

PENERAPAN DATA MINING UNTUK CLUSTERING PENYAKIT DIARE MENGGUNAKAN ALGORITMA K-MEANS (STUDI KASUS: PUSKESMAS BEBER)

Kholil AldiYatna¹, Nining Rahaningsih², Raditya Danar Dana³

¹ Teknik Informatika, STMIK IKMI Cirebon

² Komputerisasi Akuntansi, STMIK IKMI Cirebon

³ Manajemen Informatika, STMIK IKMI Cirebon

Jl. Perjuangan No.10B, Karyamulya Cirebon, Indonesia

kholilaldi02@gmail.com

ABSTRAK

Masih menjadi tantangan utama di Indonesia, termasuk di Kabupaten Cirebon, penyakit diare terus meningkatkan kekhawatiran kesehatan. Puskesmas Beber Cirebon mengalami lonjakan kasus diare yang mencolok dalam beberapa tahun terakhir. Selain jumlah penderita yang meningkat, penelitian ini juga menyoroti kurangnya pemahaman tentang pola, karakteristik, dan penyebaran diare di daerah Beber. Penelitian ini bertujuan untuk menganalisis dan mengelompokkan data diare di wilayah kerja Puskesmas Beber menggunakan teknik data mining dengan algoritma K-Means. Data dari rekam medis pasien Puskesmas Beber yang menderita diare dianalisis dan dikelompokkan ke dalam 3 cluster menggunakan algoritma K-Means. Hasilnya menunjukkan nilai K optimal adalah 3, dengan cluster 0 memiliki jumlah kasus tertinggi (386 kasus), diikuti oleh cluster 1 (132 kasus) dan cluster 2 (61 kasus). Temuan ini memberikan wawasan penting tentang pola dan faktor risiko diare di wilayah tersebut, berpotensi untuk meningkatkan diagnosis, penanganan, dan pencegahan diare. Penerapan data mining pada rekam medis juga dapat meningkatkan manajemen informasi di Puskesmas, efisiensi layanan kesehatan, dan mendukung kebijakan penanggulangan penyakit yang lebih terarah. Secara keseluruhan, penelitian ini memberikan kontribusi metodologis dan praktis yang berharga dalam upaya mengatasi masalah diare di tingkat lokal atau Puskesmas

Kata kunci : Data Mining, Algoritma Clustering, Penyakit Diare, K-Means

1. PENDAHULUAN

Diare adalah salah satu masalah kesehatan yang sering terjadi di seluruh dunia, termasuk Indonesia [1]. Penanganan dan prediksi penyakit diare menjadi hal yang sangat penting dalam upaya menjaga kesehatan Masyarakat. Puskesmas memiliki data pasien, seperti riwayat penyakit dan gejala, yang dapat digunakan untuk menganalisis dan memprediksi penyakit diare. Cirebon terletak di pesisir utara Jawa Barat, yang merupakan wilayah yang beragam dalam hal geografis dan demografis. Kepadatan penduduk yang tinggi dapat memperbesar risiko penyebaran penyakit infeksius, termasuk diare[2].

Tabel 1. Data 10 Jenis Penyakit Terbanyak di Kabupaten Cirebon Tahun 2020

No	Jenis Penyakit	Jumlah Kasus
1	Acute upper respiratory infection, unspecified	57,844
2	Essential (primary) hypertension	26,574
3	Myalgia	19,445
4	Dyspepsia	18,694
5	Acute nasopharyngitis [common cold]	15,985
6	Diarrhea and gastroenteritis of presumed infectious origin	8,703
7	Acute upper respiratory infections of multiple and unspecified sites	6,474
8	Non-insulin-dependent diabetes mellitus without complications	5,843
9	Necrosis of pulp	5,190
10	Acute pharyngitis, unspecified	3,323

Berdasarkan data kesehatan di Cirebon, diare menempati posisi ke-6 dari 10 penyakit terbanyak pada tahun 2020, hal ini menunjukkan fakta bahwa ada kebutuhan nyata untuk mengembangkan alat atau sistem yang dapat membantu dalam mengelola dan memitigasi penyakit ini.

Tabel 2. Jumlah Kasus Diare di Kab Cirebon Tahun 2021-2022

Jumlah Kasus	Tahun
8617	2021
8660	2022

Data menunjukkan bahwa jumlah kasus diare relatif stabil dalam periode tersebut, dengan 8.617 kasus pada tahun 2021 dan 8.660 kasus pada tahun 2022. Meskipun tidak terjadi perubahan yang signifikan dalam jumlah kasus diare dari tahun ke tahun, angka tersebut tetap cukup tinggi, mencapai lebih dari 8.600 kasus per tahun. Hal ini menunjukkan bahwa diare tetap menjadi masalah kesehatan yang signifikan di Cirebon dan perlu mendapatkan perhatian serius.

