

PENERAPAN ALGORITMA K-MEANS CLUSTERING UNTUK IDENTIFIKASI KELAYAKAN PENERIMA BANTUAN PROGRAM KELUARGA HARAPAN (PKH) DI DESA TAMBAKSARI CIAMIS

Fitriani, Rudi Kurniawan, Tati Suprpti

Program Studi Teknik Informatika S1, STMIK IKMI Cirebon
STMIK IKMI Cirebon, Jalan Perjuangan No. 10B Majasem, Kota Cirebon, Indonesia
fitrianifitriani0702@gmail.com

ABSTRAK

Program bantuan sosial telah menjadi salah satu upaya yang signifikan untuk menurunkan tingkat kemiskinan di Indonesia yang bertujuan untuk meningkatkan kesejahteraan masyarakat. Akan tetapi bantuan sosial PKH masih kurang tepat sasaran dan juga tidak sesuai dengan harapan masyarakat yang disebabkan status rakyat miskin yang tidak maksimal. Penelitian ini bertujuan untuk mengidentifikasi kelayakan penerima bantuan sosial Program Keluarga Harapan (PKH) di Desa Tambaksari Ciamis dengan menggunakan metodologi *Knowledge Discovery in Database* (KDD). Metode yang digunakan dalam penelitian ini adalah algoritma K-Means Clustering. Algoritma ini merupakan salah satu metode dalam pengelompokan data yang efektif untuk sistem pendukung pengambilan keputusan, terutama pada data yang berukuran besar. Hasil dari penelitian ini dengan menggunakan algoritma K-Means adalah adanya *cluster* data penerima PKH dengan mendapatkan hasil *Davies Bouldin Index* (DBI) sebesar 0,415 dari parameter *measure type* jenis *Numerical Measure* dengan *Euclidean Distance*. Penelitian ini menghasilkan 2 kelompok dari 211 data. *Cluster* 0 dengan jumlah 190 termasuk layak, *cluster* 1 dengan jumlah 21 termasuk tidak layak. Hasil DBI tersebut dinilai cukup baik sebab semakin dekat hasil yang diperoleh dengan angka nol maka semakin baik.

Kata kunci: PKH, K-Means Clustering, DBI, Desa Tambaksari Ciamis

1. PENDAHULUAN

Di Indonesia masalah kemiskinan adalah masalah yang sulit untuk diselesaikan [1]. Pemerintah sendiri sudah melakukan beberapa usaha untuk melakukan pengentasan kemiskinan yakni melalui beberapa program bantuan sosial, diantaranya program keluarga harapan, program bantuan langsung tunai dan lain sebagainya [2]. Pada era pemerintahan Indonesia saat ini, kemiskinan merupakan salah satu masalah penting yang menjadi perhatian pemerintah negara manapun saat ini di Indonesia [3]. Permasalahan kemiskinan ini terjadi di berbagai pelosok negara, tidak terkecuali di pedesaan. Masalah kemiskinan ini harus segera diselesaikan agar angka kemiskinan menurun. Langkah yang diambil oleh pemerintah adalah menerapkan kebijakan terkait kepada keluarga yang kurang mampu. Dalam konteks ini, pemerintah mewujudkan kebijakan tersebut melalui pelaksanaan Program Keluarga Harapan (PKH). Dalam pemilihan komponen penerima Program Keluarga Harapan (PKH) tentu sering terjadi permasalahan dalam mencari data warga. Permasalahan yang dihadapi adalah kriteria warga yang mendapatkan Program Keluarga Harapan (PKH) tidak sesuai dengan ketentuan pemerintah. Di desa Tambaksari Ciamis permasalahan kemiskinan menjadi permasalahan yang cukup tinggi, maka perlu adanya program pemberdayaan yang efektif serta adil bagi semua masyarakat. Sesuai dengan kebijakan yang dikeluarkan oleh pemerintah pusat yaitu melalui pelaksanaan Program Keluarga Harapan (PKH). Namun faktanya untuk penyaluran PKH ini belum

tepat sasaran dan belum tepat jumlah penerima bantuan PKH. Hal ini menimbulkan kecemburuan sosial antar warga di kalangan lingkungan masyarakat tersebut. Maka, penelitian ini dibutuhkan untuk mengidentifikasi kelayakan penerima Program Keluarga Harapan (PKH) di Desa Tambaksari Ciamis.

