IMPLEMENTASI DATA MINING CLUSTERING DALAM MENGELOMPOKKAN PENDUDUK PENYANDANG DISABILITAS MENGGUNAKAN ALGORITMA K-MEANS

Puji Lestari, Nana Suarna, Willy Prihartono
Teknik Informatika, STMIK IKMI Cirebon
Jalan Perjuangan No.10B, Karyamulya, Kec. Kesambi, Kota Cirebon, Indonesia
pujilestari04042004@gmail.com

ABSTRAK

Penyandang disabilitas merujuk kepada individu yang mengalami keterbatasan fisik atau mental, yang dapat menghambat mereka dalam menjalankan kegiatan sehari-hari secara normal. Tidak semua orang dilahirkan dengan kondisi fisik atau mental yang sempurna, menyebabkan sebagian dari mereka mungkin merasa kurang percaya diri atau rendah diri dalam berinteraksi sosial. Permasalahan yang dihadapi penyandang disabilitas merupakan isu kompleks, karena adanya keterbatasan pada fungsi tubuh yang tidak optimal dapat menimbulkan dampak sosial yang signifikan. Ketidakmampuan ini bisa menjadi penghalang dalam melaksanakan kegiatan harian mereka. Orang dengan kebutuhan khusus, atau disabilitas, memiliki karakteristik hidup yang berbeda dan memerlukan pelayanan khusus untuk menjamin hak mereka terhadap kehidupan yang lebih baik. Oleh karena itu, untuk memberdayakan penyandang disabilitas, penelitian perlu dilakukan berdasarkan kecamatan di kabupaten/kota di Jawa Barat. Hal ini bertujuan untuk mengidentifikasi kelompok-kelompok wilayah dengan tingkat keberagaman penyandang disabilitas yang tinggi, sedang, dan rendah. Penelitian ini menggunakan metode data mining Clustering dengan algoritma K-Means untuk menganalisis kasus penyandang disabilitas di Kabupaten/Kota di Provinsi Jawa Barat. Metode clustering bertujuan untuk mengelompokkan data berdasarkan karakteristik yang serupa. Jumlah cluster ditentukan menggunakan nilai Davies Bouldin Index (DBI). Hasil penelitian menunjukkan adanya 2 cluster optimal, di mana Cluster 1 terdiri dari 27 kabupaten/kota dan Cluster 2 terdiri dari 2 kabupaten/kota, dengan nilai DBI sebesar 0.121

Kata kunci : *Data Mining, Algoritma K-Means, Clustering, Davies Bouldin Index*

1. PENDAHULUAN

Dalam beberapa tahun terakhir, peningkatan jumlah penduduk dengan berbagai jenis disabilitas di Jawa Barat menjadi sebuah perhatian serius. Disabilitas mencakup berbagai kondisi fisik, mental, dan sensorik yang memerlukan perhatian khusus untuk memahami kebutuhan dan pelayanan yang diperlukan oleh setiap individu. Mengelompokkan individu dengan disabilitas berdasarkan jenis kategori disabilitas dapat menjadi langkah awal yang penting untuk menyusun strategi pelayanan yang lebih efektif dan terfokus. Pemahaman mendalam tentang berbagai jenis disabilitas akan membantu pemerintah dan lembaga terkait dalam mengembangkan program-program inklusif yang dapat meningkatkan kualitas hidup dan partisipasi sosial bagi semua orang di Jawa Barat. Strategi ini diharapkan dapat menghasilkan perubahan positif dalam memberikan aksesibilitas dan kesetaraan bagi individu dengan disabilitas, yang akan menghasilkan lingkungan yang lebih inklusif bagi semua lapisan masyarakat.

Masalah yang dihadapi dalam penelitian ini adalah Penting untuk memahami bahwa setiap jenis disabilitas memiliki karakteristik dan kebutuhan yang unik. Maka, tujuan dari penelitian yang dilakukan ini adalah melakukan pengelompokan (*clustering*) jumlah individu dengan disabilitas di Jawa Barat berdasarkan kategori disabilitas menggunakan algoritma k-means. Metode ini dipilih karena kemampuannya untuk mengelompokkan data menjadi beberapa kelompok yang saling seragam berdasarkan karakteristik tertentu. Selama proses ini, setiap orang dengan disabilitas akan ditempatkan dalam kelompok yang paling sesuai dengan kondisi mereka. Ini memungkinkan peneliti dan pihak terkait untuk lebih memahami kebutuhan khusus masing-masing kelompok. Karena itu, diharapkan bahwa penelitian ini dapat memberikan dasar yang lebih kuat untuk merancang program pelayanan yang lebih khusus dan efisien bagi individu dengan disabilitas di Jawa Barat. Diharapkan juga bahwa temuan pengelompokan ini akan berfungsi sebagai dasar untuk pembuatan kebijakan yang lebih inklusif yang memungkinkan individu dengan disabilitas berpartisipasi lebih aktif dalam berbagai aspek masyarakat.

Penelitian sebelumnya yang dilakukan oleh Wildani Putri, M.Afdal yang berjudul "Penerapan Algoritma K-Means Untuk Pengelompokan Data Penyandang Disabilitas Di Kabupaten Rokan Hilir". menyimpulkan Klaster yang ditemukan dalam penelitian ini terdiri dari tiga bagian, yaitu cluster_0 yang mencakup kelompok kecamatan dengan tingkat disabilitas rendah, cluster_2 yang mencakup kelompok kecamatan dengan tingkat disabilitas sedang, dan cluster_1 yang merupakan kelompok kecamatan dengan jumlah penyandang disabilitas terbanyak. Dari total 18 kecamatan di Kabupaten Rokan Hilir, terdapat satu kecamatan yang termasuk

dalam cluster tingkat tinggi, yaitu kecamatan Bangko, yang memiliki jumlah penyandang disabilitas terbanyak. Sementara itu, tiga kecamatan, yaitu Rimba Melintang, Rantau Kopar, dan Pujud, termasuk dalam cluster tingkat sedang, dan ke-14 kecamatan lainnya termasuk dalam cluster tingkat rendah. Hasil ini menunjukkan bahwa metode K-Means dapat efektif dalam mengelompokkan data penyandang disabilitas di Kabupaten Rokan Hilir. Validasi dengan menggunakan metode Davies Bouldin menghasilkan nilai sebesar 0,063, menunjukkan tingkat keberhasilan yang baik dalam pengelompokan data tersebut. [1]