Puskesmas Beber di Cirebon adalah salah satu fasilitas kesehatan yang melayani masyarakat setempat. Peningkatan yang terjadi dalam jumlah kasus diare di Kecamatan Beber, Kabupaten Cirebon, seperti yang terlihat dari data yang diperoleh dari Puskesmas Beber, sangat mencolok. Pada tahun 2022 terdapat 579 kasus diare. Data ini memberikan gambaran jelas tentang meningkatnya prevalensi diare

di wilayah tersebut, terutama di kalangan balita. Oleh karena itu, perlu adanya dukungan kolaboratif dari semua pihak dalam mengatasi masalah diare ini di masyarakat. Dalam era informasi dan teknologi yang semakin maju, penggunaan data mining dan algoritma *clustering* menjadi alat yang sangat berguna dalam menganalisis dan mengklasterkan penyakit diare. [3]

Adapun pada penelitian terdahulu yang dikemukakan oleh [4] diare adalah suatu gangguan pada saluran pencernaan yakni tinja ataupun feses berubah menjadi cair serta berlangsung 3 kali ataupun lebih dengan kurun waktu 24 jam dan bisa dibarengi muntah- muntah. Penyakit ini terjadi sebab pola konsumsi makanan yang dicemari sejumlah bakteri, virus serta parasit yang muncul pada lingkungan yang minim kebersihannya. Serta, berdasarkan data WHO tahun 2019, diare menjadi penyebab menurunkan usia harapan hidup sebesar 1,97 tahun pada penderitanya, di bawah penyakit infeksi saluran pernapasan bawah (2,09 tahun). Secara global pada tahun 2016, air minum yang tidak sehat, sanitasi buruk, dan lingkungan kurang bersih menjadi faktor utama terhadap kematian 0,9 juta jiwa termasuk lebih dari 470.000 kematian bayi yang disebabkan oleh diare menurut [5]. Algoritma *K-Means* merupakan metode yang digunakan untuk meng*clustering* atau mengelompokkan suatu data penyebaran penyakit TBC merupakan penyakit menular yang masih menjadi perhatian dunia hingga saat ini [6].

Pada penelitian ini bertujuan mengelompokkan daerah- daerah penyebaran penyebarannya. Data mining adalah proses ekstraksi pengetahuan yang berharga dari kumpulan data yang besar dan kompleks [3].

Salah satu cabang utama dari data mining adalah *clustering*, di mana algoritma *clustering* digunakan untuk mengelompokkan data ke dalam kategori atau kelas yang sesuai berdasarkan fitur-fitur yang ada. Pendekatan data mining untuk mengidentifikasi pola penyakit diare berdasarkan data riwayat medis pasien. Metode yang digunakan meliputi klasifikasi dan *clustering*. Temuan utama mencakup kemampuan algoritma klasifikasi untuk memprediksi jenis diare berdasarkan gejala [7].

Salah satu teknik dalam data mining yang dapat digunakan adalah algoritma *clustering K-Means* [3]. Algoritma ini dipilih karena efektif dalam pengelompokan data ke dalam kelompok-kelompok yang homogen berdasarkan pola-pola tertentu, sifat yang relatif sederhana dan mudah diimplementasikan, serta bertujuan mengidentifikasi pola atau kelompok data yang mungkin mencerminkan karakteristik penyakit diare. Melalui penggunaan data mining dan algoritma *clustering*, akan dapat memberikan solusi yang efektif dalam mengidentifikasi kasus diare lebih awal, mengarahkan perawatan yang tepat, dan pada akhirnya, mengurangi dampak penyakit diare pada masyarakat Cirebon.

Penelitian ini bertujuan utama untuk mengembangkan pemahaman yang lebih mendalam tentang pola dan faktor-faktor yang berkaitan dengan kasus penyakit diare di Puskesmas Bebe Kabupaten Cirebon, menggunakan pendekatan data mining dengan fokus pada algoritma *k-means clustering*. Tujuan spesifik melibatkan identifikasi kelompok pasien dengan risiko tinggi terhadap penyakit diare, pengelompokan wilayah geografis berdasarkan prevalensi penyakit, serta pengembangan strategi pencegahan dan penanganan yang lebih spesifik.

Melalui penelitian ini, diharapkan dapat ditemukan pola-pola karakteristik yang dapat membantu dalam merancang langkah-langkah intervensi yang lebih efektif. Selain itu, penelitian ini berusaha untuk mengisi kesenjangan pengetahuan yang ada dengan menerapkan algoritma *k-means* pada data kesehatan masyarakat, khususnya terkait penyakit diare. Kontribusi pada bidang Informatika dapat dilihat dari pengaplikasian metode data mining dalam konteks kesehatan masyarakat, yang dapat memberikan pandangan baru dan holistik dalam penanganan penyakit. Potensi manfaat praktis dari hasil penelitian ini mencakup penguatan sistem pencegahan penyakit di puskesmas, pengalokasian sumber daya kesehatan yang lebih efisien, serta memberikan informasi yang lebih akurat kepada masyarakat terkait risiko dan langkah-langkah preventif yang dapat diambil. Dengan demikian, penelitian ini diharapkan tidak hanya memberikan kontribusi pada pengembangan bidang Informatika, tetapi juga memiliki dampak positif yang nyata dalam upaya meningkatkan kesehatan masyarakat di Kabupaten Cirebon.

Dalam mengatasi masalah meningkatnya kasus diare di Kabupaten Cirebon, penggunaan metode algoritma menjadi sangat relevan. Salah satu metode yang efektif dalam mengelompokkan data adalah menggunakan algoritma *k-means*. Algoritma *k-means* merupakan salah satu metode *clustering* yang bertujuan untuk mengelompokkan data ke dalam beberapa kelompok berdasarkan kemiripan fitur atau karakteristik tertentu [3].

