ANALISIS DATASET PROGRAM KELUARGA HARAPAN (PKH) DESA SIDAHARJA MENGGUNAKAN ALGORITMA K-MEANS

Deka Maulana Fathul Aziz, Bambang Irawan, Agus Bahtiar
Teknik Informatika, STMIK IKMI Cirebon
Sistem Informasi, STMIK IKMI Cirebon
JI. Perjuangan No. 10 B, Majasem, Cirebon, Indonesia
dekamaulanafathulaziz@gmail.com

ABSTRAK

Program Keluarga Harapan yang selanjutnya disebut (PKH) adalah program pemberian bantuan sosial bersyarat kepada Keluarga Miskin (KM) yang ditetapkan sebagai keluarga penerima manfaat PKH. Permasalahan pada penelitian ini yaitu bagaimana mengimplementasikan metode *K-means* dalam *Clustering* data penerima PKH Desa Sidaharja dengan mencari *cluster* yang terbaik. Penelitian ini bertujuan untuk mengelompokkan data penerima PKH Desa Sidaharja dengan mencari *cluster* yang terbaik. Selain itu, penelitian ini bertujuan untuk mendapatkan parameter *Numerical Measures Type* apa yang menghasilkan kelompok terbaik pada data penerima PKH Desa Sidaharja. Pada penelitian ini digunakan algoritma *K-means* guna mendapatkan informasi dari hasil pengelompokan data penerima PKH Desa Sidaharja. Teknik analisis data menggunakan *Knowledge Discovery in Database* (KDD). Hasil dari pengelompokan menunjukan bahwa dari jumlah *cluster* (K = 2, 3, 4, 5, 6) menghasilkan K mana yang terbaik performanya yaitu yang mendekati angka 0. Nilai K yang paling mendekati angka 0 adalah (K = 6). Diperoleh DBI optimal untuk (K = 6) dengan nilai 0.230 untuk parameter *Numerical Measures Type - Euclidean Distance*. Untuk mengevaluasi kinerja algoritma *K-means* dapat dilihat dari *performance*, dimana nilai *Davies Bouldin* yang mendekati 0 mengindikasikan kualitas algoritma yang semakin baik. Dengan demikian, maka jumlah *cluster* terbaik dalam percobaan ini adalah 6 atau yang mendekati 0.

Kata kunci : *Data Mining, K-means,* DBI, *Program Keluarga Harapan, Desa Sidaharja.*

1. PENDAHULUAN

Kemiskinan merupakan salah satu permasalahan utama yang sering dihadapi oleh banyak negara di dunia, terutama Indonesia [1]. Setiap pemimpin daerah maupun pusat menjadikan kemiskinan sebagai prioritas utama yang harus diselesaikan [2]. Kurangnya pendapatan akan berakibat pada rendahnya kualitas hidup seseorang [3]. Pemerintah sendiri telah melakukan beberapa upaya untuk mengentaskan kemiskinan, hal ini dilakukan melalui berbagai program bantuan sosial [4].

Untuk menurunkan tingkat kemiskinan maka pemerintah Indonesia melalui Kementerian Sosial telah mengeluarkan Program Keluarga Harapan (PKH) [3]. Dilansir dari laman Kemensos, Program Keluarga Harapan yang selanjutnya disebut (PKH) adalah program pemberian bantuan sosial bersyarat kepada Keluarga Miskin (KM) yang ditetapkan sebagai keluarga penerima manfaat PKH. Melalui PKH, KM didorong untuk memiliki akses dan memanfaatkan pelayanan sosial dasar kesehatan, pendidikan, pangan dan gizi, perawatan, dan pendampingan, termasuk akses terhadap berbagai program perlindungan sosial lainnya yang merupakan program komplementer secara berkelanjutan. Faktor pembagian kelompok penerima bantuan PKH adalah dari keluarga miskin atau pra sejahtera, mempunyai anggota keluarga yang mencakup ibu hamil/menyusui, mempunyai anak dari usia 0 hingga 6 tahun, mempunyai anak yang berpendidikan SD, SMP, atau SMA sederajat, mempunyai anggota keluarga lanjut usia yang berusia minimal 60 tahun dan penyandang disabilitas berat [5]. Namun permasalahannya, proses penyaluran bantuan terkadang masih dianggap tidak merata atau tepat sasaran [6]. Hal ini dikarenakan banyak masyarakat yang merasa dirinya pantas mendapat PKH tetapi terlihat tidak memenuhi kriteria yang ditentukan [6]. Dari latar belakang yang telah dijelaskan, maka permasalahan pada penelitian ini yaitu bagaimana mengimplementasikan metode *K-means* dalam *Clustering* data penerima Program Keluarga Harapan (PKH) Desa Sidaharja dengan mencari *cluster* yang terbaik. Untuk menghadapi permasalahan ini, penelitian ini memanfaatkan data mining dengan menerapkan teknik *clustering* guna mengelompokkan data ke dalam kelompok yang memiliki kesamaan.