Penelitian yang dilakukan oleh Febriansyah Febriansyah dan Siti Muntari yang berjudul "Penerapan Algoritma K-Means untuk Klasterisasi Penduduk Miskin pada Kota Pagar Alam" tujuan penelitian ini adalah untuk mengatasi masalah pengumpulan data yang tidak akurat dan distribusi bantuan kepada orang miskin melalui Program Keluarga Harapan (PKH). Hasil dari penelitian ini adalah pengembangan sistem berbasis algoritma K-Means Clustering untuk mengklasifikasikan penduduk miskin di Kota Pagar Alam. Penelitian ini juga menghasilkan tiga *cluster* data kemiskinan, yaitu *cluster_0* dengan 156 orang, *cluster_1* dengan 82 orang, dan *cluster_2* dengan 233 orang [4].

Tujuan utama dari penelitian ini adalah melakukan proses pengelompokan data (*Clustering*) serta analisis mendalam terhadap informasi penerima bantuan PKH Desa Tambaksari Ciamis. Pendekatan yang digunakan dalam penelitian ini adalah dengan menerapkan algoritma K-Means untuk mengelompokkan data berdasarkan pola-pola yang ada. Dalam algoritma K-Means ini digunakan indikator *Davies Bouldin Index* (DBI) untuk menentukan jumlah *cluster* terbaik [5]. Selain itu, penelitian ini juga bertujuan untuk melakukan evaluasi terhadap kelayakan atau tidak layaknya populasi

penerima bantuan PKH untuk mendapatkan manfaat dari program PKH yang ada.

Untuk menunjang serta membantu keberhasilan dalam mengidentifikasi kelayakan penerima PKH dibutuhkan suatu kajian yang dapat mengelompokkan atribut-atribut yang ada dalam komponen keluarga penerima PKH sehingga akan terbentuknya kelompok yang sesuai dengan tingkat kesejahteraan sosial. Pengelompokkan karakteristik-karakteristik komponen ini akan menjadi panduan dalam memberikan bantuan PKH sehingga dapat dilakukan dengan lebih terfokus dan sesuai dengan kelompok yang membutuhkan berdasarkan tingkat kesejahteraan sosial. Salah satu metode yang digunakan untuk melakukan pengelompokkan atribut ini yaitu dengan menggunakan analisis *Clustering*. Pada penelitian ini akan menggunakan *Clustering* non hirarki yaitu metode pendekatan algoritma K-means. Algoritma K-means ini digunakan karena cepat diaplikasikan serta dapat mengelompokkan atribut-atribut berdasarkan kesamaan karakteristik diantara atribut-atribut tersebut [6]. Dan akan diklasifikasikan ke dalam satu atau lebih *cluster* sehingga sifat-sifatnya akan sebanding satu sama lain. Untuk menentukan apakah kelompok penerima PKH layak mendapatkan manfaat dari program ini.

2. TINJAUAN PUSTAKA

2.1. Data Mining

Data mining adalah serangkaian prosedur yang digunakan untuk mengekstraksi nilai tambahan, terutama informasi yang sebelumnya tidak dikenali secara manual, berasal dari suatu dataset. Metode yang diterapkan untuk menghasilkan informasi ini adalah dengan mengidentifikasi pola penting dari data yang telah tersimpan dalam basis data atau diambil dari basis data yang telah tersedia. Data mining ialah suatu proses untuk mengekstrak informasi bermanfaat dari gudang basis data yang besar, yang disebut Knowledge Discovery Databases (KDD) [7].

2.2. K-Means

K-Means merupakan salah satu metode pengelompokan data non-hirarki (*Clustering*) yang dapat mengorganisir data ke dalam dua kelompok atau lebih. [8]. Algoritma K-Means terkenal karena kemudahan penggunaannya serta kemampuan untuk mengelompokkan data yang besar dan mengidentifikasi data outlier dengan cepat [9].

2.3. Clustering

Metode pengelompokan data yang dikenal sebagai *Clustering* merupakan mengelompokkan data ke dalam beberapa *cluster* atau kelompok sehingga data dalam satu *cluster* memiliki kesamaan yang tinggi, sementara kesamaan data antar *cluster* memiliki kemiripan yang rendah [10].

2.4. Program Keluarga Harapan (PKH)

PKH merupakan salah satu inisiatif bantuan dari pemerintah yang melibatkan semua anggota keluarga, kepala rumah tangga harus memenuhi kualifikasi tertentu yang akan dipilih untuk mendapatkan bantuan tersebut, kualifikasi atau kriteria bagi rumah tangga yang memenuhi syarat untuk menerima bantuan tersebut [11].