K-means clustering merupakan salah satu metode data *clustering* non-hirarki yang mengelompokkan data dalam bentuk satu atau lebih *cluster*/kelompok. Data-data yang memiliki karakteristik yang sama dikelompokkan dalam satu *cluster*/kelompok dan data yang memiliki karakteristik yang berbeda dikelompokkan dengan *cluster*/kelompok yang lain sehingga data yang berada dalam satu *cluster*/kelompok memiliki tingkat variasi yang kecil. Dalam konteks analisis penyakit diare, penggunaan *k-means* dapat membantu identifikasi pola dan karakteristik tertentu yang mungkin menjadi faktor penyebab peningkatan kasus diare.

Melalui penelitian ini, hasil yang diperoleh dapat memberikan kontribusi signifikan pada pemahaman kita tentang penyakit diare di Puskesmas Bebe, Kabupaten Cirebon, dengan menggunakan pendekatan

data mining berbasis algoritma *k-means clustering*. Penelitian ini akan menghasilkan informasi yang lebih terinci mengenai pola dan faktor-faktor risiko penyakit diare, mengisi kekosongan pengetahuan dalam menerapkan teknologi informatika pada analisis kesehatan masyarakat. Para praktisi di bidang kesehatan dapat memanfaatkan temuan ini untuk mengidentifikasi kelompok risiko tinggi secara lebih efektif, menyusun strategi pencegahan yang lebih terarah, dan mengoptimalkan alokasi sumber daya. Selain itu, hasil penelitian ini dapat menjadi acuan bagi peneliti dalam pengembangan metode analisis penyakit lainnya yang melibatkan data kesehatan masyarakat. Secara keseluruhan, potensi implikasi hasil penelitian ini akan memberikan kontribusi pada perkembangan teknologi di bidang kesehatan, meningkatkan pemahaman tentang pola penyakit diare, dan memberikan dasar untuk pengembangan inovatif dalam penelitian kesehatan berbasis data di masa depan.

2. TINJAUAN PUSTAKA

2.1. Penelitian Terdahulu

Paper 1 “Analisis K-Means Clustering Pada Pemetaan Provinsi Indonesia Berdasarkan Indikator Rumah Layak Huni”[8]. Artikel penelitian ini berfokus pada penggunaan analisis kluster untuk mengelompokkan provinsi-provinsi di Indonesia berdasarkan indikator perumahan. Hasil analisis menghasilkan empat kluster yang masing-masing memiliki karakteristik yang berbeda terkait akses layak huni, air minum, penerangan listrik, dan sanitasi layak. Temuan-temuan tersebut disajikan dalam bentuk tabel, grafik, dan peta panas, yang memberikan wawasan tentang hasil pengelompokan dan distribusi berbagai variabel di berbagai wilayah.

Paper 2 “Pemetaan Kabupaten Kota Di Provinsi Jawa Timur Berdasarkan Tingkat Kasus Penyakit Menggunakan Pendekatan Agglomeratif Hierarchical Clustering”[9]. Penelitian ini membahas tentang analisis dan pemetaan kabupaten dan kota di Provinsi Jawa Timur berdasarkan karakteristik sebaran jumlah kasus penyakit pada setiap kabupaten dan kota dengan menggunakan metode hierarchical settlement clustering. Tujuan dari penelitian ini adalah untuk memandu Pemerintah Provinsi Jawa Timur dalam mengambil kebijakan perencanaan untuk meningkatkan pembangunan nasional di bidang kesehatan sehingga tingkat kesehatan masyarakat dapat diangkat ke level. Metode penelitian yang digunakan adalah studi literatur dengan menggunakan sumber referensi dan ulasan lain dari website pendukung.

Paper 3 “Data Mining Menggunakan Algoritma K-Means Clustering dalam Analisis Tingkat Potongan Harga terhadap Harga Jual Sepeda Motor Honda”[10]. Penelitian ini, membahas penggunaan teknik Data Mining dengan menggunakan Algoritma K-Means Clustering untuk menentukan tingkat potongan harga yang tepat dari setiap tipe sepeda motor Honda.

Penelitian ini juga membahas tahapan-tahapan yang dilakukan dalam proses penelitian, seperti mengidentifikasi masalah, mempelajari literatur, menentukan tujuan penelitian, menguji hasil penelitian, dan menganalisa hasil penelitian.

Paper 4 “Klasterisasi Menggunakan Metode Algoritma K-Means dalam Meningkatkan Penjualan Tupperware”[11]. Penelitian ini membahas tentang penggunaan metode Data Mining dengan menggunakan algoritma K-Means untuk klasterisasi dalam meningkatkan penjualan produk Tupperware pada toko Asrafi Raya di Pasaman Barat. Penelitian ini dilakukan dengan mengidentifikasi masalah, menganalisis masalah, menentukan tujuan, mempelajari literatur, mengumpulkan data, melakukan analisis dengan metode K-Means, mengimplementasikan dengan software RapidMiner, dan menguji hasil penelitian. Hasil penelitian ini dapat membantu pemilik toko Asrafi Raya dalam menentukan strategi penjualan pada Toko Asrafi Jaya.

Paper 5 “Implementation Of The K-Means Clustering Algorithm For The Covid-19 Vaccinated Village In The Ujung Padang Sub-District”[12]. Penelitian ini membahas tentang implementasi algoritma k-means clustering untuk mengelompokkan desa tervaksinasi Covid-19 pada Kecamatan Ujung Padang berdasarkan jumlah kasus vaksinasi Covid-19 dosis 1 dan dosis 2. Hasil dari penelitian ini berupa 3 cluster yang terdiri dari C1 tingkat desa tervaksinasi Covid-19 tertinggi terdapat 2 nagori, C2 tingkat desa tervaksinasi Covid-19 sedang terdapat 12 nagori dan C3 tingkat desa tervaksinasi Covid-19 rendah terdapat 6 nagori.