Penelitian ini bertujuan untuk mengelompokkan data penerima Program Keluarga Harapan (PKH) Desa Sidaharja dengan mencari *cluster* yang terbaik. Selain itu, penelitian ini juga bertujuan untuk mendapatkan Parameter *Numerical Measures Type* apa yang menghasilkan kelompok terbaik pada data penerima PKH Desa Sidaharja, serta untuk mendapatkan infromasi dari hasil pengelompokan data penerima PKH Desa Sidaharja. Digunakan indikator *Indeks Davies-Bouldin* (DBI) untuk menentukan jumlah *cluster* terbaik [7]. Melalui tahapan analisis tersebut, diharapkan penyaluran bantuan PKH ini menjadi lebih optimal dan tepat sasaran diberikan kepada masyarakat yang benar-benar membutuhkan [6] khusunya di Desa Sidaharja.

Data mining adalah proses yang menggunakan metode statistik, matematika, kecerdasan buatan dan

pembelajaran mesin untuk mengidentifikasi informasi yang berguna dan temukan pengetahuan tersembunyi dari berbagai database [8]. *Clustering* adalah proses pengelompokan suatu kelompok objek berdasarkan sejumlah kemiripan [9] Pada dasarnya, *clustering* adalah pendekatan untuk mencari dan mengelompokkan data berdasarkan kesamaan karakteristik di antara satu set data dengan yang lain [10]. Pada Penelitian data mining melibatkan tahapan pengolahan data yang salah satunya sering disebut dengan *Knowledge Discovery in Database* (KDD) [11]. Algoritma *K-means* dipilih karena Algoritma ini memiliki keuntungannya adalah mudah diterapkan dan dijalankan, relatif cepat, mudah disesuaikan dan paling banyak digunakan dalam aktivitas data mining [12]. Pada penelitian ini algoritma *K-means* dioptimasi menggunakan *Davies Bouldin Index* (DBI) untuk menentukan jumlah *cluster* yang optimal [9]. *Indeks Davies-Bouldin* (DBI) adalah pengukuran yang populer untuk mengevaluasi kinerja pengelompokan dengan membagi klaster [9].

2. TINJAUAN PUSTAKA

2.1. Penelitian Terdahulu

Salah satu penelitian terkait yang sudah behasil dilakukan diantaranya penelitian oleh Dwi Kurnia Utami, Novica Irawaci, dan Sumantri pada tahun 2023 berjudul "Analisis Metode *K-means* pada *Cluster*ing Penerimaan Bantuan PKH Desa Pulau Rakyat Tua" Masalah dalam penelitian ini adalah distribusi bantuan Program Keluarga Harapan (PKH) di Desa Pulau Rakyat Tua yang sering dianggap tidak tepat sasaran. Hasil penelitian ini menunjukkan bahwa penggunaan algoritma *K-means cluster*ing dalam klasterisasi penerima bantuan Program Keluarga Harapan (PKH) di Desa Pulau Rakyat Tua menghasilkan dua klaster: 29 penerima bantuan PKH yang layak dan 1 penerima bantuan PKH yang tidak layak [13].

Kemudian, penelitian oleh Said Abdul Azis, Sarjon Defit, dan Yuhandri Yunus pada tahun 2021 berjudul "Klasterisasi Dana Bantuan Pada Program Keluarga Harapan (PKH) Menggunakan Metode *K-means*" Masalah dalam penelitian ini adalah adanya kebutuhan untuk melakukan pengelompokan desa berdasarkan tingkat kemiskinan menggunakan metode *cluster*ing, namun untuk melakukan pengelompokan tersebut diperlukan data-data masyarakat miskin atau data terpadu kesejahteraan sosial sebagai acuan. Hasil penelitian ini membahas penggunaan algoritma *K-means* untuk mengelompokkan penyaluran dana bantuan Program Keluarga Harapan (PKH) berdasarkan empat tahapan dengan hasil klasterisasi menjadi 3 kelompok: kelompok C1 (Rumah Tangga Hampir Miskin), C2 (Rumah Tangga Miskin), dan C3 (Rumah Tangga Sangat Miskin) [3].

Dan, penelitian oleh Nurahman dan Jetri Susanto pada tahun 2023 berjudul "Klasterisasi Data Penerima Bantuan Langsung Tunai Menggunakan Algoritma *K-means*" Masalah dalam penelitian ini adalah untuk mengetahui seberapa berhaknya penerima BLT desa Pelangsian tahun 2019 yang terdampak virus Covid-19 dengan melihat beberapa *dataset* yang telah dikumpulkan, seperti rata-rata penghasilan per bulan, jumlah anggota keluarga laki-laki dan perempuan yang tinggal di rumah, serta dilihat dari atap rumah dan dinding tempat tinggal penerima BLT desa tersebut. Hasil klasterisasi data ini dapat membantu dalam penyaluran bantuan kepada kelompok yang membutuhkan, dan penelitian ini juga memberikan rekomendasi untuk penggunaan *dataset* yang lebih besar dalam penelitian mendatang. Selain itu, penelitian ini juga mencakup referensi terkait studi-studi terkait COVID-19 dan analisis data [11].

