

Penentuan Jumlah Cluster Optimal Menggunakan Davies Bouldin Index pada Algoritma K-Means untuk Menentukan Kelompok Penyakit

Ibnu Fiki Fauzi¹, Mochzen Gito Resmi², Teguh Iman Hermanto³

^{1,2,3} Program Studi Teknik Informatika, Sekolah Tinggi Teknologi Wastukencana
Purwakarta
ibnufiki38@wastukencana.ac.id

ABSTRAK

Puskesmas Jatiluhur memberikan pelayanan kesehatan bagi masyarakat setiap harinya. sekitar 60 orang lebih dari berbagai wilayah yang terdapat di lingkungan kecamatan Jatiluhur. Seiring dengan bertambahnya jumlah pasien tersebut, maka bertambah pula data pasien setiap harinya, sehingga dari data tersebut hanya digunakan sebagai arsip untuk melihat riwayat penyakit pasien. Salah satu informasi data yang perlu diperhatikan yaitu mengenai kelompok penyakit pasien di lingkungan Puskesmas Jatiluhur. Untuk mempermudah proses pengelolaan data yang banyak diperlukan teknik menganalisis data, yaitu menggunakan teknik data mining. Dalam penelitian ini dilakukan teknik data mining dengan menggunakan algoritma k-means. Algoritma k-means merupakan salah satu algoritma data mining yang banyak digunakan dalam penelitian pengelompokan. Penelitian kali ini akan mengaplikasikan Davies Bouldin Index (DBI) sebagai salah satu cara menentukan jumlah cluster paling optimal untuk menentukan kelompok penyakit pasien di lingkungan Puskesmas kecamatan Jatiluhur. Data yang digunakan pada penelitian ini adalah data rekam medis pasien berobat. Pengujian dilakukan dengan membentuk beberapa cluster diantaranya K2, K3 dan K4 untuk mencari evaluasi pengujian dengan nilai DBI terkecil, dimana setelah dilakukan proses pengolahan data diketahui nilai terkecil adalah 0,406 yang terdapat pada cluster yang terbentuk sebanyak K3 dengan hasil clustering diperoleh cluster_0 sebanyak 28 anggota cluster_1 sebanyak 7 anggota dan cluster_2 sebanyak 102 anggota. Hasil pengelompokan penyakit tersebut dapat dijadikan dasar penyuluhan kesehatan berdasarkan prioritas kelompok penyakit yang banyak diderita.

Kata Kunci: *Data Mining, Davies Bouldin Index, K-Means, Puskesmas, Rekam Medis*

ABSTRACT

Puskesmas Jatiluhur provides health services to the community every day. more than 60 people from various areas in Jatiluhur sub-district. Along with the increasing number of patients, patient data also increases every day, so that the data is only used as an archive to view patient medical history. One of the information data that needs attention is regarding the disease group of patients in

the Puskesmas Jatiluhur environment. To simplify the data management process, many analyze data techniques are needed, namely by using data mining techniques. In this study, data mining techniques were carried out using the k-means algorithm. The k-means algorithm is a data mining algorithm that is widely used in clustering research. This study will apply the Davies Bouldin Index (DBI) as a way to determine the most optimal number of clusters to determine the disease group of patients at the Puskesmas District Jatiluhur. The data used in this study were medical record data of patients undergoing treatment. The test was carried out by forming several clusters including K2, K3 and K4 to look for test evaluations with the smallest DBI values, where after processing the data the smallest value was obtained, namely 0.406 which was found in the clusters formed as many as K3 with clustering results obtained as many as 28 members cluster_0, cluster_1 7 members and cluster_2 as many as 102 members. The results of these grouping diseases can be used as a basis for health counseling based on the priority groups of diseases from which many suffer.

Keywords: *Data Mining, Davies Bouldin Index, K-Means, Medical Record, Puskesmas*

1. PENDAHULUAN

Indonesia memiliki salah satu instansi kesehatan yaitu Puskesmas yang merupakan instansi pemerintah dengan fungsi memberikan pelayanan kepada orang yang membutuhkan pengobatan. Menurut Permenkes RI Nomor 75 Tahun 2014 Puskesmas adalah fasilitas pelayanan kesehatan yang menyelenggarakan upaya kesehatan masyarakat dan upaya kesehatan perseorangan tingkat pertama, dengan lebih mengutamakan upaya promotif dan preventif, untuk mencapai derajat kesehatan masyarakat yang setinggi-tingginya di wilayah kerjanya. Puskesmas merupakan salah satu instansi kesehatan yang berada ditingkat kecamatan, Puskesmas dapat menjangkau masyarakat di setiap daerah.

Kabupaten Purwakarta terletak di wilayah provinsi Jawa Barat, dengan 17 kecamatan. Setiap kecamatan Purwakarta memiliki puskesmas, dan beberapa kelurahan memiliki puskesmas sendiri. Berdasarkan dari data dasar Puskesmas provinsi Jawa Barat kementerian kesehatan Republik Indonesia tahun 2020 yang berhasil dihimpun, setidaknya ada 20 Puskesmas yang terdapat di kabupaten Purwakarta, termasuk Puskesmas Jatiluhur yang berada di kecamatan Jatiluhur. Tata kerja yang tertib, rapi, dan teliti diperlukan untuk menghasilkan informasi yang cepat, akurat, dan tepat waktu sesuai kebutuhan untuk mewujudkan pelayanan kesehatan yang baik. Seperti instansi kesehatan pada umumnya.

Puskesmas Jatiluhur memberikan pelayanan kesehatan bagi masyarakat setiap harinya, sekitar 60 orang lebih dari berbagai wilayah yang terdapat di lingkungan kecamatan Jatiluhur. Seiring dengan bertambahnya jumlah pasien tersebut, maka bertambah pula data pasien pada setiap harinya, sehingga dari data tersebut hanya digunakan sebagai arsip saja untuk melihat riwayat penyakit pasien. Salah satu informasi data yang perlu diperhatikan yaitu mengenai kelompok penyakit pasien di lingkungan Puskesmas Jatiluhur. Sehingga nantinya dapat membantu pihak Puskesmas Jatiluhur maupun pihak terkait untuk dijadikan informasi baru yang berguna sebagai bahan atau dasar penyuluhan kesehatan kepada masyarakat. Untuk mempermudah proses pengelolaan data yang banyak, suatu instansi kesehatan seperti halnya Puskesmas Jatiluhur diperlukan teknik menganalisis data untuk mengetahui kelompok penyakit pasien di

lingkungan Puskesmas Jatiluhur. Salah satunya yaitu menggunakan teknik data *mining*, sehingga data yang banyak tersebut dapat dimanfaatkan secara optimal.