Paper 6 “Penerapan Metode Algoritma K-Means Clustering Untuk Pemetaan Penyebaran Penyakit Demam Berdarah Dengue (DBD)”[13]. Penelitian ini membahas penerapan algoritma K-Means clustering untuk pemetaan wilayah yang terkena Demam Berdarah Dengue . Selain itu, penelitian ini mengidentifikasi tiga cluster berdasarkan tingkat penyebaran DBD, dengan nilai kinerja sebesar 1.044, menunjukkan bahwa jumlah cluster 3 adalah jumlah yang optimal. Selain itu, penelitian ini juga menyajikan gejala penyakit DBD, proses data mining, dan aplikasi K-Means clustering dalam analisis kesehatan masyarakat.

Paper 7 “Klasifikasi Data Penduduk Pada Pemilihan Umum Di Kota Binjai Menggunakan Algoritma K-Means (Studi Kasus : KPU Kota Binjai)”[14]. Penelitian ini membahas tentang penerapan metode data mining dalam pemilihan umum di Kota Binjai. Penelitian ini dimulai dengan tahap identifikasi masalah untuk mengamati dan mencari permasalahan yang dihadapi oleh KPU Kota Binjai. Selanjutnya, dilakukan pengumpulan teori-teori terkait dengan data mining dan data penduduk. Tahap pengujian metode dilakukan untuk menguji metode yang digunakan dalam pengelompokan data yang tepat. Selanjutnya, dilakukan perancangan sistem untuk mempermudah pengguna dalam memahami

sistem yang dibangun. Jurnal ini juga mencantumkan beberapa referensi terkait dengan penggunaan metode data mining dalam berbagai konteks lainnya.

2.2. Data Mining

Data mining dalam penelitian ini adalah proses menggali informasi berharga dari sekumpulan data mentah peminjaman buku perpustakaan STMIK IKMI Cirebon dengan mengaplikasikan teknik pengelompokan data (clustering) khususnya algoritma k-means. Algoritma k-means digunakan untuk membagi dan mengelompokkan data peminjaman buku ke dalam beberapa kelompok/klaster berdasarkan kesamaan pola atau kemiripan karakteristik. Hasil dari penerapan data mining ini diharapkan dapat memberikan informasi penting terkait pola peminjaman dan minat baca pengunjung perpustakaan.[15]

2.3. Clustering

Clustering dalam penelitian ini adalah proses pengelompokan data pasien berdasarkan rekam medis di Puskesmas M. Thaha Bengkulu Selatan ke dalam beberapa kelompok atau klaster. Tujuan clustering adalah untuk membagi data pasien ke dalam kelompok-kelompok yang memiliki kesamaan karakteristik tertentu, seperti jenis penyakit yang diderita, gejala, atau faktor risiko.[16]

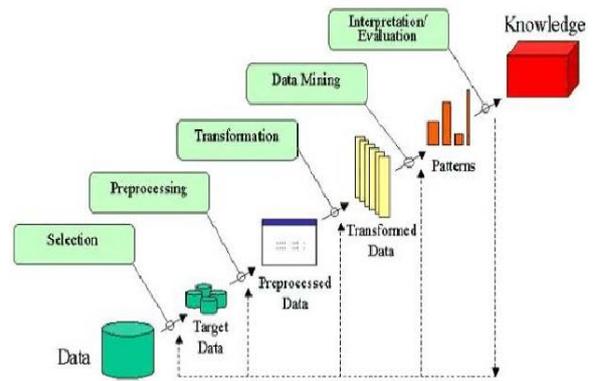
2.4. Algoritma K-Means

Algoritma K-Means adalah teknik analisis data yang digunakan dalam penelitian ini untuk mengelompokkan data pasien rehabilitasi narkoba ke dalam beberapa kelompok berdasarkan kesamaan pola atau karakteristik perilaku. Algoritma ini bekerja dengan memulai pembentukan cluster awal dan secara iteratif mengoreksi cluster tersebut hingga tidak ada perubahan yang signifikan. Dalam penelitian ini, algoritma K-Means digunakan untuk mengklasifikasikan pasien berdasarkan variabel seperti usia, jenis narkoba yang digunakan, dan lama penggunaannya. Penerapan algoritma ini diharapkan dapat memberikan wawasan penting untuk pengambilan keputusan dalam program rehabilitasi narkoba.[17]

3. METODE PENELITIAN

3.1. Tahapan Metode Penelitian

Gambar 1 Membahas tahapan metode Knowledge Discovery in Database (KDD). Penelitian ini menggunakan pendekatan kuantitatif, penelitian ini berfokus pada analisis data kuantitatif yang melibatkan pengolahan data pasien penyakit diare. Data penelitian ini bersumber pada data pasien di Puskesmas Beber, Cirebon mengenai demografis, geografis, Riwayat medis, gejala, dan data lainnya yang berkaitan dengan penyakit diare. Data tersebut akan dibersihkan dan dipreproses untuk memastikan kualitas data yang baik sebelum digunakan dalam analisis data mining.