2.2. Landasan Teori

2.2.1. *Data Mining*

Data mining adalah proses yang menggunakan metode statistik, matematika, kecerdasan buatan dan pembelajaran mesin untuk mengidentifikasi informasi yang berguna dan temukan pengetahuan tersembunyi dari berbagai database [8].

2.2.2. *Clustering*

Clustering adalah proses pengelompokan suatu kelompok objek berdasarkan sejumlah kemiripan [9] pada dasarnya, *clustering* adalah pendekatan untuk mencari dan mengelompokkan data berdasarkan kesamaan karakteristik di antara satu set data dengan yang lain [10].

2.2.3. *K-means*

Algoritma *K-Means* merupakan metode pengelompokan iteratif yang membagi *dataset* ke dalam sejumlah k *cluster* yang telah ditentukan sebelumnya. Tingkat ketelitian yang tinggi terhadap ukuran objek membuat algoritma ini efisien dan dapat diandalkan untuk memproses objek dalam jumlah besar. Kelebihan lainnya adalah algoritma ini tidak berpengaruh terhadap urutan objek [10].

2.2.4. *Indeks Davies Bouldin* (DBI)

Indeks Davies Bouldin (DBI) merupakan metrik yang umum digunakan untuk mengevaluasi efektivitas pengelompokan dengan memperhitungkan perbedaan antar klaster. DBI memberikan indikasi tentang sejauh mana kualitas pengelompokan, mempertimbangkan baik kesamaan di dalam klaster maupun perbedaan di antara klaster [9].

2.2.5. *Knowledge Discovery in Database* (KDD)

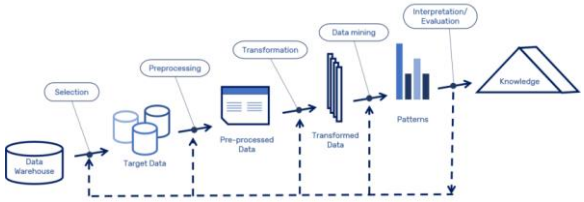
Knowledge Discovery in Database (KDD) adalah pendekatan yang diterapkan untuk mendapatkan informasi yang berasal dari basis data yang tersedia. Hasil temuan dari proses ini dapat dimanfaatkan untuk membentuk basis pengetahuan yang digunakan dalam konteks pengambilan keputusan [10].

2.2.6. Program Keluarga Harapan (PKH)

Dilansir dari laman Kemensos, Program Keluarga Harapan yang selanjutnya disebut (PKH) adalah program pemberian bantuan sosial bersyarat kepada Keluarga Miskin (KM) yang ditetapkan sebagai Keluarga Penerima Manfaat (KPM) PKH. Melalui PKH, KM didorong untuk dapat mengakses dan memanfaatkan pelayanan sosial dasar di bidang kesehatan, pendidikan, pangan dan gizi, perawatan dan pendampingan, termasuk akses terhadap berbagai program perlindungan sosial lainnya yang merupakan program komplementer secara berkelanjutan.

3. METODE PENELITIAN

Pendekatan desain dalam penelitian ini mengadopsi *Knowledge Discovery in Database* (KDD), yang melibatkan proses seni sintesis dan penemuan ilmiah, interpretasi, serta visualisasi data dari berbagai jenis. Penemuan pengetahuan di dalam gudang data (KDD) menjadi aspek krusial dalam mengungkap pola data, menghasilkan temuan yang akurat, inovatif, dapat ditindaklanjuti, dan berwawasan.



Gambar 1. Tahapan Proses KDD [14]

Proses *Knowledge Discovery in Database* (KDD) dapat diuraikan sebagai berikut :

1. *Selection*
   Data penerima PKH Desa Sidaharja merupakan data tahun 2023. Data yang digunakan dalam penelitian ini sebanyak 188 *record*. Pada tahapan ini, terjadi penyeleksian atribut untuk menentukan atribut mana yang akan digunakan dalam proses data mining.
2. *Pre-processing*
   Pada tahapan ini, hanya fokus pada data yang akan diolah. Ini juga termasuk membersihkan data untuk memperbaiki duplikat, nilai yang hilang, dan memperbaiki kesalahan seperti kesalahan ketik.
3. *Transformation*
   Merupakan proses transformasi untuk data terpilih, sehingga data tersebut layak dan sesuai untuk diolah dalam data mining.
4. *Data Mining*
   Proses data mining ini melibatkan penerapan algoritma atau pencarian pengetahuan, dan dalam penelitian ini, digunakan algoritma *K-means clustering* pada *tools rapidminer* versi 10.2.
5. *Interpretation / Evaluation*
   Tahapan ini merupakan tahap akhir dari proses data mining. Tahap ini dilakukan setelah data diolah dengan algoritma yang dipilih menggunakan *tools rapidminer* dan tujuannya adalah untuk melakukan evaluasi terhadap hasil proses data mining.