Penelitian sebelumnya menurut Kristiyo Indriadi Wardoyo, Maryaningsih, Jhoanne Fredricka yang berjudul "Klasterisasi Siswa Penyandang Disabilitas Berdasarkan Tingkat Tunagrahita Menggunakan Algoritma K-Means". Merangkum hasil klasterisasi siswa yang dilakukan dengan menggunakan sampel data sebanyak 73 siswa melalui aplikasi, ditemukan bahwa 47,9% siswa termasuk dalam Cluster 1 (tingkat tunagrahita ringan), 37% siswa termasuk dalam Cluster 2 (tingkat tunagrahita sedang), dan 15,1% siswa termasuk dalam Cluster 3 (tingkat tunagrahita berat). Hasil ini diperoleh melalui perhitungan nilai euclidean distance dengan memilih jarak terdekat. [2]

Penelitian yang dilakukan menurut Ade Indah Sari, Heru Satria Tambunan, Widodo Saputra, Irfan Sudahri Damanik, Ilham Syahputra Saragih yang berjudul "Implementasi Algoritma K-Means Dalam Pengelompokan Penyandang Disabilitas Menurut Kecamatan Kabupaten Simalungun". Merangkum, data yang diinput merupakan jumlah penyandang disabilitas dari tahun 2008 hingga 2017, yang terbagi dalam 31 kecamatan dan diklasifikasikan ke dalam tiga klaster, yaitu klaster tinggi, sedang, dan rendah. Melalui perhitungan k-means, ditemukan bahwa terdapat 8 kecamatan dalam klaster tinggi, 15 kecamatan dalam klaster sedang, dan 8 kecamatan dalam klaster rendah. Proses implementasi dilakukan dengan menggunakan aplikasi RapidMiner 5.3 untuk membantu menghasilkan nilai yang akurat. [3]

Tujuan penelitian ini adalah untuk mengidentifikasi dan mengelompokkan individu dengan disabilitas di Jawa Barat berdasarkan kategori disabilitas menggunakan algoritma k-means, dengan fokus pada pemahaman distribusi, karakteristik, dan potensi strategi pelayanan metode yang lebih efisien untuk meningkatkan mutu hidup mereka.

Metode penelitian ini mengggunakan aplikasi Rapidminner dalam memanfaatkan untuk mengimplementasikan algoritma k-means clustering. Proses penelitian akan dimulai dengan pemrosesan data individu disabilitas di Jawa Barat, kemudian dilanjutkan dengan penggunaan k-means untuk mengelompokkan individu berdasarkan kategori disabilitas. Parameter k (jumlah klaster) akan dioptimalkan melalui nilai DBI untuk memastikan pengelompokan yang optimal. Setelah pengelompokan selesai, akan dilakukan analisis karakteristik masing-masing kelompok untuk mendapatkan pemahaman yang lebih dalam tentang distribusi dan kebutuhan individu dalam setiap kategori disabilitas. Harapan dari hasil penelitian ini adalah dapat memberikan pemahaman yang berharga dalam perancangan strategi pelayanan yang lebih efektif bagi masyarakat disabilitas di Jawa Barat.

Penggunaan algoritma k-means dalam penelitian ini menghasilkan temuan bahwa, individu dengan disabilitas di Jawa Barat dapat dibagi menjadi kelompok-kelompok yang memiliki karakteristik serupa berdasarkan kategori disabilitas. Analisis distribusi jumlah individu di setiap kelompok memberikan gambaran yang lebih terperinci tentang kebutuhan dan profil masing-masing kelompok. Selain itu, optimasi parameter k melalui nilai DBI untuk menghasilkan jumlah cluster yang optimal, memastikan validitas pengelompokan. Hasil ini diharapkan dapat menjadi dasar untuk menyusun strategi pelayanan yang lebih terfokus dan inklusif bagi setiap kelompok disabilitas di Jawa Barat, meningkatkan aksesibilitas, pendidikan, dan dukungan sosial sesuai dengan karakteristik unik masing-masing kelompok.

2. LANDASAN TEORI

Landasan teori adalah kerangka konseptual atau teoritis yang dipergunakan sebagai acuan pokok dalam suatu penelitian. Dalam situasi penelitian ini, sejumlah dasar teori yang bisa dijadikan landasan melibatkan hal-hal berikut:

2.1. Data Mining

Data mining atau yang sering disebut KDD (*Knowledgw Discovery in Database)* adalah proses yang mencakup pengumpulan dan pemanfaatan riwayat data untuk menemukan pola dan hubungan dalam kumpulan data yang besar. Penambangan data menghasilkan informasi yang dapat digunakan untuk membuat keputusan di masa mendatang. Salah satu teknik yang dikenal dalam penambangan data adalah pengelompokan atau pengelompokan data (clustering).[4]

2.2. Clustering

Cluster adalah kumpulan data yang dikelompokkan karena adanya kesamaan antara sebagian data di dalamnya, namun juga memiliki beberapa perbedaan dengan kelompok lain. Berbeda dengan klasifikasi, pengelompokan melibatkan variabel objek dalam pembentukan kelompoknya. Pengelompokan tidak bertujuan untuk Menyelesaikan tugas klasifikasi melibatkan kegiatan seperti meramalkan nilai dari suatu variabel objek. Sebaliknya, algoritma clustering berupaya memisahkan kumpulan informasi yang ada menjadi kelompok-kelompok yang homogen atau sejenis. Tingkat kesamaan data di dalam suatu kelompok akan menghasilkan nilai yang semakin besar, sementara perbedaan antar kelompok akan menghasilkan nilai yang semakin kecil.[5]