Gambar 1. Tahapan Metode Penelitian

3.2. Sumber Data

Sumber data yang digunakan dalam penelitian ini berasal dari Puskesmas Beber Cirebon. Data yang digunakan dalam penelitian ini adalah data pasien yang mencakup berbagai informasi terkait dengan penyakit diare. Jenis data ini mencakup:

- Nama Penderita, adalah nama pasien yang pernah dilakukan treatment pengobatan karena gejala diare
- Zinc, adalah salah satu jenis obat, sebagai salah satu treatment dalam menanggulangi gejala penyakit diare.
- Jenis Kelamin, adalah gender dari pasien yang pernah dilakukan treatment pengobatan karena gejala diare
- Desa / Kelurahan, adalah alamat pasien penderita gejala diare yang dikelompokkan berdasarkan kelurahan
- Oralit, adalah salah satu jenis obat, sebagai salah satu treatment dalam menanggulangi gejala penyakit diare
- Tahun, adalah periode tahun dimana pasien yang pernah dilakukan treatment pengobatan karena gejala diare.

3.3. Teknik Pengumpulan Data

Pengumpulan data medis dilakukan dengan metode dokumentasi, cara yang digunakan untuk memperoleh data dan informasi dalam bentuk dokumen, tulisan, angka yang berupa laporan serta keterangan yang dapat mendukung penelitian. Metode dokumentasi dalam konteks penelitian ini merujuk pada penggunaan data sekunder yang sudah terdokumentasi sebelumnya, dalam hal ini data rekam medis pasien penyakit diare dari Puskesmas Beber. Penggunaan metode dokumentasi memungkinkan peneliti untuk mengakses dan menganalisis informasi yang telah tercatat dalam bentuk dokumen atau rekaman, tanpa perlu melakukan pengumpulan data baru secara langsung dari lapangan. Dalam hal ini, data pasien yang didiagnosis menderita penyakit diare telah terdokumentasi dengan rapi dalam rekam medis. Pengumpulan data dilakukan dengan melakukan pengambilan data historis langsung kepada pihak bersangkutan yang ada di Puskesmas Beber pada periode tahun 2022. Metode dokumentasi dipilih

karena data pasien penyakit diare yang digunakan berasal dari Puskesmas Beber, yang telah mencatat dengan cermat informasi medis pada rekam medis pasien selama periode Januari hingga Desember 2022. Keputusan ini diambil untuk memanfaatkan sumber data yang telah ada tanpa memerlukan waktu tambahan yang terkait dengan pengumpulan data primer. Selain itu, data yang terdokumentasi dengan baik dalam rekam medis Puskesmas Beber diyakini memiliki tingkat kualitas dan konsistensi yang tinggi, mengurangi risiko kesalahan atau ketidakakuratan dalam analisis. Format data yang telah terstruktur dengan baik memudahkan penerapan teknik analisis lanjutan, seperti data mining menggunakan algoritma *K-Means*, tanpa perlu melakukan langkah tambahan dalam persiapan data. Dengan demikian, penggunaan data sekunder ini memungkinkan peneliti untuk fokus pada analisis lebih lanjut untuk mengungkap pola dan wawasan baru terkait dengan penyakit diare yang diteliti.

3.4. Teknik Analisa Data

Pada penelitian kali ini menggunakan Teknik analisa data menggunakan Algoritma *K-Means* digunakan untuk mengelompokkan data secara partisi yang memisahkan data kedalam kelompok yang berbeda pada data penderita penyakit Diare. Aturan dalam penggunaan algoritma *K-Means* yaitu menentukan terlebih dahulu jumlah kluster, atribut yang digunakan bersifat numerik. Dalam setiap *cluster* memiliki karakter yang sama namun antar *cluster* diciptakan tingkat variasi yang kecil, sehingga memberikan karakteristik yang berbeda. Proses *clustering* mempergunakan algoritma *K-Means* secara umum mempunyai tahapan, yakni:

- a. Menentukan jumlah *cluster* (*k*).
- b. Tentukan *centroid* tiap *cluster* secara acak.
- c. Kemudian hitung jarak setiap objek ke *centroid* yang sudah ditentukan
- d. Mengelompokkan setiap data berdasarkan jarak terdekat antara data dengan masing-masing *centroid*.
- e. Perbarui kembali *centroid* setiap *cluster*
- f. Ulangi langkah ke 3 sampai 5 hingga setiap *cluster* tidak ada yang berubah.

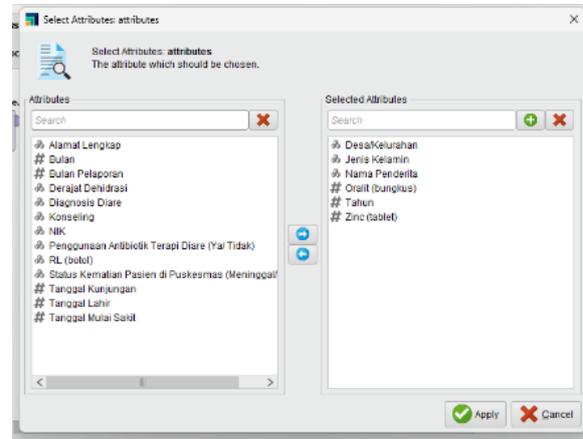
Algoritma *K-Means* dipilih untuk analisis data karena efektif dalam mengelompokkan data ke dalam *cluster* berdasarkan kedekatan fitur, serta relatif cepat dalam menangani volume data yang besar. Dalam konteks skripsi ini, algoritma *K-Means* memungkinkan untuk mengidentifikasi pola penyakit diare yang mungkin ada berdasarkan kesamaan gejala atau faktor risiko yang terkait.