4. HASIL DAN PEMBAHASAN

Analisis data menggunakan teknik *Knowledge Discovery in Database* (KDD).

4.1. *Dataset*

Dataset yang digunakan dalam penelitian ini adalah data penerima Program Keluarga Harapan (PKH) Desa Sidaharja. Jumlah *record* sebanyak 188 *record*, *dataset* tersebut berisi atribut-atribut yang dibutuhkan dalam penelitian sebagaimana terlihat dalam gambar 2 di bawah ini.

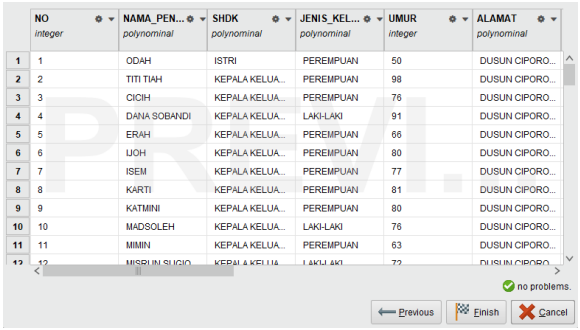
No	Nama_Penerima	Shdk	Jenis_Kelamin	.....	Penghasilan
1	Odah	Istri	Perempuan	.....	500000
2	Titi Tiah	Kepala Keluarga	Perempuan	.....	1000000
3	Cicih	Kepala Keluarga	Perempuan	.....	0
4	Dana Sobandi	Kepala Keluarga	Laki-Laki	.....	1500000
5	Erah	Kepala Keluarga	Perempuan	.....	0
6	Ijoh	Kepala Keluarga	Perempuan	.....	0
7	Isem	Kepala Keluarga	Perempuan	.....	0
8	Karti	Kepala Keluarga	Perempuan	.....	0
9	Katmini	Kepala Keluarga	Perempuan	.....	1000000
10	Madsoleh	Kepala Keluarga	Laki-Laki	.....	0
.....	.....	.....	.....	.....	.....
188	Haryati	Istri	Perempuan	.....	500000

Gambar 2. Dataset Penelitian dari PKH

4.2. *Selection*

Langkah pertama adalah mencari lokasi file Itu sebelumnya dibuat dalam format xlxs dengan operator *Read Excel* yang berfungsi baca data xlxs.

Gambar 3. Operator Read Excel dari Rapidminer



	NO integer	NAMA_PEN... polynominal	SHDK polynominal	JENIS_KEL... polynominal	UMUR integer	ALAMAT polynominal
1	1	ODAH	ISTRI	PEREMPUAN	50	DUSUN CIPORO...
2	2	TITI TIAH	KEPALA KELUA...	PEREMPUAN	98	DUSUN CIPORO...
3	3	CICIH	KEPALA KELUA...	PEREMPUAN	76	DUSUN CIPORO...
4	4	DANA SOBANDI	KEPALA KELUA...	LAKI-LAKI	91	DUSUN CIPORO...
5	5	ERAH	KEPALA KELUA...	PEREMPUAN	66	DUSUN CIPORO...
6	6	IJOH	KEPALA KELUA...	PEREMPUAN	80	DUSUN CIPORO...
7	7	ISEM	KEPALA KELUA...	PEREMPUAN	77	DUSUN CIPORO...
8	8	KARTI	KEPALA KELUA...	PEREMPUAN	81	DUSUN CIPORO...
9	9	KATMINI	KEPALA KELUA...	PEREMPUAN	80	DUSUN CIPORO...
10	10	MADSOLEH	KEPALA KELUA...	LAKI-LAKI	76	DUSUN CIPORO...
11	11	MIMIN	KEPALA KELUA...	PEREMPUAN	63	DUSUN CIPORO...

Gambar 4. Hasil Operator Read Excel dari Rapidminer

Pada gambar 4 menunjukkan datanya yang diimpor, maka tombolnya ada di operator *Read Excel* tidak lagi memiliki tanda seru berwarna kuning. Artinya operator sudah memasukkan datanya dan siap memprosesnya.

Kemudian menggunakan operator *Set Role* untuk mengidentifikasi *id* dalam *dataset*.

Gambar 5. Operator Set Role dari Rapidminer

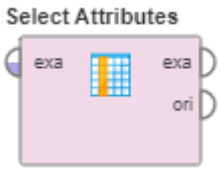
Tabel 1. Parameter Set Role

No.	Parameter	Isi	
1.	attribute name	no	
2.	target role	id	
set additional roles		attribute name	target role
		no	id

