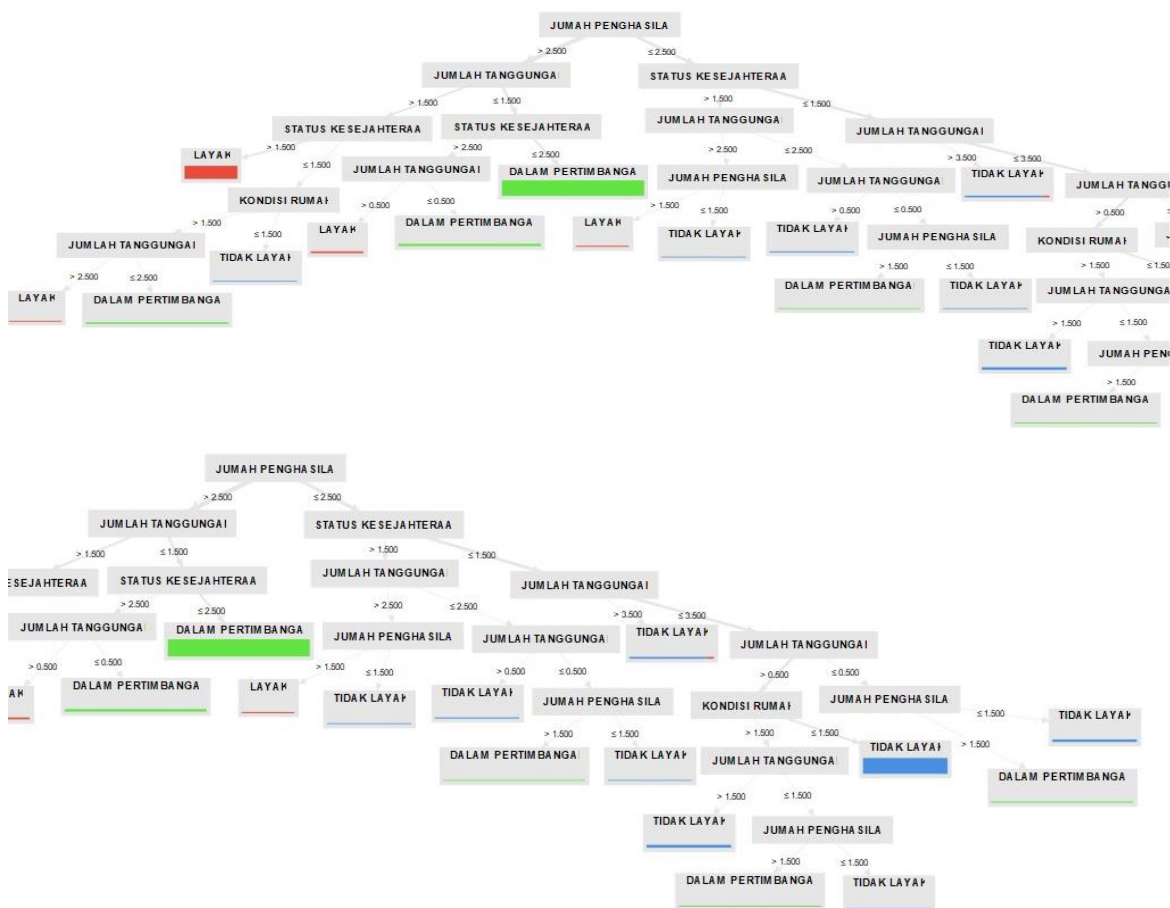
$$Entropy\ Total\ (S) = \sum_{i=1}^n -P_i \times \log_2 P_i \dots\dots\dots(3.2)$$

$$= ((-377/1357) * \log_2 (377/1357)) + ((-482/1357) * \log_2 (482/1357)) + ((-498/1357) * \log_2 (498/1357)) \\ = 1,5745$$

b. **Gain**

$$Entropy(S) - \sum_{i=1}^n \left| \frac{S_i}{S} \right| \times Entropy(S_i) \dots\dots\dots(3.3)$$

3.3 Hasil Akhir Pohon Keputusan (*Decision Tree*)



Gambar 3. 5 Hasil Pohon Keputusan Algoritma C4.5

Permodelan pohon keputusan di atas dibuat menggunakan *tools Rapidminer*. Berdasarkan hasil pohon keputusan di atas, atribut yang paling menentukan dan menjadi akar (*root*) adalah atribut “Jumlah Penghasilan” dengan nilai *split* setiap *node* atribut ditentukan berdasarkan nilai *weight*. Nilai akurasi yang dihasilkan dari model *decision tree* di atas adalah sebesar 99,71%.