2.5. RapidMiner

Perangkat lunak RapidMiner bersifat terbuka. RapidMiner sebuah solusi untuk analisis data mining, text mining, dan prediksi. RapidMiner merupakan software mandiri yang dirancang untuk analisis data dan data mining yang dapat diintegrasikan pada produknya sendiri. Ditulis dalam bahasa Java, sehingga dapat bekerja di semua sistem operasi [12].

2.6. Davies Bouldin Index (DBI)

David L. Davies dan Donald W. Bouldin menerbitkan sebuah teknik dan menamakannya *Davies Bouldin Index* (DBI) yang dipergunakan untuk menilai *cluster*. Skor Index Davies-Bouldin memiliki skema penilaian *cluster* internal yang menentukan skor *cluster* baik atau buruk berdasarkan jumlah skor *cluster* dan kedekatan antar skor *cluster*. Semakin kecil nilai DBI, semakin baik kekompakan atau pemisahannya, sebaliknya untuk nilai DBI yang besar [13].

3. METODE PENELITIAN

3.1. Sumber Data

Penelitian ini menggunakan sumber data sekunder. Data yang digunakan dalam penelitian ini berasal dari data penerima bantuan PKH yang diperoleh dari pemerintah Desa Tambaksari. Data yang diperoleh dari pihak pelaksana PKH. Data yang diperoleh sebanyak 211 data.

3.2. Teknik Pengumpulan Data

Teknik pengumpulan data yang digunakan peneliti untuk mendapatkan data dalam penelitian ini yaitu gabungan antara penelitian pustaka dan penelitian lapangan. Observasi secara langsung di lokasi penelitian guna memperoleh informasi data yang sesuai dan lebih akurat mengenai hal-hal yang terkait dengan Program Keluarga Harapan (PKH) di Desa Tambaksari Ciamis. Teknik observasi ini digunakan untuk memperhatikan dan mengamati tentang bagaimana dampak bagi kehidupan penerima manfaat sebelum dan sesudah penerima PKH di Desa Tambaksari mendapatkan bantuan. Teknik dokumentasi ini dilakukan untuk pengumpulan dokumen dokumen yang berkaitan dengan pelaksanaan PKH di Desa Tambaksari Tambaksari. Dokumen dokumen tersebut mencakup berbagai aspek, seperti daftar lengkap penerima manfaat yang tercatat dalam program, data yang menggambarkan alokasi bantuan

3.3. Teknik Analisis Data

Pada penelitian ini metode yang digunakan adalah metode *Knowledge Discovery Process* (KDD) sebagai proses dimana teknik data mining digunakan guna mengidentifikasi informasi serta pola yang relevan dalam sekumpulan data, proses data mining ini melibatkan penggunaan algoritma untuk menemukan pola yang dapat mengidentifikasi dalam suatu proses database. Berikut adalah langkah dalam proses *Knowledge Discovery in Database Process* (KDD):

a. Data Selection

Dalam tahap selection, data diidentifikasi dari sumber yang relevan. Atribut yang digunakan meliputi data penerima bantuan seperti nama, alamat, RT/RW, status tempat tinggal, pendapatan, luas lahan, jenis lantai, dinding, fasilitas memasak, MCK, air, dan pembuangan air besar. Data diambil dari 211 penerima bantuan di Desa Tambaksari, Ciamis.

b. Preprocessing

Dalam tahap preprocessing, dilakukan pembersihan data, penghapusan data tidak relevan, penanganan data kosong, dan penggabungan data dari berbagai sumber. Data kategorik dikonversi ke bentuk yang dapat diolah oleh algoritma K-Means *Clustering*.

c. Data Transformation

Dalam preprocessing, fitur baru diekstraksi untuk meningkatkan performa *Clustering*. Subset fitur paling relevan dipilih. Dilakukan transformasi subdataset ke format seragam untuk menyederhanakan proses data mining. Atribut dinamakan ulang dengan kode untuk memudahkan pengolahan data. 14 atribut diolah di RapidMiner, 9 atribut dikonversi ke numerik menggunakan *Nominal to Numerical*.

d. Data Mining

Dalam penentuan jumlah K yang optimal untuk K-Means, digunakan *Davies Bouldin Index* (DBI). K-Means *Clustering* diterapkan pada data hasil preprocessing menggunakan RapidMiner. Dilakukan analisis hasil *Clustering* untuk mengidentifikasi kelompok penerima PKH. Nilai K dioptimalkan dengan *Cluster Distance Performance* dan DBI. Nilai K yang dicoba: 2, 3, 4, 5, berdasarkan *Numerical Measure* jenis *Euclidean Distance* dengan max run 5, 10, 15, 20, 25.

e. Evaluasi dan Interpretasi hasil

Hasil *Clustering* dievaluasi dengan menganalisis karakteristik tiap *cluster* seperti profil kelayakan penerima PKH. Hasil *Clustering* diinterpretasikan menyimpulkan kelayakan penerima bantuan PKH di Desa Tambaksari Ciamis dan memberikan rekomendasi yang sesuai.