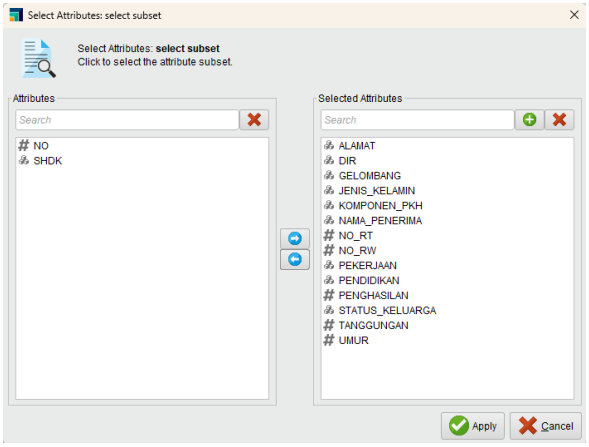
Row No.	NO	NAMA_PENERIMA
1	1	ODAH
2	2	TITI TIAH
3	3	CICIH
4	4	DANA SOBANDI
5	5	ERAH

Gambar 6. Hasil Operator Set Role dari Rapidminer

Kemudian menggunakan operator *Select Attributes*, Operator ini memilih *subset* Atribut dari *ExampleSet* dan menghapus atribut lainnya, sebagaimana terlihat dalam gambar 7 di bawah ini.



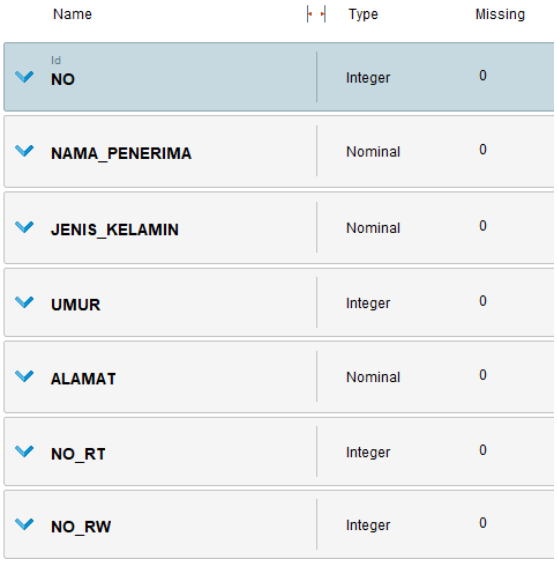
Gambar 7. Operator Select Atributes dari Rapidminer



Gambar 8. Hasil Operator Select Attributes dari Rapidminer

4.3. *Prepocessing*

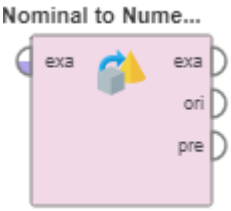
Dikarenakan tidak ada data yang hilang atau memiliki nilai yang *missing* dalam *dataset*, oleh karena itu, tahapan *prepocessing* tidak perlu dilakukan. Hasil statistik *dataset*, sebagaimana terlihat pada gambar 9 menunjukkan bahwa tidak ada atribut yang memiliki nilai yang *missing*. Pemeriksaan langsung per-record untuk menilai konsistensi *dataset* mengindikasikan bahwa nilai-nilai dalam *dataset* konsisten.



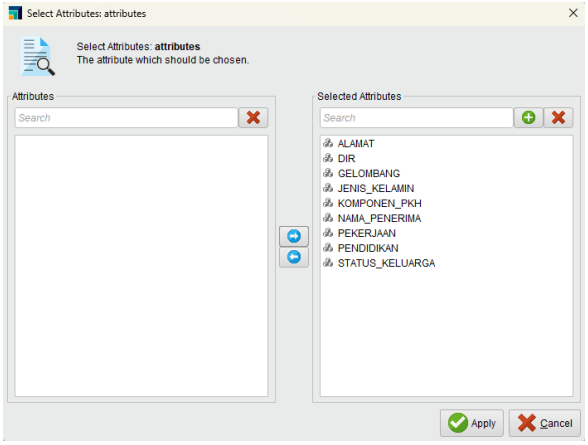
Gambar 9. Statistik Dataset dari Rapidminer

4.4. *Transformation*

Dikarenakan *dataset* ada yang menggunakan tipe data *nominal* sementara algoritma *K-means* memerlukan tipe data *numeric*, maka transformasi diperlukan untuk mengubah data *nominal* menjadi *numerical*. Hal ini dapat dilakukan dengan menggunakan operator *Nominal to Numerical*, sebagaimana terlihat dalam gambar 10 di bawah ini.



Gambar 10. Operator Nominal to Numerical dari Rapidminer

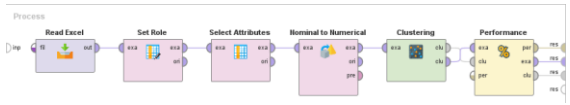


Gambar 11. Hasil Operator Nominal to Numerical dari Rapidminer

4.5. *Data Mining*

Penentuan jumlah *cluster* adalah langkah yang harus dilakukan. Nilai K yang digunakan melibatkan identifikasi jumlah klaster optimal serta jumlah anggota optimal dalam setiap klaster. Pengelompokan data dilakukan dengan menggunakan algoritma *K-means* dengan *Cluster Distance Performance* dan dioptimasi dengan *Davies Bouldin Index* (DBI) pada *Tools Rapidminer* versi 10.2.