2.3. Davies Bouldin Index

Prinsip mendasar dalam pendekatan pengukuran Davies-*Bouldin Index* (DBI) adalah untuk maksimalkan jarak antara pusat-pusat kluster dan sebaliknya, minimalkan jarak antara pusat-pusat kluster. Semakin kecil nilai yang dihasilkan dari perhitungan DBI, semakin optimal skema pengelompokan yang dicapai. [5]

2.4. Penelitian Terdahulu

Penelitian pertama oeh Ade Indah Sari, Heru Satria Tambunan, Widodo Saputra, Irfan Sudahri Damanik, Ilham Syahputra Saragih yang berjudul "Implementasi Algoritma K-Means Dalam Pengelompokan Penyandang Disabilitas Menurut Kecamatan Kabupaten Simalungun". Penelitian ini bertujuan untuk mengklasifikasikan wilayah penyandang disabilitas berdasarkan distrik, menggunakan teknik data mining dengan metode K-Means *clustering*. Metode K-Means adalah salah satu teknik pengelompokan yang berfungsi untuk membagi dataset menjadi beberapa kelompok. Sumber data yang digunakan dalam penelitian ini berasal dari Badan Pusat Statistik (BPS) Kabupaten Simalungun. Data yang dimasukkan mencakup jumlah penyandang disabilitas dari tahun 2008 hingga 2017, dengan 31 kecamatan yang terbagi menjadi 3 klaster, yaitu klaster tinggi, sedang, dan rendah. Hasil perhitungan K-Means menunjukkan bahwa terdapat 8 kecamatan dalam klaster tinggi, 15 kecamatan dalam klaster sedang, dan 8 kecamatan dalam klaster rendah. Proses implementasi menggunakan aplikasi RapidMiner 5.3 dilakukan untuk membantu mendapatkan nilai yang akurat.[3]

Penelitian kedua oleh Kristiyo Indriadi Wardoyo, Maryaningsih, Jhoanne Fredricka yang berjudul "Klasterisasi Siswa Penyandang Disabilitas Berdasarkan Tingkat Tunagrahita Menggunakan Algoritma K-Means". Tujuan penelitian ini menempatkan para siswa dengan kebutuhan khusus tunagrahita dibagi menjadi dua tingkat, yaitu ringan dan sedang, dengan melakukan observasi terhadap nilai IQ dan pencapaian akademik siswa. Langkah ini diambil untuk meningkatkan efisiensi dan efektivitas proses belajar mengajar. Meskipun demikian, menempatkan siswa berkebutuhan khusus tunagrahita ke dalam kelas yang tepat dapat menjadi tantangan bagi pihak sekolah, karena hingga saat itu belum ada aplikasi yang membantu pengelompokan data siswa tersebut. Oleh karena itu, diperkenalkan sebuah aplikasi klasterisasi siswa penyandang disabilitas di Sekolah Luar Biasa Negeri 1 Kota Bengkulu, yang menggunakan algoritma K-Means. Aplikasi ini dikembangkan menggunakan bahasa pemrograman *Visual Basic Net* dan *Database SQL Server*. Hasil klasterisasi siswa, yang didapatkan melalui sampel data dari 73 siswa menggunakan aplikasi, menunjukkan bahwa 47,9% termasuk dalam *Cluster* 1 (tingkat tunagrahita ringan), 37% termasuk dalam *Cluster* 2 (tingkat tunagrahita sedang), dan 15,1% termasuk dalam *Cluster* 3 (tingkat tunagrahita berat), berdasarkan perhitungan nilai *euclidean distance* dengan mengambil jarak terdekat.[2]

Penelitian ketiga oleh M. Atiqul Mutaqin, Gunawan, Wresti Andriyani yang berjudul "Klasterisasi Data Disabilitas Menggunakan Algoritma K-Means". Tujuan Penelitian ini bertujuan untuk mengevaluasi hasil pengelompokan kelas siswa dengan disabilitas menggunakan metode K-Means clustering. Teknik data mining, khususnya metode K-Means *clustering*, digunakan dalam penelitian ini untuk membagi dataset menjadi beberapa kelompok. Data yang digunakan berasal dari Sekolah Luar Biasa N Kota Tegal, mencakup sampel kelas siswa disabilitas pada tahun 2021-2022 yang dikategorikan ke dalam tiga klaster, yaitu klaster ringan, sedang, dan berat, berdasarkan penilaian *Intelligence Quotient* (IQ). Dari hasil perhitungan K-Means, ditemukan 11 siswa termasuk dalam klaster ringan, 9 siswa termasuk dalam klaster sedang, dan 10 siswa termasuk dalam klaster berat. Proses pengelompokan dilakukan menggunakan *Microsoft Excel*, dan aplikasi RapidMiner digunakan untuk membantu mendapatkan nilai yang akurat.[6]

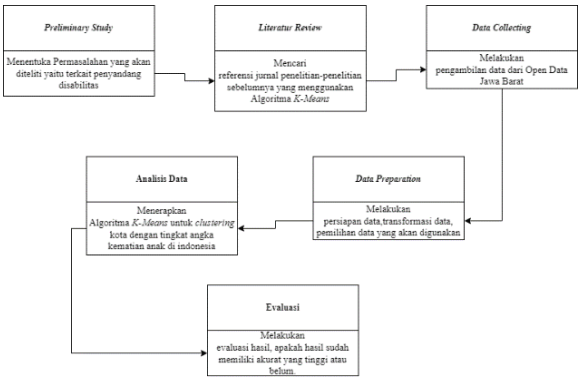
Penelitian keempat oleh Wildani Putri, M.Afdal yang berjudul "Penerapan Algoritma K-Means Untuk Pengelompokan Data Penyandang Disabilitas Di Kabupaten Rokan Hilir". Tujuan penelitian ini yaitu mengelompokan menyusun data mengenai penyandang disabilitas di Kabupaten Rokan Hilir, memberikan informasi relevan kepada pemerintah dan instansi terkait untuk mendukung pengambilan keputusan yang tepat. Metode pengelompokkan yang diterapkan dalam penelitian ini menggunakan algoritma K-Means. Klaster yang dihasilkan terdiri dari tiga kelompok, yaitu *cluster*_0 yang mencakup kecamatan dengan tingkat disabilitas rendah, *cluster*_2 yang mencakup kecamatan dengan tingkat disabilitas sedang, dan *cluster*_1 yang mencakup kecamatan dengan jumlah penyandang disabilitas terbanyak. Dari total 18 kecamatan di Kabupaten Rokan Hilir, satu kecamatan termasuk dalam klaster tingkat tinggi, yaitu kecamatan Bangko, sementara tiga kecamatan termasuk dalam klaster tingkat sedang, yaitu kecamatan Rimba Melintang, Rantau Kopar, dan Pujud, serta 14 kecamatan lainnya termasuk dalam klaster tingkat rendah. Hasil penelitian menunjukkan bahwa metode K-Means efektif dalam mengelompokkan data penyandang disabilitas di Kabupaten Rokan Hilir. Validasi dengan menggunakan metode Davies Bouldin menghasilkan nilai sebesar 0,063, menandakan keberhasilan yang baik dalam pengelompokan data tersebut.[1]