Data *mining* adalah teknik pengolahan data berskala besar yang dapat menghasilkan pengetahuan baru (Setiyani et al., 2020). Data *mining* bertujuan dan berguna untuk menemukan pola data dengan jumlah data yang besar (Azhari et al., 2021). Salah satu metode yang dapat digunakan dalam data *mining* adalah metode *clustering*, yang merupakan teknik klasifikasi tanpa pengawasan yang mengklasifikasikan *dataset* berdasarkan kriteria kesamaan tertentu (Rustam et al., 2018). Algoritma *k-means* adalah algoritma *clustering* klasik berdasarkan partisi yang mudah dipahami, sederhana, dan konvergensi cepat.

Untuk melakukan penentuan tingkat kualitas hasil *clustering*, dilakukan evaluasi *clustering*. penentuan jumlah *cluster* terbaik pada penelitian ini menggunakan *Davies Bouldin Index* sebagai evaluasi hasil *clustering*. *Davies Bouldin Index* (DBI), yang diperkenalkan pertama kali oleh David L. Davies dengan Donald W. Bouldin pada tahun 1979, adalah salah satu cara untuk mengukur validitas atau jumlah *cluster* paling optimal dalam metode pengelompokan, di mana kohesi adalah jumlah dari kedekatan data terhadap titik pusat *cluster* dari *cluster* yang diikuti (Muningsih et al., 2021). Dengan menggunakan evaluasi DBI dilakukan skema evaluasi internal *cluster* yang dimana kualitas hasil *cluster* berdasarkan dari banyaknya dan kedekatan antara data hasil *cluster*.

Penelitian terdahulu yang dilakukan oleh Rahayu Anggraini, Elin Haerani, Jasril, Iis Afrianty dalam penelitiannya yang berjudul "Pengelompokan Penyakit Pasien Menggunakan Algoritma *K-means*". Analisis ini mengelompokan penyakit pasien, Data yang digunakan terdiri dari 3875 *record* dan 5 atribut yaitu Jenis Kelamin, Jenis Peserta, Diagnosa, Status Pulang, Alamat. Dari pengujian menggunakan algoritma *k-means* menunjukkan hasil *clustering* yaitu *cluster* 1 sebanyak 710 data sedangkan *cluster* 2 dengan jumlah 3165 data. Hasil dari penelitian menunjukkan bahwa penggunaan 2 *cluster* menjadi *cluster* terbaik (Anggraini et al., 2022).

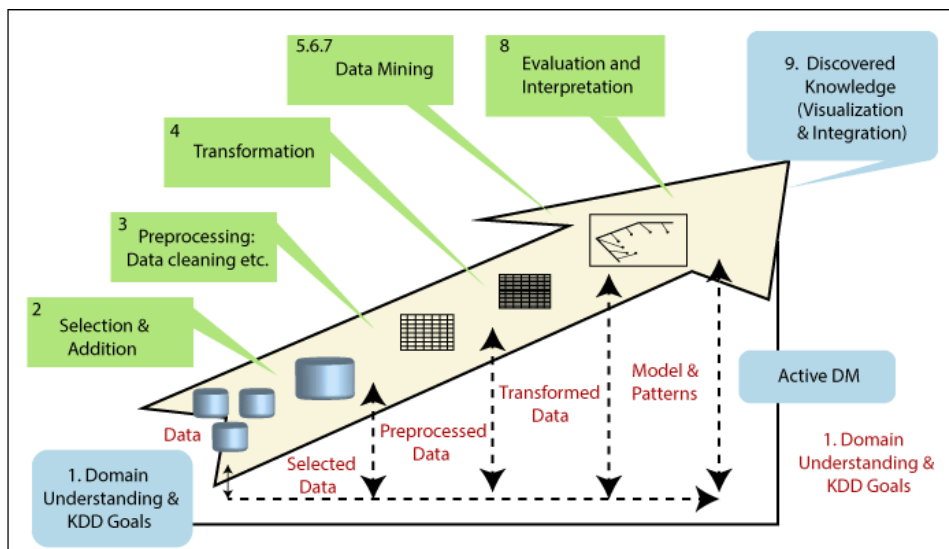
Penelitian ini bertujuan untuk menentukan *cluster* optimal dengan menggunakan *Davies Bouldin Index* pada algoritma *k-means*, serta pembentukan kelompok penyakit pasien dari hasil penentuan *cluster* optimal yang tersebar di lingkungan Puskesmas Jatiluhur dengan algoritma *k-means*, sehingga dapat dihasilkan informasi sebagai pengetahuan baru, serta dijadikan sebagai dasar penyuluhan kesehatan berdasarkan prioritas kelompok penyakit yang banyak diderita.

2. TINJAUAN PUSTAKA

Pada penelitian ini, tinjauan pustaka mencakup pemahaman serta pengetahuan mengenai proses metode data *mining* yang digunakan, yaitu *Knowledge Discovery in Database* (KDD), Algoritma *K-Means* dan *Davies Bouldin Index* (DBI).

2.1 *Knowledge Discovery in Database* (KDD)

Suatu proses komputasi yang menggunakan algoritma matematika untuk membuat ekstraksi data dan memperhitungkan tindakan yang mungkin dilakukan di masa depan biasa disebut *Knowledge Discovery in Database* (KDD). Hasil dari KDD adalah pengetahuan yang sebelumnya tidak diketahui, potensial, dan bermanfaat (Muttaqin & Defriani, 2020). Pada Gambar 1 terdapat tahapan proses yang ada pada KDD.



Gambar 1. Proses *Knowledge Discovery in Database* (KDD)
(Sumber: (Muttaqin & Defriani, 2020))

2.2 Algoritma *K-Means*

Algoritma *k-means* adalah salah satu algoritma *clustering* yang paling sederhana yang didasarkan pada nilai titik *centroid*, yang merupakan titik pusat *cluster* (Rohmatullah et al., 2020). Algoritma *k-means* menggabungkan setiap objek yang diamati ke dalam kelompok dengan *mean* terdekat. Tujuan algoritma adalah untuk menemukan pusat kelompok dalam data sebanyak iterasi perbaikan yang dilakukannya. Jumlah nilai *K* akan ditentukan untuk membentuk jumlah *cluster*, kemudian pencarian nilai titik *centroid*, pengelompokan data berdasarkan jumlah *cluster* yang telah ditentukan. Jika titik *centroid* berubah maka akan dilakukan lagi pemilihan titik *centroid* hingga diketahui tidak ada perubahan pada titik *centroid* (Nurahman et al., 2022). Berikut merupakan gambaran langkah-langkah pada algoritma *k-means* (Muliono & Sembiring, 2019).