4. HASIL DAN PEMBAHASAN

Implementasi algoritma K-Means *Clustering* dirancang menggunakan tools RapidMiner V10.0. pada penelitian ini digunakan tahapan *Knowledge*

Discovery in Database (KDD) yang terdiri dari beberapa tahap, yaitu sebagai berikut :

4.1. Data

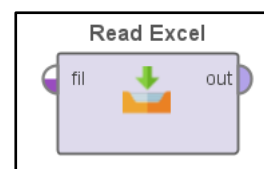
Data berasal dari file excel "dataset PKH Tambaksari". Sumber data adalah BNBA penyaluran bansos PKH tahap 3 tahun 2022 dan data sensus penduduk Desa Tambaksari 2021. Total 211 data penerima PKH Desa Tambaksari dengan 14 atribut terkait penentuan kelayakan penerima PKH. Data atribut penerima PKH dapat dilihat pada Tabel 1 sebagai berikut :

Tabel 1. Atribut Data

No	Atribut	Type Data
1	Nomor	Integer
2	Nama penerima	Polynomial
3	Alamat	Polynomial
4	RT	Integer
5	RW	Integer
6	Status tempat tinggal	Polynomial
7	Pendapatan	Integer
8	Luas lahan	Integer
9	Jenis lantai	Polynomial
10	Jenis dinding	polynomial
11	Energiuntuk memasak	polynomial
12	Fasilitas MCK	Polynomial
13	Sumber air	polynomial
14	Fasilitas BAB	polynomial

4.2. Data Selection

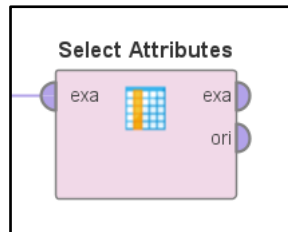
Pada tahap seleksi data, dilakukan pemilihan data BNBA penyaluran bansos PKH tahap 3 tahun 2022 dan data sensus penduduk Desa Tambaksari 2021 untuk implementasi K-Means *Clustering* guna mengidentifikasi kelayakan penerima PKH. Proses seleksi berdasarkan atribut pengukur tingkat kemiskinan. Dihasilkan satu dataset terdiri dari 14 atribut dan 211 data penerima PKH. Agar dataset penerima PKH desa Tambaksari ada di repository RapidMiner, maka perlu dilakukan import data. Operator yang digunakan untuk import data adalah Read Excel. Operator *read excel* digunakan untuk membaca file berbentuk excel. Proses pemodelan data dapat dilihat pada Gambar 1 di bawah ini.



Gambar 1. Operator Read Excel dari RapidMiner

Attributes filter type diubah ke subset agar atribut yang tidak digunakan dapat dihapus. Dari 14 atribut data penerima PKH, yang dihapus adalah 4 yaitu alamat, no, RT, RW yang digunakan adalah 10 jenis dinding, energi memasak, fasilitas BAB, MCK, jenis lantai, luas lahan, nama penerima, pendapatan, status

tempat tinggal, sumber air. Operator *Select Attributes* dapat dilihat pada Gambar 2 di bawah ini.



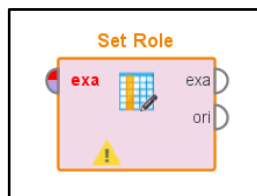
Gambar 2. Operator *Select Attributes* dari RapidMiner

Pada operator *Select Attributes* terdapat parameter yang perlu disesuaikan. *Select Attributes* yang berfungsi untuk memilih subset atribut dari *ExampleSet* dan menghapus atribut lainnya yang tidak akan digunakan sebagai pengukur tingkat kemiskinan yang akan digunakan sebagai alat ukur kelayakan penerima PKH. Parameter dalam *Select Attributes* dapat dilihat pada Tabel 2 di bawah ini.