Penelitian kelima oleh Fitriana Harahap yang berjudul "Perbandingan Algoritma K Means dan K Medoids Untuk Clustering Kelas Siswa Tunagrahita". Tujuan membandingkan hasil pengelompokan kelas siswa tunagrahita menggunakan metode K-Means dan K-Medoids Clustering. Kedua metode menghasilkan tiga klaster. Dengan menggunakan metode K-Means

Clustering, ditemukan 8 siswa tunagrahita ringan, 14 siswa tunagrahita sedang, dan 14 siswa tunagrahita berat. Sementara itu, metode K-Medoids *Clustering* menunjukkan adanya 7 siswa tunagrahita ringan, 19 siswa tunagrahita sedang, dan 10 siswa tunagrahita berat. Validasi dilakukan dengan menggunakan nilai Davies Bouldin Index (DBI), di mana nilai DBI untuk K-Means adalah 0,161, dan untuk K-Medoids adalah 0,281. Dengan demikian, dapat disimpulkan bahwa pengelompokan menggunakan metode K-Means Clustering memberikan hasil yang lebih baik dibandingkan dengan metode K-Medoids *Clustering*, karena menghasilkan nilai DBI yang lebih kecil, yaitu 0,16.[7]

3. METODE PENELITIAN

Peneliti menggunakan jenis penelitian kuantitatif algoritma (K-Means). Metode penelitian yang dilakukan dengan cara ini memiliki tahapan *Preliminary Study, Literatur Review, data Collecting, Data Preparation, Analisis Data, dan Evaluasi* dengan penjelasan sebagai berikut :



Gambar 1. Metode Penelitian

Berdasarkan gambar diatas, dapat dideskripsikan metode penelitian seperti tabel dibawah ini.

Tabel 1. Deskripsi Aktivitas Metode Penelitian

Tahapan	Aktivitas	Deskripsi Aktivitas
Preliminary Study	Analisis Masalah	Menentukan permasalahan yang akan yang akan diteliti yaitu terkait penyandang disabilitas.
Literatur Review	Melakukan Studi Literatur	Mencari referensi jurnal penelitian-penelitian sebelumnya yang menggunakan Algoritma K-Means
Data Collecting	Pengumpulan Data	Melakukan pengambilan data dari Open Data Jawa Barat
Data Preparation	Persiapan data dan pemilihan data	Melakukan persiapan data,transformasi data, pemilihan data yang akan digunakan
Analisis Data	Melakukan penerapan Algoritma K-Means	Menerapkan Algoritma K-Means untuk clustering kota dengan tingkat angka kematian anak di indonesia
Evaluasi	Evaluasi Hasil	Melakukan evaluasi hasil, apakah hasil sudah memiliki akurat yang tinggi atau belum.

4. HASIL DAN PEMBAHASAN

4.1. Hasil Penelitian

Penelitian ini mencapai beberapa temuan signifikan dan memberikan pemahaman tentang pengelompokan data dalam konteks penelitian ini. Penelitian ini melibatkan metode analisis dengan langkah-langkahnya *Data Selection, Data Prepocessing, Data Transformation, Data Mining, Evaluation* dan Knowledge. Hasil clustering ini memberikan gambaran yang lebih jelas tentang pola dan relasi diantara data sebelumnya yang kompleks.

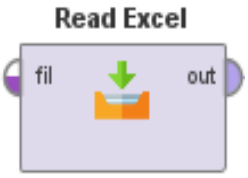
4.1.1. Data

Data Data yang digunakan penelitian ini diperoleh melalui sumber Open Data Jabar dengan format tampilan excel yang terstruktur dengan judul Data Penyandang Disabilitas Tahun 2015-2022. Data ini terdiri dari 1.296 dataset dan memiliki 6 atribut yaitu Nama Provinsi, Nama, Kabupaten/Kota, Kategori Disabilitas, Jumlah Penduduk, Satuan, Tahun beserta penjelasannya dapat dilihat pada Tabel 2. Penyandang Disabilitas berikut.

Tabel 2. Penyandang Disabilitas

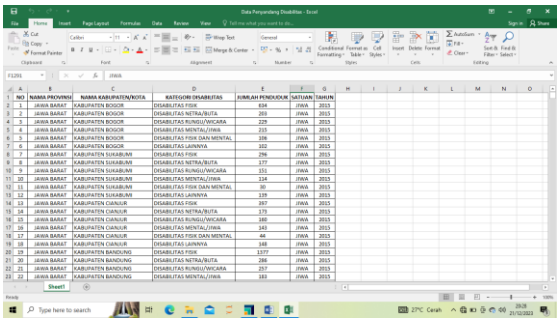
No	Tipe	Atribut	Keterangan
1	Nama Provinsi	*Polynominal*	Untuk atribut yang memiliki lebih dari satu kategori
2	Nama Kabupaten/Kota	*Polynominal*	Untuk atribut yang memiliki lebih dari satu kategori
3	Kategori Disabilitas	*Polynominal*	Untuk atribut yang memiliki lebih dari satu kategori
4	Jumlah Penduduk	*Integer*	Untuk file dengan nilai integer (angka tanpa koma)
5	Satuan	*Polinominal*	Untuk atribut yang memiliki lebih dari satu kategori
6	Tahun	*Polinominal*	Untuk atribut yang memiliki lebih dari satu kategori