Desain model proses hingga tahap Data Mining sebagaimana terlihat dalam gambar 12 di bawah ini.

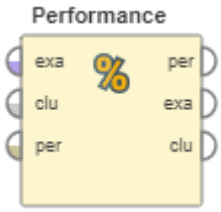


Gambar 12. Desain K-means dari Rapidminer

Operator ini melakukan pengelompokan menggunakan algoritma *K-means*, sebagaimana terlihat dalam gambar 13 di bawah ini.

Gambar 13. Cluster K-means dari Rapidminer

Clustering mengelompokkan contoh yang mirip satu sama lain. Karena tidak ada atribut label yang diperlukan, *clustering* dapat digunakan pada data yang tidak berlabel dan merupakan algoritma pembelajaran mesin tanpa pengawasan. Algoritma *K-means* menentukan satu set k *cluster* dan menetapkan setiap contoh ke dalam satu *cluster*. Klaster terdiri dari contoh yang serupa. Kemiripan antara examples didasarkan pada ukuran jarak di antara mereka.



Gambar 14. Performance dari Rapidminer

Operator ini digunakan untuk evaluasi kinerja metode pengelompokan berbasis *centroid*. Operator ini memberikan daftar nilai kriteria kinerja berdasarkan *centroid cluster*. Model ini juga memiliki informasi mengenai *centroid* dari setiap *cluster*. Operator *Cluster Distance Performance* mengambil model *cluster centroid* dan kumpulan *cluster* ini sebagai input dan mengevaluasi kinerja model berdasarkan *centroid cluster*. Ada dua ukuran kinerja yang didukung: Rata-rata dalam jarak *cluster* dan *Indeks Davies-Bouldin*. Ukuran-ukuran kinerja ini dijelaskan dalam parameter.

4.1.1. Hasil Eksperimen Klaster (Numerical Measures Type - Euclidean Distance)

Hasil pengelompokan didapatkan melalui lima uji coba yang dilakukan dengan menggunakan nilai K sebanyak 2, 3, 4, 5, dan 6 untuk *Numerical Measures Type = Euclidean Distance*. Tahap berikutnya adalah menghitung nilai DBI untuk setiap pengelompokan percobaan tersebut. Hasil dari nilai DBI sebagaimana terlihat dalam tabel 2 di bawah ini.

Tabel 2. Klaster yang dihasilkan
(Numerical Measures Type - Euclidean Distance)

Numerical Measures	K	DBI
Euclidean Distance	2	0.535
	3	0.383
	4	0.338
	5	0.259
	6	0.230

Dalam Tabel 2, evaluasi DBI dari algoritma *K-means* menunjukkan nilai 0.230 dengan jumlah k=6. Sementara itu, untuk k=2, DBI mencapai 0.535. k=3 memiliki nilai DBI sebesar 0.383. Jumlah *cluster* k=4 dan k=5, masing-masing memiliki nilai DBI sebesar 0.338 dan 0.259.

4.1.2. Hasil Eksperimen Klaster (Numerical Measures – Camberra Distance)

Hasil pengelompokan didapatkan melalui lima uji coba yang dilakukan dengan menggunakan nilai K sebanyak 2, 3, 4, 5, dan 6 untuk *Numerical Measures Type = Camberra Distance*. Tahap berikutnya adalah menghitung nilai DBI untuk setiap pengelompokan percobaan tersebut. Hasil dari nilai DBI sebagaimana terlihat dalam tabel 3 di bawah ini.

Tabel 3. Klaster yang dihasilkan
(Numerical Measures Type - Canberra Distance)

Numerical Measures	K	DBI
Canberra Distance	2	0.855
	3	0.855
	4	0.855
	5	0.829
	6	0.855

Dalam Tabel 3, evaluasi DBI dari algoritma *K-means* menunjukkan nilai 0.829 dengan jumlah k=5. Sementara itu, untuk k=2, k=3, k=4 dan k=6, masing-masing memiliki nilai DBI yang sama yaitu sebesar 0.885.

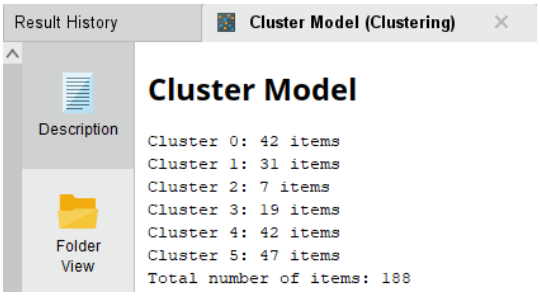
4.1.3. Hasil Eksperimen Klaster (Numerical Measures Type – Correlation Similarity)

Hasil pengelompokan didapatkan melalui lima uji coba yang dilakukan dengan menggunakan nilai K sebanyak 2, 3, 4, 5, dan 6 untuk *Numerical Measures Type = Correlation Similarity*. Tahap berikutnya adalah menghitung nilai DBI untuk setiap pengelompokan percobaan tersebut. Hasil dari nilai DBI sebagaimana terlihat dalam tabel 4 di bawah ini.