Tabel 2. Parameter *Select Attributes*

No.	Parameter	Isi
1.	Attribute filter type	Subset
2.	Attributes	Select Attributes ...
		Nama penerima
		Pendapatan
		Luas lahan
		Jenis lantai
		Jenis dinding
		Energiuntuk memasak
		Status tempat tinggal
		Fasilitas MCK
		Sumber air
		Fasilitas BAB

Setelah atribut telah disesuaikan dengan apa yang diperlukan langkah selanjutnya adalah menentukan atribut mana yang akan dijadikan id pada dataset. Dalam menentukan id digunakan operator *Set Role*. Pemodelan operator *Set Role* dapat dilihat pada Gambar 3 di bawah ini.



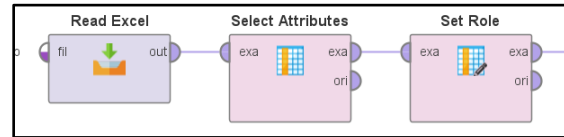
Gambar 3. Operator *Set Role* dari RapidMiner

Untuk mengubah satu atribut menjadi id diperlukan parameter pada operator *Set Role*, *attribute name* diisi dengan Nama Penerima dan *target role* diisi dengan id. Parameter *Set Role* dapat dilihat pada Tabel 3 di bawah ini.

Tabel 3. Parameter *Set Role*

No.	Parameter	Isi
1.	Attribute name	Nama penerima
2.	Target role	id

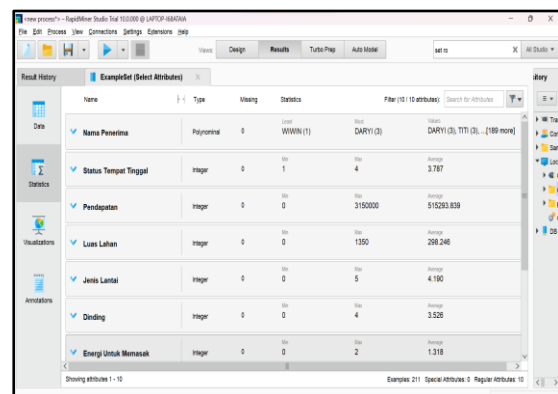
Proses pemodelan pada tahap selection di RapidMiner, dapat dilihat pada Gambar 4 di bawah ini.



Gambar 4. Proses Pemodelan Tahap *Selection Data*

4.3. Preprocessing

Preprocessing data dilakukan untuk memastikan akurasi analisis. Dataset penerima PKH dianalisis untuk menghilangkan data tidak lengkap atau tidak relevan. Dilakukan normalisasi data untuk menyamakan skala dan unit pengukuran atribut, agar setiap atribut berkontribusi seimbang. Data diproses dengan *Select Attributes* untuk memfilter berdasarkan kondisi tertentu, misalnya menghapus data yang hilang atau tidak valid. Setelah *Select Attributes*, tidak ada *missing attributes* pada dataset. Proses pengecekan *missing* dapat dilihat pada Gambar 5 dibawah ini.

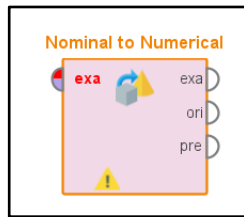


Gambar 5. Pengecekan *Missing Value* pada Data Penerima PKH Tambaksari Ciamis

Data penerima PKH yang diolah menggunakan RapidMiner tidak memiliki *missing value* pada data atribut yang ditandai dengan angka 0 di bagian *Missing*.

4.4. Data Transformation

Pada tahap *transformation*, dilakukan pemberian bobot nilai pada data asli. Data nominal diubah ke bentuk *Numerical* agar mempermudah perhitungan K-Means *Clustering* pada 211 data. Proses *transformation* menggunakan operator *Nominal to Numerical* dapat dilihat pada gambar 6 di bawah ini.



Gambar 6. Operator *Nominal to Numerical* dari RapidMiner

Parameter pada operator *Nominal to Numerical* yang digunakan dapat dilihat pada Tabel 4 di bawah ini.

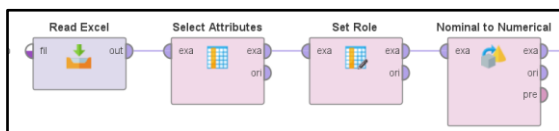
Tabel 4. Parameter *Nominal to Numerical*

No.	Parameter	Isi
1.	Attribute filter type	Subset
2.	Attributes	Select attributes
		Nama penerima
		Pendapatan
		Luas lahan
		Jenis lantai
		Jenis dinding
		Energi untuk memasak
		Status tempat tinggal
		Fasilitas MCK
		Sumber air
		Fasilitas BAB
3.	Coding type	Unique integers

Penggunaan operator *Nominal to Numerical* ini bertujuan untuk mengubah data kategorikal menjadi *Numerical*. Hasil dari operator *Nominal to Numerical* dapat dilihat pada Gambar 7 di bawah ini.