Dari Tabel 2, pada aplikasi RapidMiner untuk membaca data tersebut menggunakan operator *Read Excel*. Dapat dilihat pada Gambar 2



Gambar 2. Operator *Read Excel* Pada RapidMiner

Operator *Read Excel* dalam RapidMiner Digunakan untuk mengimpor data dari file *Excel* ke dalam RapidMiner. Dengan menggunakan operator ini, Anda dapat memuat file Excel yang berisi data yang telah dipilih. Ini adalah tampilan dari berkas dalam *Microsoft Excel* :

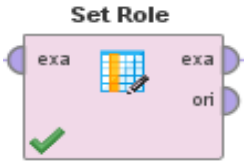


Gambar 3. Data Penyandang Disabilitas Dalam File Excel

Gambar 3 merupakan data file yang digunakan dalam tahap read excel. Data tersebut terdapat atribut Nama Provinsi, Nama, Kabupaten/Kota, Kategori Disabilitas, Jumlah Penduduk, Satuan, dan Tahun dari rentang tahun 2015-2022 dengan total keseluruhan data berjumlah 1.296 data.

4.1.2. Data Selection

Tahapan selection digunakan untuk menyeleksi atau memilih data yang akan diolah. Data yang diolah yaitu data penyandang disabilitas. Tahapan ini dilakukan untuk memilih data yang akan diproses di RapidMiner, tahapan ini dilakukan di *Microsoft Excel* dan di RapidMiner. Pada *Microsoft Excel* menghapus data yang bernilai *null* dan atribut yang tidak diperlukan pada file penyandang disabilitas tahun 2015-2022. Pada RapidMiner menggunakan operator Set Role untuk menentukan id pada dataset, dapat dilihat pada Gambar 4.



Gambar 4. Operator *Set Role* Pada RapidMiner

Parameter pada operator *Set Role* yang digunakan dapat dilihat pada Gambar 4 dengan menggunakan atribut name yaitu No menjadi target role yaitu id terlihat pada Tabel 3 berikut.

Tabel 3. Parameter operator Set Role

No.	Parameter	Isi
1.	*attribute name*	*No*
2.	*target role*	Id

Dari hasil pembacaan operator Set Role didapat informasi pada Tabel 4 berikut.

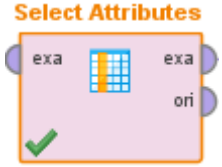
Tabel 4. Hasil Pembacaan Operator Set Role

No	Uraian	Keterangan
1	No	*Id*
2	Nama Provinsi	*Polynominal*
3	Nama Kabupaten/Kota	*Polynominal*
4	Kategori Disabilitas	*Integer*
5	Jumlah Penduduk	*Polynominal*
6	Satuan	*Polynominal*
7	Tahun	*Polynominal*

Karena dataset sebelum dimasukkan ke RapidMiner sudah terseleksi beberapa di Excel maka yang tadinya 7 atribut menjadi 6 atribut saat dimasukkan RapidMiner, dikarenakan mempunyai nilai yang sama dan nilai null.

4.1.3. Data preprocessing

Sebelum melaksanakan Data Mining, langkah awal yang penting adalah menjalani serangkaian proses, termasuk membersihkan duplikasi data, memeriksa inkonsistensi dalam data, dan memperbaiki kesalahan yang mungkin muncul, seperti kesalahan dalam mencetak (tipografi). Di samping itu, proses enrichment data dilaksanakan untuk memperluas data yang telah ada dengan menyertakan informasi tambahan yang relevan dan diperlukan dalam Kontak Penemuan Pengetahuan (KDD), termasuk data atau informs eksternal yang dapat meningkatkan kualitas dan keberagaman data. Dalam tahap ini diperoleh data sebanyak 1.296 record pada tahun 2015-2022 dengan 6 atribut yang akan digunakan yaitu Nama Provinsi, Nama Kabupaten/Kota, Kategori Disabilitas, Jumlah Penduduk, Satuan, dan Tahun. Berikut adalah proses processing pada RapidMiner menggunakan operator Select Attributes dapat dilihat pada Gambar 5 berikut.



Gambar 5. Operator *Select Attributes* pada RapidMiner

Operator *Select Attributes* dalam RapidMiner digunakan untuk memilih, menghapus, atau mengatur atribut dalam dataset. Ini memungkinkan mengubah

tipe data, dan melakukan transformasi lainnya pada atribut sebelum proses analisis data. Parameter yang digunakan pada operator *Select Attributes* dapat dilihat pada Tabel 5 berikut.

Tabel 5. Parameter Pada Operator *Select Attributes*

No.	Parameter	Isi
1.	*Type*	*include attributes*
2.	*attribute filter type*	*a subset*
3.	*select subset*	*Nama Provinsi*
		Nama Kabupaten/Kota
		Kategori Disabilitas
		Jumlah Penduduk
		Satuan
		Tahun

Hasil dari penggunaan operator *Select Attributes* diperoleh informasi pada Tabel 6 berikut.

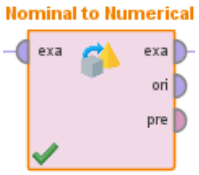
Tabel 6. Hasil Operator i

No	Nama Atribut	Jenis Data	Missing
1	Nama Provinsi	*Polynominal*	0
2	Nama Kabupaten/Kota	*Polynominal*	0
3	Kategori Disabilitas	*Integer*	0
4	Jumlah Penduduk	*Polynominal*	0
5	Satuan	*Polynominal*	0
6	Tahun	*Integer*	0

Berdasarkan hasil statistik dari dataset dengan 6 atribut, seperti yang terlihat pada Tabel 6, dapat disimpulkan bahwa tidak ada atribut yang mengandung nilai yang hilang. Untuk mengevaluasi konsistensi dataset, dilakukan pemeriksaan langsung pada setiap rekord, dan hasilnya menunjukkan bahwa dataset tersebut konsisten dalam nilai-nilainya.