3.4 Nilai Akurasi C4.5

(Apply Model) × PerformanceVector (Performance) ×

☒ Table View ☐ Plot View

accuracy: 99.71%

	true TIDAK LAYAK	true DALAM PERTIMBANGAN	true LAYAK	class precision
pred. TIDAK LAYAK	496	0	2	99.60%
pred. DALAM PERTIMBANGAN	2	482	0	99.59%
pred. LAYAK	0	0	375	100.00%
class recall	99.60%	100.00%	99.47%	

Gambar 3. 6 Nilai Akurasi C4.5

Hasil nilai akurasi pada proses perhitungan data menggunakan metode C4.5 yang dihasilkan menggunakan tools *rapidminer* adalah sebesar 99,71%. Dimana nilai precision dari kategori "Tidak Layak" = 99,60%, "Dalam Pertimbangan" = 99,59%, dan "Layak" = 100%. Sedangkan nilai *recall* dari kategori "Tidak Layak" = 99,60%, "Dalam Pertimbangan" = 100%, dan "Layak" = 99,47%. Hal ini membuktikan hasil data yang dihasilkan sangat baik.

4. KESIMPULAN

Berdasarkan hasil penelitian terhadap data penduduk Desa Talang Belido pada tahun 2023 hingga bulan Desember 2024, yang mencakup 1357 kepala keluarga, dapat disimpulkan hasil penelitian ini berdasarkan perhitungan *Clustering K-Means* melibatkan 5 iterasi dan menghasilkan 3 kelompok prioritas penerima bantuan Program Keluarga Harapan (PKH) yaitu *Cluster 1* sebagai prioritas 1 (Layak) terdapat 377 data kepala keluarga, *Cluster 2* sebagai prioritas 2 (Dalam Pertimbangan) terdapat 482 data kepala keluarga, serta *Cluster 3* sebagai prioritas 3 (Tidak Layak) terdapat 498 data kepala keluarga. Hasil dari perhitungan algoritma C4.5 membuktikan urutan atribut yang paling berpengaruh dalam pembuatan hasil pohon keputusan adalah jumlah penghasilan, jumlah tanggungan, status kesejahteraan, dan yang terakhir kondisi rumah. Nilai akurasi hasil perhitungan dengan metode *Clustering K-means* yang dihitung berdasarkan nilai *DBI (Davies-Bouldin Index)* sebesar 0,187 atau setara 81,3% serta *silhouette score* sebesar 0,75. Sedangkan nilai akurasi dari permodelan pohon keputusan (*Decision Tree*) dengan algoritma C4.5 menghasilkan nilai akurasi sebesar 99,71%

REFERENCES

- [1] Badan Pusat Statistik Rembang, "Data dan Informasi Kemiskinan Kabupaten Rembang 2017-2021." [Online]. Available: <https://rembangkab.bps.go.id/id/publication/2022/12/06/467c50362bd6b45fad1ffcda/data-dan-informasi-kemiskinan-kabupaten-rembang-2017-2021.html>
- [2] M. Amali and A. Devita, "Faktor-Faktor yang Mempengaruhi Kemiskinan di Provinsi Jambi," *J-MAS (Jurnal Manaj. dan Sains)*, vol. 7, no. 2, p. 1247, 2022, doi: 10.33087/jmas.v7i2.537.
- [3] Badan Pusat Statistik, "Persentase Penduduk Miskin Maret 2024 turun menjadi 9,03 persen," 2024, [Online]. Available: <https://www.bps.go.id/id/pressrelease/2024/07/01/2370/persentase-penduduk-miskin-maret-2024-turun-menjadi-9-03-persen-.html>
- [4] M. Effendy, "Muhadjir Sebut Target Penurunan Kemiskinan 7,5 Persen Sulit Dicapai." [Online]. Available: <https://www.cnnindonesia.com/nasional/20240222173830-32-1066137/muhadjir-sebut-target-penurunan-kemiskinan-75-persen-sulit-dicapai>
- [5] N. Aminudin, I. Ayu, and P. Sari, "Sistem Pendukung Keputusan (Dss) Penerima Bantuaprogram Keluarga Harapan (Pkh) Pada Desa Bangun Rejo Kec.Punduh Pidada Pesawaran Dengan Menggunakan Metode Analytical Hierarcy Process (Ahp)," *J. TAM (Technol. Accept. Model)*, vol. 5, no. 2, pp. 66–72, 2015.
- [6] K. K. PMK, "Temukan Bansos Tidak Merata di Kantong Kemiskinan, Menko PMK: Data Lapangan Harus Disempurnakan." [Online]. Available: <https://www.kemenkopmk.go.id/temukan-bansos-tidak-merata-di-kantong-kemiskinan-menko-pmk-data-lapangan-harus-disempurnakan>
- [7] E. Dwiguna and A. Bahtiar, "Penerapan Data Mining Untuk Menentukan Penerima Bantuan Blt Menggunakan Metode Clustering K-Means Pada Desa Pamulihan," *JATI (Jurnal Mhs. Tek. Inform.,* vol. 8, no. 2, pp. 1382–1388, 2024, doi: 10.36040/jati.v8i2.9029.