Gambar 7. Hasil *Data Transformation* dari Kategorikal ke *Numerical*

Proses pemodelan pada tahap di RapidMiner, dapat dilihat pada Gambar 8 di bawah ini.

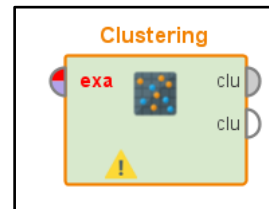


Gambar 8. Proses Pemodelan *Data Transformation*

4.5. Data Mining

Proses *clustering* menggunakan algoritma K-Means dilakukan pada 211 data penerima PKH Desa Tambaksari untuk mengidentifikasi kelayakan

penerima PKH dan mengetahui anggota mana yang layak dan tidak. Data diolah di RapidMiner v10.0 dengan algoritma K-Means agar diperoleh *cluster* data kelayakan penerima PKH. Pada tahap data mining, teknik yang digunakan adalah pengklasteran data dengan mengimplementasikan algoritma K-Means. Operator K-Means berfungsi untuk mengelompokkan data pada *cluster* yang telah ditentukan. Tampilan operator *Clustering* dengan menggunakan K-Means dapat dilihat pada Gambar 9 di bawah ini.



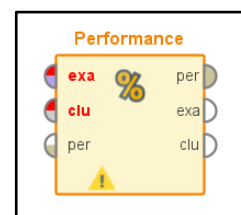
Gambar 9. Operator K-Means dari RapidMiner

Pada operator *Clustering* terdapat parameter yang perlu disesuaikan. Parameter yang digunakan pada operator K-Means yang digunakan dapat dilihat pada Tabel 5 di bawah ini.

Tabel 5. Parameter K-Means

No.	Parameter	Isi
1.	k	2 - 5
2.	Max run	5 - 20
3.	Measure types	Numerical Measure
4.	Numerical Measure	Euclidean Distance
		Correlation Similarity
5.	Max optimization steps	100

Operator performance yang berfungsi untuk mengevaluasi kinerja model yang memberikan daftar nilai secara otomatis sesuai dengan nilai yang diberikan pada *Clustering*. Performance menggunakan operator *cluster distance performance* dapat dilihat pada Gambar 10 di bawah ini.



Gambar 10. Operator *Cluster Distance Performance* dari RapidMiner

Pada *cluster distance performance* terdapat parameter yang harus disesuaikan dengan kebutuhan olah data dapat dilihat pada Tabel 6 di bawah ini.

Tabel 6. Parameter *Cluster Distance Performance*

No.	Parameter	Isi
1.	Main criterion	Davies Bouldin
2.	Main criterion only	
3.	Normalize	
4.	maximize	

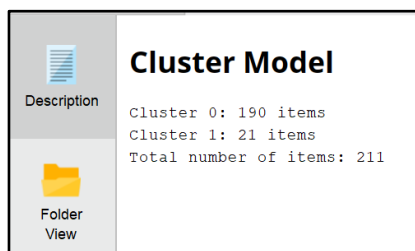
4.6. Evaluasi dan Interpretasi

Pada penelitian ini, untuk mendapatkan kelompok terbaik penentuan kelayakan penerima bantuan PKH dilakukan eksperimen untuk mencari nilai DBI terkecil. Hasil eksperimen dapat dilihat pada Tabel 7 sebagai berikut :

Tabel 7. Hasil Eksperimen

Measure Type Numerical Measure	Clustering K	Max Runs	DBI
Euclidean Distance	2	5	0.415
	3	10	0.596
	4	15	0.469
	5	20	0.424
Correlation Similarity	2	5	0.589
	3	10	1.310
	4	15	2.024
	5	20	3.127

Setelah proses dijalankan, dapat dilihat hasil pada *cluster* model, untuk pilihan Description dapat dilihat pada Gambar 11 di bawah ini.

Gambar 11. Hasil *Cluster Model* dari RapidMiner

Dari hasil *cluster* model didapatkan nilai DBI yang dapat di lihat pada Gambar 12 di bawah ini.



Gambar 12. Nilai DBI terbaik dari RapidMiner

Dari hasil proses eksperimen mencari k optimal berdasarkan DBI dan *measure type* bahwa identifikasi kelayakan penerima bantuan PKH dibagi menjadi 2 *cluster*. Dengan *cluster* 0 termasuk kedalam layak menerima bantuan PKH dan *cluster* 1 tidak layak menerima bantuan PKH.