4.1.4. Data Transformation

Pada tahap ini peneliti merubah data non-numeric menjadi *numeric*. Tahap penelitian ini menggunakan operator *Nominal to Numerical* dapat ditampilkan pada Gambar 6 berikut.



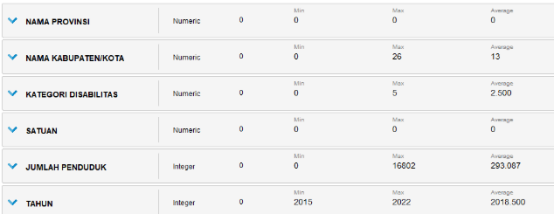
Gambar 6. Operator *Nominal to Numerical* pada RapidMiner

Parameter yang digunakan pada operator Nominal to Numerical yaitu pada attribute *filter type* memilih *subset* dan *attributes* yang akan dijadikan Numerik yaitu atribut Nama Provinsi, Nama Kabupaten/Kota, Kategori Disabilitas, dan Satuan dapat dilihat pada Tabel 7 berikut.

Tabel 7 Parameter pada Operator *Nominal to Numerical*

No.	Parameter	Isi
1.	attribut filter type	*Subset*
2.	coding type	*unique integers*
3.	Attributes	*Select Attributes*
		Nama Provinsi
		Nama Kabupaten/Kota
		Kategori Disabilitas
		Satuan

Dikarenakan atribut-atribut sebelumnya sudah mengalami selection dan hanya terpilih 4 atribut, dan atribut No menjadi id, atribut Jumlah Penduduk dan atribut Satuan sudah berbentuk numerik. Maka hanya atribut Nama Provinsi, Nama Kabupaten/Kota, Kategori Disabilita, dan Satuan yang berbentuk polynominal diubah menjadi numerik. Hasil dari penggunaan operator *Nominal to Numerical* terlihat pada Gambar 7 Berikut.



Gambar 7. Hasil Penggunaan Operator *Nominal to Numerical*

4.1.5. Data Mining

Langkah ini berfungsi untuk menentukan metode atau teknik yang paling sesuai dalam mengidentifikasi pola atau informasi menarik dalam data yang telah dipilih. Dalam penelitian ini, proses data mining diimplementasikan dengan menggunakan metode *Clustering*, dengan memanfaatkan Algoritma K-Means. Pada tahap ini juga peneliti melakukan proses clustering, menguji dan mengevaluasi hasil dari proses tersebut dalam platform RapidMiner. Operator yang digunakan adalah operator *Clustering* (K-Means) terlihat pada Gambar 8 berikut.

Gambar 8. Operator *Clustering* pada RapidMiner

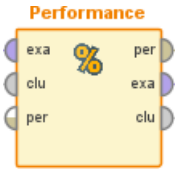
Parameter yang digunakan pada k-Means atau *Clustering* dapat dilihat pada Tabel 8 berikut.

Tabel 8. Parameter pada Operator *Clustering*

No.	Parameter	Isi
1	add *cluster* attribute	Digunakan
2	k	2
3	*max runs*	10

No.	Parameter	Isi
4	*determine good star values*	Digunakan
5	*measure types*	*MixedMeasured*
6	*Mixed measured*	*MixedMeasured*
7	*Max optimization*	100

Pada saat menggunakan operator clustering kita perlu mengetahui berapa klaster yang terbaik dalam dataset ini, lalu dalam penelitian ini menggunakan performance *Davies Bouldin Index* (DBI) terlihat pada Gambar 9 berikut.



Gambar 9. Operator Performance untuk *Davies Bouldin Index* pada RapidMiner

Parameter yang digunakan pada operator *Performance* terlihat pada Tabel 9 berikut.

Tabel 9. Parameter pada Operator *Performance*

No.	Parameter	Isi
1	*Main criterion*	*Davies Bouldin*
2	*Normalize*	Digunakan
3	*Maximize*	Digunakan

Pada operator Performance, parameter normalize dan *maximize* digunakan karena jika hanya mencentang *normalize* saja akan menjadi minus, dan menggunakan atau mencentang *maximize* hanya menghilangkan minusnya saja tidak merubah *cluster* atau lainnya.

Tahapan pada clustering dapat dilihat pada Gambar 9 berikut.

Gambar 9. Tahapan *Clustering* pada RapidMiner

4.1.6. Evaluation

Teori evaluasi yang digunakan dalam penelitian ini adalah Davies Bouldin Index (DBI). Dalam Tabel 10, terdapat jumlah cluster yang telah dimodelkan dan nilai DBI yang dihasilkan.

Tabel 10. Nilai DBI Setiap Jumlah Cluster

Cluster	Nilai DBI
K2	0.121
K3	0.133
K4	0.237
K5	0.220
K6	0.209
K7	0.202
K8	0.186
K9	0.184
K10	0.187

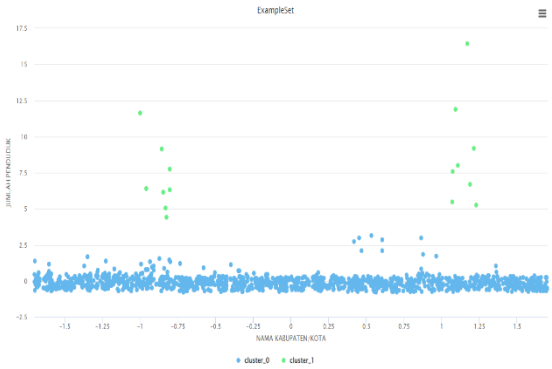
Pada Tabel 10 adalah hasil dari perhitungan nilai Davies Bouldin Indeks menggunakan aplikasi RapidMinner melalui percobaan K2, K3, K4, K5, K6, K7, K8, K9 hingga K10. Mengacu pada prinsip DBI, nilai yang dianggap baik dalam data mining adalah semakin kecil atau mendekati nol. Dalam Tabel 10 nilai DBI yang paling baik adalah K2 atau cluster 2, dengan nilai DBI yang mendekati nol yaitu 0.121. Oleh karena itu, pengelompokkan dilakukan dengan menggunakan 2 cluster.