Tabel 4. Klaster yang dihasilkan
(Numerical Measures Type – Correlation Similarity)

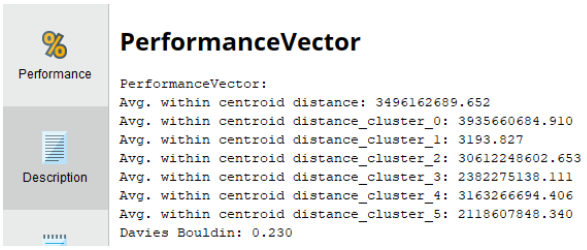
Numerical Measures	K	DBI
Correlation Similarity	2	0.581
	3	0.568
	4	0.691
	5	0.996
	6	1.742

Dalam Tabel 4, evaluasi DBI dari algoritma *K-means* menunjukkan nilai 0.568 dengan jumlah k=3. Sementara itu, untuk k=2, DBI mencapai 0.581. k=4 memiliki nilai DBI sebesar 0.691. Jumlah *cluster* k=5 dan k=6, masing-masing memiliki nilai DBI sebesar 0.996 dan 1.742.



Gambar 15. Hasil Cluster Model dari Rapidminer

Pada gambar 15 hasil pengelompokan menunjukkan adanya enam *cluster* dengan jumlah data sebagai berikut: C0 berjumlah 42 data, C1 berjumlah 31 data, C2 berjumlah 7 data, C3 berjumlah 19 data, C4 berjumlah 42 data, dan C5 berjumlah 47 data. Hal ini dapat dilihat pada gambar 15, yang menunjukkan pembagian data ke dalam enam kelompok.



Gambar 16. Hasil Performance Vector dari Rapidminer

Hasil dari pengelompokan menggunakan algoritma *K-means* dengan menggunakan *tools Rapidminer* versi 10.2 untuk mencari nilai K2 – K6 menghasilkan K mana yang terbaik performanya yaitu yang mendekati angka 0. Nilai K yang paling mendekati angka 0 adalah K6. Diperoleh DBI optimal untuk K6 dengan nilai 0.230 untuk parameter *Numerical Measures Type - Euclidean Distance* dari tiga parameter *Numerical Measures Type* yang dilakukan ekperimen yaitu *Euclidean Distance, Canberra Distance, dan Correlation Similarity*. Untuk mengevaluasi kinerja algoritma *K-means* dapat dilihat dari *performance*, dimana nilai *Davies Bouldin* yang mendekati 0 mengindikasikan kualitas algoritma yang semakin baik. Dengan demikian, maka jumlah *cluster* terbaik dalam percobaan ini adalah 6 atau yang mendekati 0.

4.6. *Interpretation / Evaluation*

Berdasarkan perbandingan antara DBI dan metode *K-means* dari K2 hingga K6, ditemukan bahwa *cluster* terkecil yang mendekati angka 0 adalah K6, dengan nilai DBI sebesar 0.230 yaitu pada parameter *Numerical Measure Type - Euclidean Distance*. Oleh karena itu, dapat disimpulkan bahwa K6 dengan nilai DBI 0.230 menunjukkan bahwa *cluster* tersebut memiliki nilai DBI yang rendah dibandingkan K lainnya, sehingga dianggap sebagai hasil *cluster* terbaik.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil uji data menggunakan *tools Rapidminer* versi 10.2 dan algoritma *K-means*, setelah menghasilkan nilai *Davies-Bouldin Index* (DBI) dengan metode *K-means* dari K-2 hingga K-6 seperti yang tercantum dalam tabel di atas, dapat disimpulkan bahwa *cluster* yang paling mendekati 0 adalah K-6, dengan nilai DBI sebesar 0.230 yaitu pada parameter *Numerical Measure Type - Euclidean Distance*. Karena K-6 memiliki nilai terendah dibandingkan dengan K lainnya, maka dapat disimpulkan bahwa hasil *cluster* terbaik adalah K-6 dengan nilai DBI 0.230 yang paling mendekati 0. Berdasarkan hasil dan pembahasan dari penelitian ini, maka terdapat beberapa saran untuk pengembangan penelitian selanjutnya yaitu penelitian ini hanya melakukan eksperimen pada nilai k2 hingga k6, parameter *numerical measures type euclidean distance, canberra distance, dan correlation similarity.*

Sehingga dapat dikembangkan untuk lebih lanjut yaitu. Berkenaan dengan jumlah K yang lebih bervariasi lagi untuk dilakukan eksperimen. Serta dapat dikembangkan untuk lebih lanjut berkenaan dengan parameter *numerical measures type* yang lain seperti *chebychev distance*, *cosine similarity*, *dice similarity* atau parameter-parameter yang lain untuk dilakukan eksperimen. Dalam konteks penelitian ini, hanya 188 *record* data yang digunakan. Untuk meningkatkan validitas temuan, disarankan agar penelitian mendatang mempertimbangkan penggunaan jumlah *dataset* yang lebih besar. Penggunaan data yang lebih besar memiliki potensi untuk menghasilkan model yang lebih beragam, bahkan dapat meningkatkan validitas hasil penelitian.