4.7. Pembahasan

Algoritma K-Means diimplementasikan dengan RapidMiner untuk mengidentifikasi kelayakan penerima PKH. Analisis data membagi penerima menjadi 2 *cluster*, *cluster* 0 untuk penerima layak dengan pendapatan <1jt, dan *cluster* 1 untuk penerima tidak layak dengan pendapatan >1jt. Hasil memberi gambaran jelas perbedaan kondisi ekonomi 2 kelompok, sehingga menentukan kelayakan PKH berdasarkan kriteria pendapatan.

Berdasarkan perbandingan nilai DBI, *cluster* optimal adalah 2 *cluster* (K=2) dengan nilai DBI terendah 0,415. Pengujian data penerima PKH menghasilkan *cluster* 0 sebanyak 190 item dan *cluster* 1 sebanyak 21 item dari total 211 item. *Cluster* 2 terpilih karena nilai DBI mendekati 0 menunjukkan performa *cluster* optimum. Hasil pengelompokan data menjadi 2 *cluster* pada penelitian ini sejalan dengan hasil penelitian terdahulu yang dilakukan oleh [14]. Penelitian tersebut melakukan pengolahan data menjadi 2 *cluster* yang menyatakan k 2 merupakan *cluster* optimum yang didapatkan. Dengan demikian, hasil pada penelitian saat ini telah sesuai dengan teori dan mekanisme *Clustering* K-Means untuk mendapatkan *cluster* optimum dalam pengelompokan data.

Berdasarkan analisis, *measure types* terbaik untuk mengidentifikasi kelayakan penerima PKH adalah numerical measures dengan *Euclidean Distance* yang menghasilkan nilai DBI paling optimal 0,415 untuk 2 *cluster*. Sedangkan nilai DBI terbaik dari *Correlation Similarity* adalah 0,589 untuk 2 *cluster*. *Euclidean Distance* terpilih karena menghasilkan nilai DBI terendah. Hasil pengelompokan data menjadi 2 *cluster* dengan *measure types* jenis *numerical measure* dengan *Euclidean Distance* pada penelitian ini sejalan dengan hasil penelitian terdahulu yang dilakukan oleh [15]. Penelitian tersebut melakukan pengolahan data berdasarkan parameter *measure types* jenis *numerical measure* dengan *Euclidean Distance* yang mendapatkan 2 *cluster* berdasarkan dengan nilai *Davies Bouldin Index* (DBI). Dengan demikian, hasil pada penelitian saat ini telah sesuai dengan teori dan mekanisme *Clustering* K-Means untuk mendapatkan *cluster* optimum dalam pengelompokan data.

5. KESIMPULAN DAN SARAN

Kesimpulan penelitian ini adalah pengujian data penerima PKH Desa Tambaksari dengan K-Means *Clustering* menghasilkan 2 *cluster* optimal berdasarkan nilai DBI terbaik 0,415, yaitu *cluster* 0 dengan 190 data penerima layak dan *cluster* 1 dengan 21 data penerima tidak layak. *Measure type* terbaik adalah *Numerical Measure* dengan *Euclidean Distance*. Saran untuk penelitian selanjutnya adalah membandingkan dengan algoritma lain dan mengembangkan *cluster* yang terbentuk dengan K-medoids. Pemerintah dapat memanfaatkan hasil klasterisasi ini untuk mengoptimalkan alokasi bantuan

PKH di Desa Tambaksari sesuai karakteristik setiap *cluster*, sehingga bantuan lebih tepat sasaran dan efisien.