4.2. Pembahasan

Penelitian ini dilakukan dengan tujuan utama untuk mengelompokkan dataset Penyandang Disabilitas menggunakan algoritma K-Means. Tujuan spesifik mengidentifikasi pola atau kesamaan berdasarkan atribut-atribut terpilih serta mengevaluasi struktur dan jarak antar cluster yang dihasilkan. Hasil pengelompokkan data Penyandang Disabilitas menggunakan algoritma K-Means telah diinterpretasikan dengan mempertimbangkan temuan yang sesuai dengan konsep, teori, dan penelitian terdahulu yang relevan dengan lingkup penelitian ini. Interpretasi ini memberikan wawasan yang mendalam tentang struktur klaster dan signifikansi temuan yang dihasilkan.

4.2.1. Pengelompokkan Menggunakan Teknik Clustering

Penelitian ini berhasil mengimplementasikan teknik clustering, dari proses tersebut menghasilkan visualisasi data dalam bentuk *scatter Bubble* pada Gambar 10 berikut.

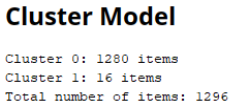


Gambar 10. *Cluster* Visualisasi *Scatter Bubble*

Gambar di atas menggambarkan persebaran klaster dalam bentuk scatter bubble yang dihasilkan dari proses pengelompokan penyandang disabilitas. Scatter berwarna biru mewakili *cluster* 0, sementara scatter berwarna hijau melambangkan *cluster* 1, dengan nilai klaster ditunjukkan pada sumbu x dan sumbu y. Kelompok pada *cluster* 0 dapat diklasifikasikan sebagai penyandang disabilitas dengan tingkat rendah, sedangkan *cluster* 1 dapat dikategorikan sebagai penyandang disabilitas dengan tingkat tinggi.

4.2.2. Klaster Yang Dihasilkan Oleh Algoritma K-Means

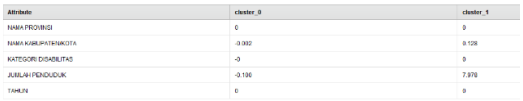
Algoritma K-Means Clustering dapat diterapkan untuk melakukan pengelompokkan data Penyandang Disabilitas dengan atribut Nama Provinsi, Nama Kabupaten/Kota, Kategori Disabilitas, Jumlah Penduduk, Satuan, dan Tahun dengan hasil Cluster Model dan jumlah anggota setiap cluster dterlihat pada Gambar 11 berikut.



Gambar 11. Hasil Jumlah Item dari 2 *Cluster*

Dari Gambar 11, menunjukkan bahwa data telah berhasil dikelompokkan menjadi 2 klaster yaitu Cluster 0 dengan jumlah 1280 anggota dan Cluster 1 berjumlah 16 anggota dari total keseluruhan sejumlah 1.296. Jumlah item yang berbeda menunjukkan adanya variasi dalam sifat atau karakteristik tersebut. Klaster dengan jumlah yang lebih tinggi menunjukkan adanya lebih banyak penyandang disabilitas.

Selanjutnya analisis nilai rata-rata centroid pada setiap cluster dari atribut yang telah ditentukan. Hasil dari nilai rata-rata atribut dapat dilihat dari Gambar 12 berikut.



Attribute	cluster_0	cluster_1
NAMA PROVINSI	0	0
NAMA KABUPATEN/KOTA	-0.002	0.128
KATEGORI DISABILITAS	-0	0
JUMLAH PENDUDUK	-0.100	7.970
TAHUN	0	0

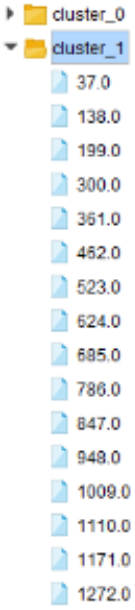
Gambar 4. 12 Hasil Rata-rata *Centroid* dari Setiap *Cluster*

Dari Gambar 12 dapat disimpulkan bahwa setiap cluster berdasarkan Penyandang Disabilitas dapat dianalisis sebagai berikut :

1 *Cluster* 0 : merupakan jumlah penyandang disabilitas rendah dengan anggota cluster berjumlah 27 kabupaten/kota yang meliputi Kabupaten Bogor, Kabupaten Sukabumi, Kabupaten Cianjur, Kabupaten Bandung, Kabupaten Garut Kabupaten Tasikmalaya, Kabupaten Ciamis, Kabupaten Kuningan, Kabupaten Cirebon, Kabupaten Majalengka, Kabupaten Sumedang, Kabupaten Indramayu, Kabupaten Subang, Kabupaten Purwakarta, Kabupaten Karawang, Kabupaten Bekasi, Kabupaten Bandung Barat, Kabupaten Pangandaran, Kota Bogor, Kota Sukabumi, Kota Bandung, Kota Cirebon, Kota Bekasi, Kota Depok, Kota Cimahi, Kota Tasikmalaya, Kota Banjar.

2 *Cluster* 1 : merupakan jumlah penyandang disabilitas yang tinggi dengan anggota cluster berjumlah 2 kabupaten/kota yang meliputi Kabupaten Ciamis dan Kota Bekasi.

   1) Dengan dihasilkannya cluster-cluster tersebut ditemukannya data penyanding disabilitas tinggi yaitu dengan nilai rata-rata atribut sebesar *Avg. within centroid distance*: 0.640 Cluster 1. Berikut beberapa keanggotaan pada cluster 1 atau data penyandang disabilitas tinggi yaitu pada Gambar 13 berikut.



Gambar 13. Penyandang Disabilitas Cluster 1

4.2.3. Klaster Yang Dihasilkan Oleh Algoritma K-Means

Selama penelitian, jarak antar klaster diukur dari titik pusat klaster terdekat dan terjauh untuk memahami distribusi atau perbedaan antar klaster berdasarkan atribut-atribut yang dianalisis dalam dataset Penyandang Disabilitas. Jumlah rata-rata dalam jarak centroid setiap cluster pada tabel 4.10.