DAFTAR PUSTAKA

[1] F. Juliawati, R. Buaton, and R. Saragih, "Pengelompokan data mining penerimaan bantuan pangan non tunai (BPNT) menggunakan metode clustering (Studi Kasus: Kantor Desa Payabakung Hamparan Perak)," *Explorer (Hayward).*, vol. 3, no. 2, pp. 69–76, 2023, [Online]. Available: https://journal.fkpt.org/index.php/Explorer/article/view/793

[2] Y. Kusnadi and M. S. Putri, "Clustering menggunakan metode k-means untuk menentukan prioritas penerima bantuan bedah rumah (Studi Kasus: Desa Ciomas Bogor)," *Jurnal Teknologi Informatika dan Komputer*. repository.bsi.ac.id, pp. 17–24, 2021. [Online]. Available: https://repository.bsi.ac.id/index.php/unduh/item/316100/2021-Genap-2021-Jurnal-TIK-UMHT-Yahdi-Kusnadi-Mardiani.pdf

[3] A. A. Said, S. Defit, and Y. Yunus, "Klasterisasi dana bantuan pada program keluarga harapan (PKH) menggunakan metode k-means," *J. Inform. Ekon. Bisnis*, vol. 3, no. 2, pp. 53–59, 2021, [Online]. Available: https://www.infeb.org/index.php/infeb/article/view/66

[4] S. Ufriani, Jasmir, and Y. Arviita, "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, pp. 342–350, 2023, [Online]. Available: https://ejournal.unama.ac.id/index.php/jakakom/article/view/726

[5] F. Febriansyah and S. Muntari, "Penerapan algoritma k-means untuk klasterisasi penduduk miskin pada kota pagar alam," *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 8, no. 1, pp. 66–77, 2023, [Online]. Available: https://ejournal.uin-suka.ac.id/saintek/JISKA/article/view/3787

[6] S. Amaliyah, Jasmir, and S. Rianti, "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, [Online]. Available: https://ejournal.unama.ac.id/index.php/jakakom/article/view/802

[7] K. Anam, D. Sudrajat, and D. A. Kurnia, "Analisis Segmentasi Pelanggan Menggunakan Metode K-Means Clustering," *J. ICT Inf. Commun. Technol.*, vol. 21, no. 2, pp. 273–278, 2022, [Online]. Available: http://ejournal.ikmi.ac.id/index.php/jict-ikmi/article/view/9

[8] A. Permadi and Y. A. Wijaya, "Pengelompokkan dataset bus menggunakan algoritma k-means," *Informatics Educ. Prof.*, vol. 7, no. 2, pp. 138–152, 2023, [Online]. Available: http://ejournal-binainsani.ac.id/index.php/ITBI/article/view/2259

[9] Y. A. Wijaya, D. A. Kurniady, E. Setyanto, W. S. Tarihoran, D. Rusmana, and R. Rahim, "Davies bouldin index algorithm for optimizing clustering case studies mapping school facilities," *TEM J*, vol. 10, no. 3, p. 1099-1103, 2021, [Online]. Available: https://www.ceeol.com/search/article-detail?id=977772

[10] W. Rosida and Y. A. Wijaya, "Klasterisasi Penyakit HIV/AIDS di Jawa Barat Menggunakan Algoritma K-Means Clustering," *Blend Sains J. Tek.*, vol. 1, no. 4, pp. 1–10, 2023, [Online]. Available: https://jurnal.ilmubersama.com/index.php/blendsains/article/view/235

[11] N. Nurahman and J. Susanto, "Klasterisasi data penerima bantuan langsung tunai menggunakan algoritma k-means," *JURIKOM (Jurnal Ris. Komputer)*, vol. 10, no. 2, pp. 461–470, 2023, [Online]. Available: http://ejurnal.stmik-budidarma.ac.id/index.php/jurikom/article/view/5807

[12] A. Asmana, Y. A. Wijaya, and M. Martanto, "Clustering data calon siswa baru menggunakan metode k-means di sekolah menengah kejuruan wahidin kota cirebon," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 6, no. 2, pp. 552–559, 2022, [Online]. Available: https://ejournal.itn.ac.id/index.php/jati/article/view/5236

[13] D. K. Utami, N. Irawati, and S. Sumantri, "Analisis metode k-means pada clustering penerimaan bantuan PKH desa pulau rakyat tua," *Sist. J. Sist. Inf.*, vol. 12, no. 3, pp. 953–961, 2023, [Online]. Available: http://sistemasi.ftik.unisi.ac.id/index.php/stmsi/article/view/3236

[14] A. Rotondo and F. Quilligan, "Evolution paths for knowledge discovery and data mining process models," *SN Comput. Sci.*, vol. 1, no. 109, pp. 2–19, 2020, doi: 10.1007/s42979-020-0117-6.