1. Menentukan jumlah *k cluster* yang akan dibentuk. Penentuan jumlah *k cluster* biasanya didasarkan pada beberapa faktor teoritis dan konseptual. Faktor-faktor ini diusulkan untuk menentukan jumlah *k cluster*.
2. Menentukan *centroid* awal (titik awal pusat *cluster*). Penentuan *centroid* awal dilakukan secara acak, yang diambil dari beberapa objek yang tersedia sebanyak jumlah dari *k cluster*. Kemudian untuk menentukan *centroid* baru pada iterasi berikutnya, dengan menggunakan rumus persamaan (1) sebagai berikut (Sari & Chotijah, 2022):

$$C = \sum_{i=1}^r (x_{i1} + x_{i2} + \dots + x_{in}) / j^n \quad (1)$$

Keterangan:

C = *centroid* nilai rata-rata dari *cluster*

x = nilai data ke- dalam *cluster*

j = jumlah anggota data pada *cluster*

3. Menghitung jarak dari setiap objek terhadap nilai *centroid* dari masing-masing data *cluster*, dengan rumus *Euclidian Distance* pada persamaan (2) sebagai berikut (Srimurdianti et al., 2023):

$$D_E = \sqrt{(x_{1i} - x_{1j})^2 + \dots + (x_{ki} - x_{kj})^2} \quad (2)$$

Keterangan:

D_E = jarak data ke i terhadap pusat *cluster* j

x_{ki} = data ke i pada atribut data ke k

x_{kj} = nilai titik pusat ke j pada atribut ke k

4. Menghitung masing-masing objek ke dalam titik pusat *cluster* (*centroid*) yang paling terdekat.
5. Melakukan iterasi ke- i , kemudian tentukan posisi *centroid* baru dengan menggunakan persamaan sebelumnya.
6. Mengulangi langkah pada poin 3 jika posisi *centroid* baru tidak sama dan iterasi atau perulangan akan berhenti jika rasio tidak lebih besar dari nilai rasio sebelumnya hingga hasil perngitungan di masing-masing data konvergen.

2.3 Davies Bouldin Index (DBI)

Davies Bouldin Index (DBI) mempunyai fungsi untuk memaksimalkan jarak *inter-cluster*, dan juga mencoba meminimalkan jarak antar-titik pada *cluster* pada saat yang bersamaan. Nilai rata-rata yang similar antara setiap *cluster* dan paling mirip merupakan ukuran dari DBI. Semakin rendah nilai DBI atau bernilai positif yang mendekati nol, maka menunjukkan skema *cluster* terbaik (Mayadi, Siti Setiawati, 2022). Berikut merupakan gambaran langkah-langkah untuk menghitung nilai *Davies Bouldin Index* (Septiani et al., 2022)(Irwan et al., 2023).

1. Menghitung *Sum of Square Within Cluster* (SSW) merupakan keterikatan antara anggota satu *cluster* atau seberapa mirip antara anggota satu dan lainnya, serta semakin kecil semakin baik dikarenakan semakin mirip. SSW dihitung untuk mengetahui matrik/kohesi/homogenitas. Kohesi adalah hubungan antara anggota kelompok dalam satu kelompok, dengan rumus persamaan (3) berikut (Amrulloh, 2023).

$$SSW = \frac{1}{m_i} \sum_{j=1}^{m_i} d(x_j, c_i) \quad (3)$$

Keterangan:

m_i = jumlah data pada *cluster* ke- i

x = data dalam *cluster*

$d(x,c)$ = jarak data ke *centroid*

x_i = data pada *cluster* tersebut

c_i = pusat *centroid cluster* ke- i

2. Menghitung *Sum of Square Between Cluster (SSB)*, yang merupakan jarak yang cukup besar antar kelompok sehingga terpisah dari kelompok lain, bertujuan untuk mengetahui separasi atau heterogenitas. Separasi adalah perbedaan antara kelompok satu dengan yang lainnya, dengan rumus persamaan (4) berikut (Sisfotenika, 2022).

$$SSB_{ij} = d(c_i, c_j) \quad (4)$$

Keterangan:

c_i = *cluster* 1

c_j = *cluster* lainnya

$d(c_i, c_j)$ = jarak antara *centroid* satu

3. Menghitung Rasio berfungsi untuk dapat mengetahui seberapa bagus nilai perbandingan *cluster* satu dengan *cluster* lainnya. Jumlah separasi harus lebih besar daripada kohesi, dengan rumus persamaan (5) berikut (Wahyudi & Pujiastuti, 2022).

$$R_{ij} = \frac{SSW_i + SSW_j}{SSB_{ij}} \quad (5)$$

Keterangan:

R_{ij} = rasio antar *cluster*

SSW_i = *cluster* 1

SSW_j = *cluster* 2

SSB_{ij} = separasi dari *cluster* 1 dan 2

4. Menghitung DBI (*Davies Bouldin Index*), dengan rumus persamaan (6) berikut (Prasetyo et al., 2023).

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{i \neq j} (R_{ij}) \quad (6)$$

Keterangan:

k = *cluster* yang ada

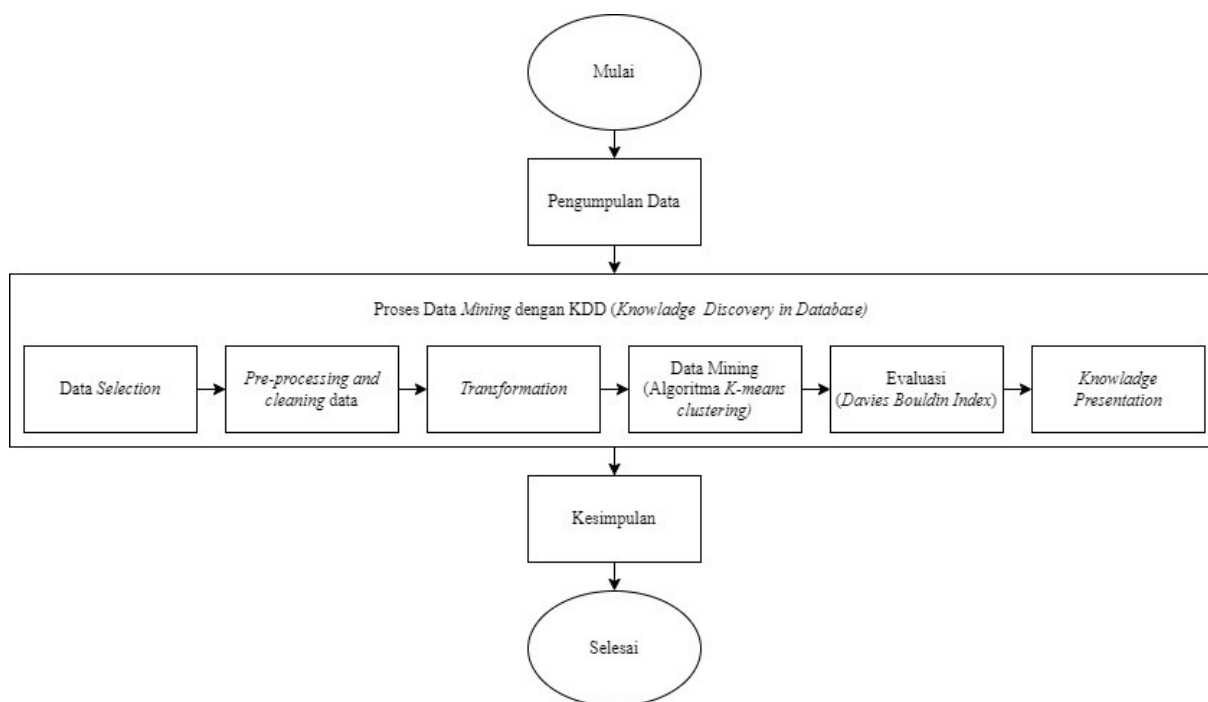
$R_{i,j}$ = rasio antar *cluster* i dan j