DAFTAR PUSTAKA

- [1] C. E. Purnomo and R. Rikendry, "Penerapan metode C4. 5 untuk klasifikasi warga miskin pada desa Mengandung Sari," *J. Teknol. dan Sist. ...*, vol. 2, no. 3, pp. 14–25, 2021, [Online]. Available: <http://jim.teknokrat.ac.id/index.php/sisteminformasi/article/view/936>
- [2] S. Ufriani, "Penerapan algoritma *clustering* k-means untuk menentukan prioritas penerima bantuan dana sosial PKH di kelurahan kampung singkep," *J. Inform. Dan Rekayasa Komput. ...*, vol. 3, no. 1, p. 342–350, 2023, [Online]. Available: <https://ejournal.unama.ac.id/index.php/jakakom/article/view/726>
- [3] I. Kurniawan and R. A. Saputra, "Penerapan algoritma C5. 0 pada sistem pendukung keputusan kelayakan penerimaan beras masyarakat miskin," *J. Inform.*, vol. 4, no. 2, p. 236–240, 2017, [Online]. Available: <https://ejournal.bsi.ac.id/ejurnal/index.php/ji/article/view/2320>
- [4] F. Febriansyah and S. Muntari, "Penerapan Algoritma K-Means untuk Klasterisasi Penduduk Miskin pada Kota Pagar Alam," *JISKA (Jurnal Inform. Sunan ...)*, vol. 8, no. 1, pp. 66–27, 2023, [Online]. Available: <https://ejournal.uin-suka.ac.id/saintek/JISKA/article/view/3787>
- [5] K. Anam, D. Sudrajat, and D. A. Kurnia, "Analisis segmentasi pelanggan menggunakan metode k-means *clustering*," *Inf. Commun. Technol.*, vol. 21, no. 2, pp. 273–278, 2022.
- [6] U. S, "Penerapan data mining dengan mengimplementasikan algoritma k-means dalam proses *clustering* untuk pengelompokan mahasiswa calon penerima beasiswa kip," *Build. Informatics, Technol. Sci.*, vol. 5, no. 1, pp. 70–85, 2023, doi: 10.47065/bits.v5i1.3411.
- [7] S. M. Sinaga, J. T. Hardinata, and M. Fauzan, "Implementasi data mining *clustering* tingkat kepuasan konsumen terhadap pelayanan go-jek," *KSATRIA J. Penerapan Sist. Inf. (Komputer Manajemen)*, vol. 2, no. 2, pp. 118–124, 2021.
- [8] G. Gustientiedina, M. H. Adiya, and Y. Desnelita, "Penerapan Algoritma K-Means Untuk *Clustering* Data Obat-Obatan," *J. Nas. Teknol. dan Sist. Inf.*, vol. 5, no. 1, pp. 17–24, 2019, doi: 10.25077/teknosi.v5i1.2019.17-24.
- [9] U. N. Putra *et al.*, "Penerapan data mining K-Means *clustering* untuk mengelompokkan berbagai jenis merk laptop," *Semin. Nas. Sist. ...*, pp. 241–249, 2021, [Online]. Available: <https://sisematik.nusaputra.ac.id/index.php/sisematik/article/view/35%0Ahttps://sisematik.nusaputra.ac.id/index.php/sisematik/article/download/35/31>
- [10] E. D. Madyatmadja, M. N. Ridho, A. R. Pratama, M. Fajri, and L. Novianto, "Penerapan visualisasi data terhadap klasifikasi tindak kriminal di Indonesia," *Infotech J. Technol. Inf.*, vol. 8, no. 1, pp. 61–68, 2022, doi: 10.37365/jti.v8i1.127.
- [11] V. C. Pamungkas, L. Muflikhah, and ..., "Klasifikasi Penerimaan Program Keluarga Harapan (PKH) Menggunakan Metode Learning Vector Quantization (Studi Kasus Desa Kedungjati)," ... *Teknol. Inf. dan ...*, vol. 3, no. 3, pp. 2659–2666, 2019, [Online]. Available: <http://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/4781>
- [12] S. Thomas and N. P, "Analisis data untuk pengelompokan mahasiswa dengan metode k-mean," *J. Inform. DAN Teknol. Inf.*, vol. 6, no. 3, pp. 118–123, 2021.
- [13] D. Hastari, F. Nurunnisa, S. Winanda, and ..., "Penerapan algoritma k-means dan k-medoids untuk mengelompokkan data negara berdasarkan faktor sosial-ekonomi dan kesehatan," ... *Nas. Penelit. dan ...*, pp. 274–281, 2023, [Online]. Available: <https://journal.irpi.or.id/index.php/sentimas/article/view/542%0Ahttps://journal.irpi.or.id/index.php/sentimas/article/download/542/365>
- [14] D. K. Utami, N. Irawati, and Sumantri, "Analisis Metode k-Means pada *Clustering* Penerimaan Bantuan PKH Desa Pulau Rakyat Tua Analysis of the k - Means Method in *Clustering* Acceptance of PKH Aid in," vol. 12, no. September, pp. 953–961, 2023.
- [15] M. N. Dayat, N. Suarna, and Y. A. Wijaya, "Analisa *Clustering* untuk mengelompokan data penayangan film bioskop menggunakan algoritma K-Means," *Intern. (Information Syst. Journal)*, vol. 6, no. 1, pp. 68–78, 2023, [Online]. Available: <https://jurnal.masoemuniversity.ac.id/index.php/internal/article/view/686>