- [8] L. Bachtiar and M. Mahradianur, "Analisis Data Mining Menggunakan Metode Algoritma C4.5 Menentukan Penerima Bantuan Langsung Tunai," *J. Inform.*, vol. 10, no. 1, pp. 28–36, 2023, doi: 10.31294/inf.v10i1.15115.
- [9] F. Salsabila and S. Maulida Intani, "Implementasi Algoritma K-Means dan C4.5 Dalam Menentukan Tingkat Penyebaran Covid-19 di Indonesia," *J. Siliwangi*, vol. 7, no. 1, pp. 25–30, 2021.
- [10] E. Fitriani, "Perbandingan Algoritma C4.5 Dan Naïve Bayes Untuk Menentukan Kelayakan Penerima Bantuan Program Keluarga Harapan," *Sistemasi*, vol. 9, no. 1, p. 103, 2020, doi: 10.32520/stmsi.v9i1.596.
- [11] Reyhan Jarsi Yoga, Basuki Rahmat, and Eka Prakarsa Mandyartha, "Cluster Analysis and Classification of Emotions in K-Pop Music with K-Means and C 4.5 Algorithm," *Neptunus J. Ilmu Komput. Dan Teknol. Inf.*, vol. 2, no. 3, pp. 147–160, 2024, [Online]. Available: <https://journal.artei.or.id/index.php/Neptunus/article/view/228>
- [12] W. Laturrizqi and O. Nurdiawan, "Implementasi Algoritma C4.5 Untuk Menentukan Penerima Bantuan Sosial," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 1, pp. 373–377, 2023, doi: 10.36040/jati.v7i1.6312.
- [13] R. Handoyo, R. Mangkudjaja, and S. M. Nasution, "Perbandingan Metode Clustering Menggunakan Metode Single Linkage dan K - Means pada Pengelompokan Dokumen," *J. SIFO Mikroskil*, vol. 15, no. 2, pp. 73–82, 2014, doi: 10.55601/jsm.v15i2.161.
- [14] Y. Indah, "Prediksi Tingkat Kepuasan Pelayanan Online Menggunakan Metode Algoritma C4.5," *J. Inform. Ekon. Bisnis*, vol. 4, pp. 59–64, 2022, doi: 10.37034/inf.v4i2.99.
- [15] D. K. Utami, N. Irawati, and S. Sumantri, "Analysis of the k-Means Method in Clustering Acceptance of PKH Aid in Pulau Rakyat Tua Village," *Sistemasi*, vol. 12, no. 3, p. 953, 2023, doi: 10.32520/stmsi.v12i3.3236.
- [16] Badan Pusat Statistik Ogan Ilir, *Indikator Kemiskinan Kabupaten Ogan Ilir*. 2022.
- [17] D. J. Guru, T. Kependidikan, and K. Pendidikan, *Pedoman pelaksanaan Program Keluarga Harapan*. 2021.
- [18] A. Essra, Rahmadani, and Safriadi, "Analisis Information Gain Attribute Evaluation Untuk Klasifikasi Serangan Intrusi," *J. Inf. Syst. Dev.*, vol. 2, no. 2, pp. 9–14, 2016.
- [19] S. D. Raihannabil, "Penerapan Metode Hierarchical Clustering untuk Klasterisasi Provinsi di Indonesia berdasarkan Indikator Status Gizi Anak Baduta (Bawah Dua Tahun) Tahun 2023," vol. 2, no. 3, pp. 424–436, 2024.