Tabel 11 Jumlah rata-rata dalam Jarak *Centroid* Setiap *Cluster*

Cluster	Rata-rata Jarak Centroid
K Pusat	0.640
K0	0.616
K1	2.616

Dari Tabel 11 dapat dijelaskan K Pusat merupakan *centoid* yang mana nilai 0.640 merupakan nilai pusat dari kalkulasi perhitungan algoritma k-means. Setelah diperoleh nilai *centoid* maka selanjutnya jarak antar obyek dapat diketahui nilainya dengan K0 atau titik awal *cluster* senilai 0.016 dan K1 atau titik akhir *cluster* senilai 2.616.

5. KESIMPULAN DAN SARAN

Kesimpulan dari hasil penelitian yang telah dilakukan diperoleh hasil bahwa penggunaan algoritma K-Means dan metode *Knowledge Discovery in Database* (KDD) dalam pemodelan dan evaluasi menggunakan nilai *Davies Bouldin Index* (DBI) dapat mengoptimalkan jumlah klaster didapat klaster terbaik yaitu pada cluster 2 dengan nilai DBI 0.121. *Clustering* ini menghasilkan pengelompokkan data penyandang

disabilitas yang berdasarkan Nama Provinsi, Nama Kabupaten/Kota, Kategori Disabilitas, Jumlah Penduduk, Satuan, dan Tahun dengan kategori rendah, dan tinggi. Dari hasil *cluster* 0 merupakan jumlah penyandang disabilitas rendah dengan anggota *cluster* berjumlah 27 kabupaten/kota yang meliputi Kabupaten Bogor, Kabupaten Sukabumi, Kabupaten Cianjur, Kabupaten Bandung, Kabupaten Garut Kabupaten Tasikmalaya, Kabupaten Ciamis, Kabupaten Kuningan, Kabupaten Cirebon, Kabupaten Majalengka, Kabupaten Sumedang, Kabupaten Indramayu, Kabupaten Subang, Kabupaten Purwakarta, Kabupaten Karawang, Kabupaten Bekasi, Kabupaten Bandung Barat, Kabupaten Pangandaran, Kota Bogor, Kota Sukabumi, Kota Bandung, Kota Cirebon, Kota Bekasi, Kota Depok, Kota Cimahi, Kota Tasikmalaya, dan Kota Banjar.

Dari hasil *cluster* 1 merupakan jumlah penyandang disabilitas yang tinggi dengan anggota cluster berjumlah 2 kabupaten/kota yang meliputi Kabupaten Ciamis dan Kota Bekasi. Dengan Performa cluster yang dihasilkan memiliki rata-rata centroid (*Avg. within centroid distance*) = 0.640, dengan rata-rata centroid cluster 0 adalah 0.616 dan rata-rata centroid *cluster* 1 adalah 2.616. Adapun saran bagi penelitian-penelitian selanjutnya yang akan disampaikan oleh penulis yaitu penulis berharap penelitian yang dilakukan ini bisa memberikan kontribusi yang lebih besar untuk penelitain-penelitian yang berhubungan dengan algoritma k-means kasus penyandang disabilitas selanjutnya adalah sebagai berikut : Pada penelitian selanjutnya disarankan untuk menggunakan datset dengan atribut yang lebih banyak berdasarkan topik yang digunakan agar hasil yang diperoleh lebih akurat.Pada penelitain selanjutnya disaranka n untuk menggunakan metode algoritma yang berbeda seperti algoritma naive bayes, algoritma k-medoids, dll. Pada penelitain selanjutnya dalam pengukuran jarak optimal sebuah *cluster* disarankan menggunakan metode lain seperti *bregmandivergences*.

DAFTAR PUSTAKA

[1] W. Putri and M. Afdal, "Penerapan Algoritma K-Means Untuk Pengelompokan Data Penyandang Disabilitas Di Kabupaten Rokan Hilir," *Indonesian Journal of Informatic Research and Software Engineering (IJIRSE)*, vol. 3, no. 1, pp. 30–38, 2023, doi: 10.57152/ijirse.v3i1.526.

[2] K. I. Wardoyo and J. Fredricka, "Klasterisasi Siswa Penyandang Tunagrahita Menggunakan Algoritma K-Means," vol. 19, no. 1, pp. 1–10, 2023.

[3] I. S. S. Ade Indah Sari, Heru Satria Tambunan, Widodo Saputra, Irfan Sudahri Damanik, "Implementasi Algoritma K-Means Dalam Pengelompokan Penyandang Disabilitas Menurut Kecamatan Kabupaten Simalungun," *Prosiding Seminar Nasional Riset Dan Information Science (SENARIS)*, vol. 2, no. 0, pp. 54–61, 2020, [Online]. Available: http://tunasbangsa.ac.id/seminar/index.php/senaris/article/view/144

[4] D. Triyansyah and D. Fitrianah, "Analisis Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Marketing," *Jurnal Telekomunikasi dan Komputer*, vol. 8, 2018, doi: 10.22441/incomtech.v8i2.4174.

[5] Anggreini Novita Lestari, "TEKNIK CLUSTERING DENGAN ALGORITMA K-MEDOIDS UNTUK MENANGANI STRATEGI PROMOSI DI POLITEKNIK TEDC BANDUNG," *Jurnal Teknologi Informasi dan Pendidikan*, vol. 12, no. 2, 2019, [Online]. Available: http://tip.ppj.unp.ac.id

[6] M. Atiqul Mutaqin and W. Andriyani, "Klasterisasi Data Disabilitas Menggunakan Algoritma K-Means," *Ijir*, vol. 3, no. 1, pp. 25–35, 2022.

[7] F. Harahap, "Perbandingan Algoritma K Means dan K Medoids Untuk Clustering Kelas Siswa Tunagrahita," *TIN: Terapan Informatika Nusantara*, vol. 2, no. 4, pp. 191–197, 2021.