Max = dicari rasio antara *cluster* yang terbesar

3. METODOLOGI

Pada penelitian ini terdapat 3 tahapan, tahapan pertama melakukan pengumpulan data, tahapan kedua melakukan proses data *mining* dengan menggunakan *Knowledge Discovery in Database (KDD)* dan tahapan yang terakhir yaitu kesimpulan. Berikut merupakan tahapan metode penelitian dapat dilihat pada Gambar 2.

Penentuan Jumlah Cluster Optimal Menggunakan Davies Bouldin Index pada Algoritma K-Means untuk Menentukan Kelompok Penyakit



Gambar 2. Tahapan Penelitian

3.1 Pengumpulan Data

Sumber data yang digunakan pada penelitian ini merupakan data rekam medis pasien berobat di Puskesmas kecamatan Jatiluhur, dari rentang waktu bulan juni 2019 sampai dengan bulan agustus 2019, dengan jumlah data sebanyak 2496 *record*. Berikut merupakan contoh *dataset* rekam medis pasien berobat, dapat dilihat pada Tabel 1.

Tabel 1. Dataset Awal Rekam Medis Pasien Berobat

Pasien	Alamat	Tahun lahir	Jenis kelamin	Status pasien	Poli	Status bayar	Diagnosa
1	Cikao Bandung	2016	L	Lama	MTBS	Umum	<i>Gastroemteritis</i>
2	Cikao Bandung	1973	L	Lama	Lansia	BPJS	ISPA
3	Cikao Bandung	1963	P	Lama	Lansia	BPJS	<i>Myalgia</i>
:	:	:	:	:	:	:	:
2496	Mekargalih	1957	P	Lama	Lansia	BPJS	<i>Arthralgia</i>

3.2 Proses Data Mining

Proses data *mining* menggunakan *Knowledge Discovery in Database* dengan tahapan yang dilakukan sebagai berikut:

1. *Data Selection*, Pada tahapan ini dilakukan penyeleksian atribut terhadap data awal pasien berobat. Penyeleksian data ini bertujuan untuk memilih atribut mana saja yang akan digunakan ataupun yang tidak akan digunakan. Atribut yang akan digunakan yaitu atribut Alamat dan Diagnosa, selanjutnya dilakukan pengolahan dengan analisis data *mining* dan atribut data yang tidak digunakan akan dibuang.

2. *Pre-processing/Cleaning Data*, *dataset* yang telah dibangun pada tahapan sebelumnya kemudian dilakukan pembersihan untuk mendapatkan *dataset* yang lebih baik. Pada dasarnya data rekam medis pasien terdapat data yang lengkap dan juga yang tidak lengkap, termasuk data yang memiliki nilai yang kosong (*missing value*). Pada proses data *mining* data seperti itu akan memberikan hasil yang kurang baik pada akhirnya. Maka dari itu diperlukan satu tahap yaitu tahap *pre-processing* atau tahap *cleaning* data yang harus dilakukan pada data pasien, agar data yang tidak diperlukan maupun kosong dapat dihilangkan.
3. *Transformation Data*, Pada tahap *transformation* data menggunakan pendekatan/teknik atribut *construction* dimana pembentukan atribut baru dari atribut yang sudah ada untuk membantu meningkatkan ketelitian/ketepatan dan pemahaman struktur data. Pembentukan atribut berdasarkan atribut hasil dari *pre-processing/cleaning* data yaitu atribut Alamat dan Diagnosa dengan membentuk hasil kalkulasi jumlah diagnosa penyakit dari setiap daerah di ruang lingkup Puskesmas Kecamatan Jatiluhur. Berikut merupakan data yang telah dilakukan transformasi, dapat dilihat pada Tabel 2.

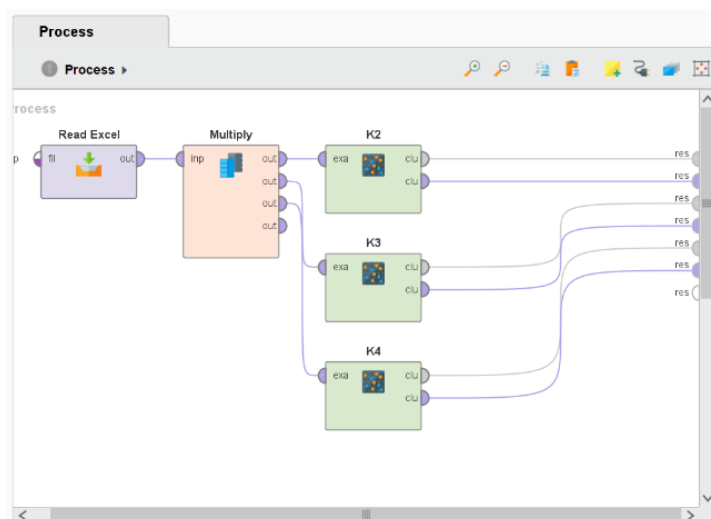
Tabel 2. Data yang Telah Dikakukan Transformasi