Hasil interpretasi dari penggunaan algoritma *K-Means* dapat memberikan wawasan tentang kelompok-kelompok penyakit diare yang mungkin ada dalam dataset, serta faktor-faktor yang berkontribusi pada masing-masing kelompok tersebut. Generalisasi dari temuan ini dapat membantu dalam pengembangan strategi pencegahan dan penanganan penyakit diare

secara lebih efektif, dengan memperhatikan variasi dalam gejala dan faktor risiko yang mungkin terjadi pada populasi secara keseluruhan yang umum digunakan untuk mengukur kinerja model klusterisasi.

4. HASIL DAN PEMBAHASAN

4.1. Data Selection



Gambar 2. Select Attribute

Pada Gambar 2 adalah tahapan pemilihan attribute yang akan digunakan.



Gambar 3. Set Role

Pada Gambar 3 adalah tahapan pemilihan salah satu attribute untuk dikategorikan sebagai id, dalam tahap ini atribut Nama penderita dikategorikan sebagai id dan menggunakan bantuan atribut set role.

Name	Type	Missing	Statistics
Nama Penderita	Polynomial	114	Count: 2968 (1) AMANDA (4)
Jenis Kelamin	Polynomial	114	Count: 1 (284) P (285) P (285) L (284)
Desa/Kelurahan	Polynomial	114	Count: 1 (15) BEBER - 3209132011 (141) KONDANGBAR - 3209132
Oralit (bungkus)	Integer	114	Count: 6 (7) 6
Zinc (tablet)	Integer	312	Count: 8 (10) 8.807
Tahun	Integer	3	Count: 1 (123) 31.172

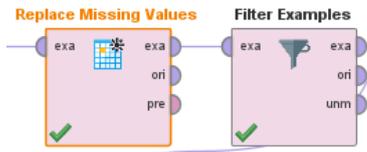
Gambar 4. Hasil Set Role

Pada Gambar 4 adalah hasil set role.

4.2. Preprocessing

Pada tahap preprocessing adalah tahapan pembersihan data dari hasil data pada Gambar 4. Preprocessing dilakukan karena terdapat beberapa atribut pada dataset memiliki nilai missing. Untuk mengatasi nilai missing digunakan operator replace missing value ,dan operator filter examples, dengan mengubah parameter attribute menjadi Zinc dan

default menjadi zero agar missing di dalam dataset sudah tidak memiliki nilai missing.



Gambar 5. Replace Missing Values

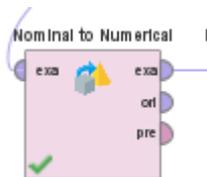
Gambar 5 adalah operator yang digunakan untuk melakukan pembersihan data atau Data Cleaning dengan menggunakan Replace Missing Value dan Filter Example.

nama	jenis kelamin	desa	zinc	oralt	tahun
Nama Penderita	jenis kelamin	Desa	Zinc (tablet)	Oralt (bungkus)	Tahun
Zero (zero)	jenis kelamin	Desa	Zinc (tablet)	Oralt (bungkus)	Tahun
Jenis Kelamin	jenis kelamin	Desa	Zinc (tablet)	Oralt (bungkus)	Tahun
Desa/Kelurahan	jenis kelamin	Desa	Zinc (tablet)	Oralt (bungkus)	Tahun
Oralt (bungkus)	jenis kelamin	Desa	Zinc (tablet)	Oralt (bungkus)	Tahun
Tahun	jenis kelamin	Desa	Zinc (tablet)	Oralt (bungkus)	Tahun

Gambar 6. Hasil Data Cleaning

Gambar 6 adalah hasil pembersihan data dan sudah tidak ada missing value.

4.3. Transformation



Gambar 7. Nominal to Numerical

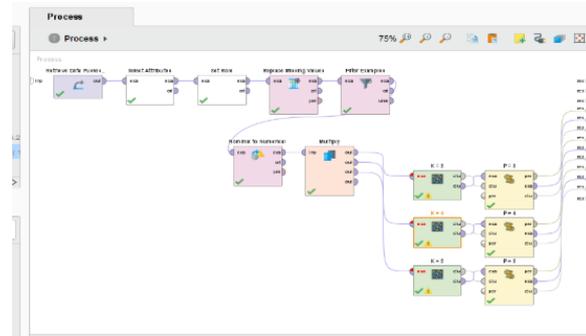
Row No.	Nama Pendi...	Jenis Kelamin	Desa/Kelura...	Zinc (tablet)	Oralt (bung...	Tahun
1	NURHAYATI	0	0	0	6	33
2	RIYANDI	1	1	0	6	31
3	SAFANJUR	0	2	0	6	27
4	RIKA	0	3	0	6	29
5	SUHENI	0	1	0	6	31
6	M LABIB	1	1	10	6	7
7	AHMAD SAMA	1	3	10	6	10
8	ARIN DWI	0	1	10	6	12
9	WAHYU	1	1	10	6	25
10	MUTIA PUTRI	0	3	10	6	14
11	TINI	0	4	0	6	48
12	DURAHIMAH	1	0	0	6	41
13	HERWANTO	1	0	10	6	10
14	MERI YUNISA	0	1	0	6	24
15	PAGIL PANG...	1	2	0	6	21

Gambar 8. Hasil Transformasi

Pada Gambar 7 adalah tahapan transformasi data Nominal to Numerical, digunakan untuk mengubah tipe atribut yang bersifat non-numerik menjadi tipe numerik. Dalam dataset penyakit Diare atribut jenis kelamin, desa/kelurahan, dirubah menjadi numerik.