Diagnosa	Jumlah kasus	Diagnosa	Jumlah kasus	Diagnosa	Jumlah kasus
<i>Abses</i>	47	Kanker Payudara	4	<i>Faringitis</i>	12
<i>Amenorea</i>	11	Katarak	32	<i>Febris</i>	116
<i>Anemia</i>	14	Kelainan Struktur Tulang	1	<i>Fibrilasi Atrium</i>	27
<i>Appendicitis</i>	2	<i>Konjungtivitis</i>	36	<i>Fibroadenoma</i>	4
<i>Arthralgia</i>	46	<i>Konstipasi</i>	5	<i>Fraktur Tibia</i>	1
Asma	62	Koreng	1	<i>Gastritis</i>	17
<i>Asma Bronkial</i>	7	Limpa <i>Denopati</i>	11	<i>Gastroenteritis</i>	70
<i>Autoimun Disease</i>	4	<i>Lipoma</i>	1	GERD	6
B20	4	<i>Low Back Pain</i>	36	<i>Gout</i>	23
Batu Ginjal	1	Malaria	18	<i>Hearing Loss Unspecified</i>	1
<i>Bell Palsy</i>	1	<i>Mastoiditis</i>	2	<i>Hemanggiyoma</i>	4
<i>Benign Neoplasm</i>	2	<i>Menometroragia</i>	4	<i>Hemonoid</i>	1
<i>Benign Prostatic Hyperplasia</i>	3	<i>Myalgia</i>	55	<i>Hemoptoe</i>	3
<i>Bronkitis</i>	32	<i>Myopia</i>	9	<i>Hemoroid</i>	8
<i>Bronkitis Kronis</i>	2	<i>Nasofaringitis Akut</i>	2	<i>Hepatitis B</i>	4
<i>Bronkopneumonia</i>	29	<i>Neprolithiasis</i>	6	<i>Hernia</i>	6
<i>Cardio Vascular Disease</i>	2	<i>Neurodermatitis</i>	15	<i>Hernia Inguinal</i>	5
<i>Cephalgia</i>	55	<i>Oedema</i>	3	<i>Hernia Nukleus Pulposus</i>	4
<i>Cervitis</i>	2	Osteoarthritis	54	<i>Hipertensi</i>	149
<i>Ches Pain</i>	8	<i>Otagia</i>	1	<i>Hordeolum</i>	8
<i>Chronic Kidney Diseas</i>	8	<i>Otitis Media Akut</i>	16	<i>Hypertensive Heart Diseases</i>	34
<i>Chronic Liver Disease</i>	1	<i>Otitis Media Kronik</i>	5	<i>Impetigo</i>	8
<i>Clavus</i>	13	<i>Otitis Media Perforata</i>	1	Iritasi Mata	1
<i>Colitis</i>	2	<i>Otitis Media Sufuratif Kronik</i>	1	ISPA	163

Penentuan Jumlah Cluster Optimal Menggunakan Davies Bouldin Index pada Algoritma K-Means untuk Menentukan Kelompok Penyakit

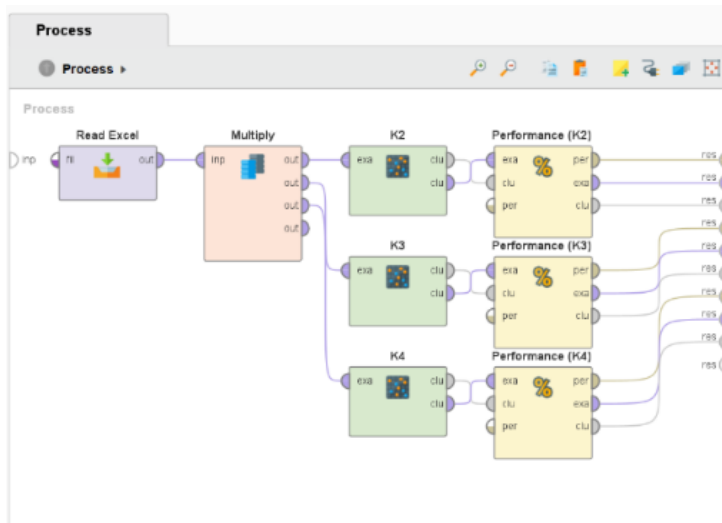
Diagnosa	Jumlah kasus	Diagnosa	Jumlah kasus	Diagnosa	Jumlah kasus
<i>Common Cold</i>	129	<i>Papiloma</i>	3	TB <i>Abdenoma</i>	1
<i>Congestive Heart Failure</i>	16	<i>Paralisis Nervus Fasialis</i>	1	TB Kelenjar	3
<i>Coronary Artery Disease</i>	22	<i>Paronikia</i>	3	TB Paru	111
<i>Coronary Heart Disease</i>	3	<i>Parotitis</i>	8	TBC	32
<i>Cystitis</i>	29	Paru <i>Obstruktif Kronik</i>	25	<i>Thalasemia</i>	1
<i>Dacriocititis</i>	1	<i>Pertusis</i>	1	<i>Tic Fasialis</i>	1
Demam <i>Tipoid</i>	3	<i>Prolaps Uteri</i>	1	<i>Tinea</i>	20
Demam Berdarah <i>Dangue</i>	2	<i>Pseudophakia</i>	3	<i>Tinea Capitis</i>	1
<i>Demensia</i>	1	<i>Ptyriasis Rosea</i>	5	<i>Tinea Corporis</i>	2
<i>Dermatitis</i>	86	<i>Rheumatoid Arthritis</i>	1	<i>Tinea Seboroik</i>	1
<i>Dermatitis Kontak Alergi</i>	38	<i>Rhinitis</i>	25	<i>Tinitus Kronis</i>	4
<i>Dermatitis Numularis</i>	1	<i>Rhino Faringitis Akut</i>	6	<i>Tonsilo Faringitis Akut</i>	7
<i>Dermatitis Seboroik</i>	4	<i>Scabies</i>	5	<i>Tonsilofaringitis</i>	31
<i>Diabetes Melitus</i>	68	<i>Selulitis</i>	10	<i>Tumor Laryns</i>	1
<i>Difteri</i>	1	<i>Sequale Stroke</i>	1	<i>Typoid Fever</i>	6
<i>Disentri</i>	2	<i>Sindrome Nefrotik</i>	1	<i>Urtikaria</i>	34
<i>Dislipidemia</i>	12	<i>Sinusitis</i>	3	<i>Varicella</i>	15
<i>Dyspepsia</i>	132	<i>Skizofrenia</i>	31	<i>Varises</i>	1
<i>Epilepsi</i>	15	<i>Stomatitis</i>	24	<i>Vertigo</i>	39
<i>Epistaksis</i>	3	<i>Stroke</i>	28	<i>Vomitus</i>	6
<i>Facial Nerve Disorder</i>	2	<i>Takikardia</i>	6	<i>Vulnus Excoriasi</i>	4
<i>Vulnus Scissum</i>	3	<i>Vulvus Lecaratum</i>	9		

4. Data Mining (Algoritma *K-means*), pada tahapan ini dilakukan proses pemodelan data *mining* terhadap *dataset* rekam medis pasien yang telah melalui proses tahapan sebelumnya dengan menggunakan algoritma *k-means* untuk dilakukan penentuan kelompok penyakit. Pada tahapan pemodelan data *mining* ini akan dilakukan *testing* pembentukan beberapa *cluster* yaitu K2, K3 dan K4 untuk mengetahui *cluster* yang paling optimal. Pada tahapan ini pemodelan data *mining* dengan algoritma *k-means* akan di implementasikan menggunakan aplikasi *RapidMiner* dengan operator yang digunakan sebagai berikut: Satu operator *Read Excel* digunakan untuk melakukan *import file* data rekam medis berekstensi *.xlsx (Excel)*, satu operator *Multiply* digunakan untuk memperbanyak pengujian pada operator *k-means* dan 3 operator *k-means*, digunakan untuk pengelompokan data dengan algoritma *k-means* dengan masing-masing operator dibentuk *cluster* sebanyak K2, K3 dan K4. Berikut merupakan hasil rancangan model *k-means* pada aplikasi *RapidMiner*, dapat dilihat pada Gambar 3.



Gambar 3. Rancangan Model *K-Means*

5. Evaluasi (*Davies Bouldin Index*), Pada tahapan evaluasi bertujuan untuk dilakukan pengujian terhadap kualitas dari hasil *cluster* yang terbentuk yaitu K2, K3 dan K4 dengan menggunakan *Davies Bouldin Index*. Untuk mendapatkan nilai *Davies Bouldin Index*, dilakukan penambahan operator *Performance Distance* pada *modelling k-means clustering* dalam aplikasi *RapidMiner* digunakan untuk mengetahui nilai *Davies Bouldin Index* pada setiap *cluster* yang terbentuk. Berikut merupakan rancangan model *k-means* dengan dengan penambahan operator *Performance Distance*, dapat dilihat pada Gambar 4.



Gambar 4. Modelling *K-Means* dengan Operator *Performance Distance* (DBI)

6. *Knowledge Presentation*, Tahapan selanjutnya adalah memvisualisasikan data atau *knowledge presentation*, dimana data rekam medis pasien yang telah dilakukan proses data *mining* akan divisualisaikan agar lebih mudah dipahami dan diharapkan dapat diambil tindakan berdasarkan analisis.

3.3 Kesimpulan

Kesimpulan merupakan tahapan menyimpulkan dari keseluruhan hasil penelitian yang telah dilakukan.

4. HASIL DAN PEMBAHASAN

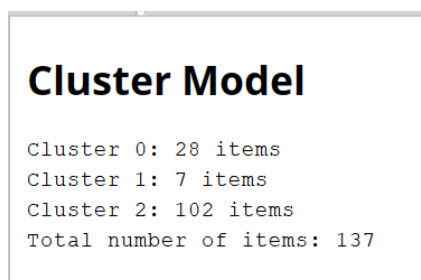
Proses pengolahan data dengan menggunakan algoritma *k-means clustering* dan dilakukan evaluasi pengujian dari setiap *cluster* yang di bentuk yaitu K1, K2 dan K3 dengan menggunakan *Davies Bouldin Index*, maka didapatkan hasil nilai *Davies Bouldin Index* dari setiap *cluster* yang dibentuk. Berikut merupakan hasil nilai *Davies Bouldin Index*, dapat dilihat pada Tabel 3.

Tabel 3. Hasil Nilai *Davies Bouldin Index* (DBI) Pada Setiap *Cluster*

<i>Cluste</i> yang terbentuk	Nilai DBI
K2	0,408
K3	0,406
K4	0,409

Berdasarkan nilai DBI dari sejumlah K uji yang terdapat pada Tabel 3. Dapat diketahui bahwa nilai DBI yang paling kecil terdapat pada K = 3 yaitu sebesar 0,406. Maka dapat dikatakan *cluster* paling optimal adalah tiga *cluster*, karena semakin kecil nilai DBI yang diperoleh (*non-negatif* ≥ 0), maka semakin baik *cluster* yang diperoleh dari proses peng*clusteran k-means* yang dilakukan (Ramadhani et al., 2022).

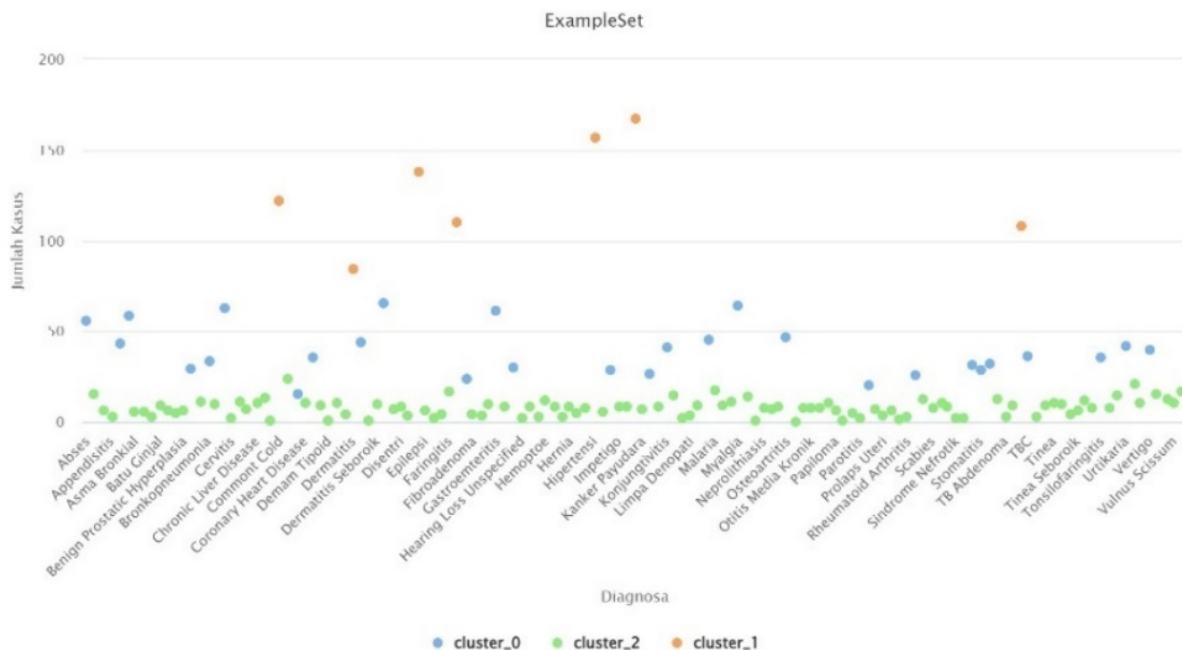
Dari hasil pengujian terhadap *cluster* yang paling optimal dengan menggunakan evaluasi *Davies Bouldin Index*, *cluster* optimal terdapat pada *cluster* yang terbentuk sebanyak 3 kelompok. Berikut merupakan hasil data rekam medis setelah dilakukan proses *k-means clustering* untuk dilakukan pengelompokan penyakit, dengan *cluster* yang terbentuk sebanyak 3 kelompok dengan jumlah anggota dari masing-masing cluster. Dapat dilihat pada Gambar 5.