Gambar 8 adalah hasil setelah transformasi data Nominal to Numerical.

4.4. Data Mining



Gambar 9. Model Penelitian

Pada Gambar 9. adalah tahapan yang dilakukan untuk pengelompokkan data diare dengan algoritma K-Means dengan jumlah cluster yang dipilih sejumlah 3, pemilihan jumlah kluster ini bertujuan untuk menemukan kluster dengan klasifikasi warga dengan tingkat penderita diare tinggi, sedang dan rendah.

Cluster Model

Cluster 0: 386 items
 Cluster 1: 61 items
 Cluster 2: 132 items
 Total number of items: 579

Gambar 10. Hasil Cluster

Pada Gambar 10 adalah hasil Cluster yang dihasilkan.

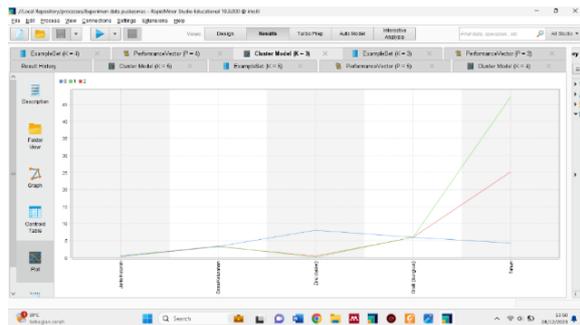
4.5. Interpretation/Evaluation

Performance
PerformanceVector
Avg. within centroid distance: 0.361
Avg. within centroid distance_cluster_0: 10.147
Avg. within centroid distance_cluster_1: 6.309
Avg. within centroid distance_cluster_2: 16.979
Device Soudan: 0.116

Gambar 11. Data Hasil PerformanceVector

Pada Gambar 11. adalah nilai DBI dengan metode K-Means dengan jumlah k=3 menghasilkan nilai 0,116, seperti pada gambar dibawah ini, dan juga dapat ditarik kesimpulan bahwa nilai DBI yang paling mendekati 0 merupakan hasil cluster terbaik.

Gambar 12 adalah gambar plot yang dapat simpulkan bahwa garis grafik yang berwarna hijau memiliki data anggota yang paling banyak



Gambar 12. Plot

4.6. Hasil

Pada hasil penelitian ini juga menunjukkan bahwa algoritma *K-Means* mampu membagi data pasien diare ke dalam beberapa kelompok berdasarkan karakteristik yang serupa. Menghasilkan 3 *cluster*, yang menunjukkan bahwa *cluster 0* memiliki insiden atau jumlah kasus penyakit diare tertinggi yaitu sebanyak 386 kasus, diikuti oleh *cluster 1* sebanyak 132 kasus dan *cluster 2* sebanyak 61 kasus. Dengan demikian penelitian ini mampu memberikan kontribusi signifikan pada pemahaman tentang penyakit diare di Puskesmas Beber, Kabupaten Cirebon, dengan menggunakan pendekatan data mining berbasis algoritma *k-means clustering*. Penelitian ini menghasilkan informasi yang lebih terinci mengenai pola dan faktor-faktor risiko penyakit diare, mengisi kekosongan pengetahuan dalam menerapkan teknologi informatika pada analisis kesehatan masyarakat. Para praktisi di bidang kesehatan dapat memanfaatkan temuan ini untuk mengidentifikasi kelompok risiko tinggi secara lebih efektif, menyusun strategi pencegahan yang lebih terarah, dan mengoptimalkan alokasi sumber daya.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian yang telah dirangkum sebelumnya, dapat disimpulkan bahwa pemodelan klaster menggunakan *K-Means* dinilai berhasil dengan baik berdasarkan nilai DBI sebesar 0,116 yang mendekati 0. Berdasarkan eksperimen dengan $K=3$ sampai 5, nilai DBI terkecil diperoleh pada $K=3$ yaitu 0,116 sehingga jumlah cluster optimal ada pada $K=3$. Perlu juga dilakukan implementasi hasil penelitian di Puskesmas Beber Cirebon dengan melibatkan pihak terkait untuk memastikan dampak nyata bagi upaya pencegahan diare. Penelitian lebih lanjut dapat mengintegrasikan data mining dengan SIG untuk visualisasi pola spasial penyakit diare di Cirebon.

DAFTAR PUSTAKA

- [1] A. V. Sutanto and Y. Fitriana, *Asuhan pada Kehamilan*. 2015.
- [2] S. Herman *et al.*, “No 主観的健康感を中心とした在宅高齢者における健康関連指標に関する分散構造分析Title,” *Jur. Tek. Kim. USU*, vol. 3,

no. 1, pp. 18–23, 2019.