Gambar 5. Hasil *Cluster Model* dengan Jumlah Anggota *Cluster*

Dari Gambar 5, dapat diketahui untuk *cluster*0 memiliki 28 anggota diagnosa penyakit, *cluster* 1 memiliki 7 anggota diagnosa penyakit dan *cluster*2 memiliki 102 anggota diagnosa penyakit.

Berikut merupakan visualisasi hasil sebaran kelompok penyakit dengan menggunakan *scatter/bubble chart*, dapat dilihat pada Gambar 6.



Gambar 6. Visualisasi Sebaran Kelompok Penyakit

Dari hasil visualisasi pada Gambar 6, dapat diketahui dari hasil analisis bahwa setiap *cluster* untuk kategori kelompok penyakit berdasarkan jumlah kasus tingkat sebaran di bedakan menjadi:

1. *Cluster 0*, kelompok penyakit dengan tingkat sebaran sedang
2. *Cluster 1*, kelompok penyakit dengan tingkat sebaran tinggi
3. *Cluster 2*, kelompok penyakit dengan tingkat sebaran rendah

Untuk data anggota kelompok penyakit berdasarkan *clusternya* dapat dilihat pada Tabel 4 berikut.

Tabel 4. Anggota Kelompok Penyakit Berdasarkan Clusternya

Cluster	Jumlah anggota	Diagnosa penyakit
0	28	<i>Abses, Arthralgia, Asma, Bronkitis, Bronkopneumonia, Cephalgia, Coronary Artery Disease, Cystitis, Dermatitis Kontak Alergi, Diabetes Melitus, Fibrilasi Atrium, Gastroemteritis, Gout, Hypertensive Heart Diseases, Katarak, Konjungtivitis, Low Back Pain, Myalgia, Osteoarthritis, Paru Obstruktif Kronik, Rhinitis, Skizofrenia, Stomatitis, Stroke, TBC, Tonsilofaringitis, Urtikaria, Vertigo</i>
1	7	<i>Common Cold, Dermatitis, Dyspepsia, Febris, Hipertensi, ISPA, TB Paru</i>
2	102	<i>Amenorea, Anemia, Appendisitis, Asma Bronkial, Autoimun Disease, B20, Batu Ginjal, Bell Palsy, Benign Neoplasm, Benign Prostatic Hyperplasia, Bronkitis Kronis, Cardio Vascular Disease, Cervitis, Ches Pain, Chronic Kidney Diseas, Chronic Liver Disease, Clavus, Colitis, Congestive Heart Failure, Coronary Heart Disease, Dacriocititis, Demam Tipoid, Deman Berdarah Dangué, Demensia, Dermatitis Numularis, Dermatitis Seboroik, Difteri, Disentri, Dislipidemia, Epilepsi, Epistaksis, Facial Nerve Disorder, Faringitis, Fibroadenoma, Fraktur</i>

Penentuan Jumlah Cluster Optimal Menggunakan Davies Bouldin Index pada Algoritma K-Means untuk Menentukan Kelompok Penyakit

Cluster	Jumlah anggota	Diagnosa penyakit
		<i>Tibia, Gastritis, GERD, Hearing Loss Unspecified, Hemanggiyoma, Hemonoid, Hemoptoe, Hemoroid, Hepatitis B, Hernia, Hernia Inguinal, Hernia Nukleus Pulposus, Hordeolum, Impetigo, Iritasi Mata, Kanker Payudara, Kelainan Struktur Tulang, Konstipasi, Koreng, Limpa Denopati, Lipoma, Malaria, Mastoiditis, Menometroragia, Myopia, Nasofaringitis Akut, Nephrolithiasis, Neurodermatitis, Oedema, Otagia, Otitis Media Akut, Otitis Media Kronik, Otitis Media Perforata, Otitis Media Sufuratif Kronik, Papiloma, Paralisis Nervus Fasialis, Paronikia, Parotitis, Pertusis, Prolaps Uteri, Pseudophakia, Ptyriasis Rosea, Rheumatoid Arthritis, Rhino Faringitis Akut, Scabies, Selulitis, Sequale Stroke, Sindrome Nefrotik, Sinusitis, Takikardia, TB Abdenoma, TB Kelenjar, Thalasemia, Tic Fasialis, Tinea, Tinea Capitis, Tinea Corporis, Tinea Seboroik, Tinnitus Kronis, Tonsilo Faringitis Akut, Tumor Laryns, Typhoid Fever, Varicella, Varises, Vomitus, Vulnus Excoriasi, Vulnus Scissum, Vulvus Lecaratum</i>

Dari Tabel 4 diatas, dapat diketahui bahwa diagnosa penyakit *Common Cold, Dermatitis, Dyspepsia, Febris, Hipertensi, ISPA, TB Paru* memiliki tingkat kasus yang paling banyak persebarannya di lingkungan Puskesmas Jatiluhur, maka dapat dijadikan sebagai prioritas untuk penyuluhan terhadap kelompok penyakit tersebut guna menekan tingkat persebarannya.

5. KESIMPULAN

Berdasarkan dari hasil pengolahan data rekam medis yang telah dipaparkan sebelumnya dengan menggunakan algoritma *k-means* dengan evaluasi pengujian menggunakan *Davies Bouldin Index* (DBI) pada setiap *cluster* yang dibentuk yaitu K2, K3 dan K4. Maka dapat disimpulkan bahwa algoritma *k-means* dengan evaluasi pengujian dengan nilai DBI dapat menentukan jumlah *cluster* optimal, dengan *cluster* yang terbentuk sebanyak K3. Hasil *clustering* yang diperoleh berupa pengelompokan penyakit berdasarkan jumlah kasus sebaran dengan kategori rendah, sedang dan tinggi. Dikarenakan keterbatasan waktu dan tenaga, peneliti menyadari bahwa hasil penelitian ini masih jauh dari kata sempurna. Karena hal tersebut untuk penelitian berikutnya bisa dilakukan komparasi maupun kombinasi dengan metode yang lain.

6. DAFTAR RUJUKAN

- Amrulloh, K. (2023). *Comparison Between Davies-Bouldin Index and Silhouette Coefficient Evaluation Methods in Retail Store Sales Transaction Data Clusterization Using K-Medoids Algorithm*. 1952–1961. <https://doi.org/10.46254/sa03.20220384>
- Anggraini, R., Haerani, E., Jasril, J., & Afrianty, I. (2022). Pengelompokan Penyakit Pasien Menggunakan Algoritma K-Means. *JURIKOM (Jurnal Riset Komputer)*, 9(6), 1840. <https://doi.org/10.30865/jurikom.v9i6.5145>
- Azhari, S. M., Pudjiantoro, T. H., & Santikarama, I. (2021). Klasterisasi Outlet Berdasarkan Data Penjualan Dengan Menggunakan Algoritma K-Medoids. *JUMANJI (Jurnal Masyarakat Informatika Unjani)*, 5(2), 69. <https://doi.org/10.26874/jumanji.v5i2.93>
- Irwan, I., Sanusi, W., Anwar, A. S., & Rahman, A. (2023). The Implementation of Spatial Model with K-Means Clustering Method to Cluster Flood Affected Areas in Bone Regency. *ARRUS Journal of Social ...*, 3(2), 186–195.

<https://sainsmat.org/index.php/soshum/article/view/1771%0Ahttps://sainsmat.org/index.php/soshum/article/download/1771/1114>

- Mayadi, Siti Setiawati, W. P. (2022). *Pelita teknologi*. 17(2), 1–11.
- Muliono, R., & Sembiring, Z. (2019). Data Mining Clustering Menggunakan Algoritma K-Means Untuk Klasterisasi Tingkat Tridarma Pengajaran Dosen. *CESS (Journal of Computer Engineering, System and Science)*, 4(2), 2502–2714.
- Muningsih, E., Maryani, I., & Handayani, V. R. (2021). Penerapan Metode K-Means dan Optimasi Jumlah Cluster dengan Index Davies Bouldin untuk Clustering Propinsi Berdasarkan Potensi Desa. *Jurnal Sains Dan Manajemen*, 9(1), 95–100. <https://ejournal.bsi.ac.id/ejurnal/index.php/evolusi/article/view/10428/4839>
- Muttaqin, M. R., & Defriani, M. (2020). Algoritma K-Means untuk Pengelompokan Topik Skripsi Mahasiswa. *ILKOM Jurnal Ilmiah*, 12(2), 121–129. <https://doi.org/10.33096/ilkom.v12i2.542.121-129>
- Nurahman, N., Purwanto, A., & Mulyanto, S. (2022). Klasterisasi Sekolah Menggunakan Algoritma K-Means berdasarkan Fasilitas, Pendidik, dan Tenaga Pendidik. *MATRIK: Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer*, 21(2), 337–350. <https://doi.org/10.30812/matrik.v21i2.1411>
- Prasetyo, M., Ansori, N. A., Firdausi, Y., & Wahid, M. F. (2023). Implementation of K-Means Clustering for Analysis Students English Proficiency. *Science and Technology /*, 1(1), 31–35.
- Ramadhani, S., Azzahra, D., & Z, T. (2022). Comparison of K-Means and K-Medoids Algorithms in Text Mining based on Davies Bouldin Index Testing for Classification of Student's Thesis. *Digital Zone: Jurnal Teknologi Informasi Dan Komunikasi*, 13(1), 24–33. <https://doi.org/10.31849/digitalzone.v13i1.9292>
- Rohmatullah, A., Rahmalia, D., & Pradana, M. S. (2020). Klasterisasi Data Pertanian Di Kabupaten Lamongan Menggunakan Algoritma K-Means Dan Fuzzy C Means. *Jurnal Ilmiah Teknosains*, 5(2), 86–93. <https://doi.org/10.26877/jitek.v5i2.4254>
- Rustam, S., Santoso, H. A., & Supriyanto, C. (2018). Optimasi K-Means Clustering Untuk Identifikasi Daerah Endemik Penyakit Menular Dengan Algoritma Particle Swarm Optimization Di Kota Semarang. *ILKOM Jurnal Ilmiah*, 10(3), 251–259. <https://doi.org/10.33096/ilkom.v10i3.342.251-259>
- Sari, M. A. P., & Chotijah, U. (2022). Pengelompokan Anggota Divisi Himpunan Mahasiswa Jurusan Pada Universitas "Xyz" Dengan Metode K-Means Clustering. *Antivirus: Jurnal Ilmiah Teknik Informatika*, 16(1), 52–62. <https://doi.org/10.35457/antivirus.v16i1.2139>
- Septiani, I. W., Fauzan, A. C., & Huda, M. M. (2022). Implementasi Algoritma K-Medoids Dengan Evaluasi Davies-Bouldin-Index Untuk Klasterisasi Harapan Hidup Pasca Operasi Pada Pasien Penderita Kanker Paru-Paru. *Jurnal Sistem Komputer Dan Informatika (JSON)*, 3(4), 556. <https://doi.org/10.30865/json.v3i4.4055>
- Setiyani, L., Wahidin, M., Awaludin, D., & Purwani, S. (2020). Analisis Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Data Mining Naïve Bayes: Systematic

Penentuan Jumlah Cluster Optimal Menggunakan Davies Bouldin Index pada Algoritma K-Means untuk Menentukan Kelompok Penyakit

Review. *Faktor Exacta*, 13(1), 35. <https://doi.org/10.30998/faktorexacta.v13i1.5548>

Sisfotenika, J. (2022). *Pengelompokkan Kebutuhan Air Bersih Di Indonesia Periode 2012-2017 Menggunakan Algoritma K-Means Clustering of Clean Water Needs in Indonesia for the 2012-2017 Period Using the K-Means Algorithm*. 12(2), 203–212.

Srimurdianti, A., Sukamto, Setiawan, W., Esyudha, E., & Pratama. (2023). *JEPIN (Jurnal Edukasi dan Penelitian Informatika) Data Mining untuk Pengelompokan Saham pada Sektor Energi dengan Metode K-Means*. 9(1), 76–81.

Wahyudi, M., & Pujiastuti, L. (2022). Comparison of K-Means Clustering and K-Medoids in Clustering Fresh Milk Production in Indonesia. *Jurnal Bumigora Information Technology (BITE)*, 4(2), 243–254. <https://doi.org/10.30812/bite.v4i2.2104>