- [3] S. Hasyrif, Rismayani, and S. Asrul, “PROSIDING SEMINAR ILMIAH SISTEM INFORMASI DAN TEKNOLOGI INFORMASI Pusat Penelitian dan Pengabdian Pada Masyarakat (P4M) STMIK Dipanegara Makassar Data Mining Menggunakan Algoritma *K-Means* Pengelompokan Penyebaran Diare Di Kota Makassar,” *SISITI Semin. Ilm. Sist. Inf. dan Teknol. Inf.*, vol. VIII, no. 1, pp. 73–82, 2019.
- [4] Y. R. Nasution and M. Eka, “Penerapan Algoritma *K-Means* Clustering Pada Aplikasi,” *J. Ilmu Komput. dan Inform.*, vol. 02, no. 01, pp. 77–88, 2018.
- [5] A. A. Iryanto, T. Joko, and M. Raharjo, “Literature Review: Faktor Risiko Kejadian Diare Pada Balita Di Indonesia,” *J. Kesehat. Lingkung.*, vol. 11, no. 1, pp. 1–7, 2021, doi: 10.47718/jkl.v11i1.1337.
- [6] T. B. Santos, “Aplikasi Data Mining untuk Clustering Daerah Penyebaran Penyakit Diare di DKI Jakarta Menggunakan Algoritma *K-MEANS*,” *J. Ilm. FIFO*, vol. 11, no. 2, p. 131, 2019, doi: 10.22441/fifo.2019.v11i2.003.
- [7] E. Nendes, “Analisa Dan Desain Sistem Informasi Penjualan Furniture Berbasis E-Commerce Studi Kasus: Pt. Anya Living,” ... *Indones. J. Inf. Syst.*, 2019, [Online]. Available: <https://jom.fti.budiluhur.ac.id/index.php/IDEAL-IS/article/view/1070>
- [8] A. Septianingsih, “Analisis *K-Means* Clustering Pada Pemetaan Provinsi Indonesia Berdasarkan Indikator Rumah Layak Huni,” *J. Lebesgue J. Ilm. Pendidik. Mat. Mat. dan Stat.*, vol. 3, no. 1, pp. 224–241, 2022, doi: 10.46306/lb.v3i1.116.
- [9] A. Septianingsih, “Pemetaan Kabupaten Kota Di Provinsi Jawa Timur Berdasarkan Tingkat Kasus Penyakit Menggunakan Pendekatan Agglomeratif Hierarchical Clustering,” *J. Lebesgue J. Ilm. Pendidik. Mat. Mat. dan Stat.*, vol. 3, no. 2, pp. 367–386, 2022, doi: 10.46306/lb.v3i2.139.
- [10] R. Mauliadi, “Data Mining Menggunakan Algoritma *K-Means* Clustering dalam Analisis Tingkat Potongan Harga Terhadap Harga Jual Sepeda Motor Honda,” *J. Inform. Ekon. Bisnis*, vol. 4, pp. 7–9, 2022, doi: 10.37034/infv4i4.156.
- [11] I. P. Mulyadi, “Klasterisasi Menggunakan Metode Algoritma *K-Means* dalam Meningkatkan Penjualan Tupperware,” *J. Inform. Ekon. Bisnis*, vol. 4, pp. 172–179, 2022, doi: 10.37034/infv4i4.164.
- [12] D. Sinta Saputri, G. Maha Putra, M. Fitri Larasati, P. Studi Sitem Informasi, and S. Tinggi Manajemen Informatika dan Komputer Royal Kisaran, “Implementation of the *K-Means* Clustering Algorithm for the Covid-19 Vaccinated Village in the Ujung Padang Sub-District,” *J. Tek. Inform.*, vol. 3, no. 2, pp. 261–

- 267, 2022, [Online]. Available: <https://doi.org/10.20884/1.jutif.2022.3.2.165>
- [13] M. A. Sembiring, "Penerapan Metode Algoritma K-Means Clustering Untuk Pemetaan Penyebaran Penyakit Demam Berdarah Dengue (Dbd)," *J. Sci. Soc. Res.*, vol. 4, no. 3, p. 336, 2021, doi: 10.54314/jssr.v4i3.712.
- [14] W. I. S. Sinaga, R. Buaton, and H. Sembiring, "Klasifikasi Data Penduduk Pada Pemilihan Umum Di Kota Binjai Menggunakan Algoritma K-Means (Studi Kasus: KPU Kota Binjai)," *Explorer (Hayward)*, vol. 3, no. 2, pp. 95–102, 2023, [Online]. Available: <https://journal.fkpt.org/index.php/Explorer/article/view/794%0Ahttps://journal.fkpt.org/index.php/Explorer/article/download/794/429>
- [15] H. L. Sari and I. Y. Beti, "Penerapan Data Mining Dalam Pengelompokan Buku Yang Dipinjam Menggunakan Algoritma K-Means," *KLIK Kaji. Ilm. Inform. dan Komput.*, vol. 3, no. 6, pp. 925–933, 2023, doi: 10.30865/klik.v3i6.826.
- [16] P. Algoritma, K. D. A. N. K. Dalam, D. I. Puskesmas, and M. T. Bengkulu, "Pengelompokan Data Pasien Berdasarkan Rekam Medis," vol. 4307, no. 3, pp. 580–586, 2023.
- [17] E. Yolanda, "Penerapan Algoritma K-Means Clustering Untuk Pengelompokan Data Pasien Rehabilitasi Narkoba," *KLIK Kaji. Ilm. Inform. dan Komput.*, vol. 4, no. 1, pp. 182–191, 2023, doi: 10.30865/klik.v4i1.1107.