IMPLEMENTASI ALGORITMA K-MEANS CLUSTERING UNTUK KLASIFIKASI DIAGNOSIS PENYAKIT LAMBUNG

Tuti Hartini [1], Ade Irma Purnamasari [2], Agus Bahtiar [3], Kaslani [4]
[1,2] Teknik Informatika, STMIK IKMI Cirebon
[3] Sistem Informasi, STMIK IKMI Cirebon
[4] Komputerisasi Akuntansi, STMIK IKMI cirebon
Jl. Perjuangan No.10B, Karyamulya, Kec. Kesambi, Kota Cirebon, Jawa Barat 45135
tuti86681@gmail.com

ABSTRAK

Keterbatasan akses layanan Kesehatan spesialis untuk diagnosis penyakit lambung masih menjadi tantangan utama di wilayah pedesaan seperti desa Sukamulya. Hal ini, menyebabkan keterlambatan penanganan dan potensi memburuknya kondisi Kesehatan Masyarakat setempat. Penelitian ini bertujuan untuk mengembangkan model deteksi penyakit lambung menggunakan Algoritma K-means Clustering yang disesuaikan dengan kondisi Kesehatan Masyarakat setempat. Metode yang digunakan adalah Algoritma K-means Clustering dengan menganalisis data dari 300 pasien yang dikumpulkan selama 3 bulan di puskesmas Desa Sukamulya. Attribut yang digunakan mencangkup usia, jenis gejala, dan diagnosis awal untuk pengelompokkan penyakit lambung seperti Gastritis, ulkus peptikum, GERD, dan maag. Hasil penelitian ini menunjukkan nilai *Davies Bouldin Index* (DBI) terbaik ada pada kluster 6 dengan nilai DBI -0.488, dengan distribusi anggota: kluster 0: 68 items, cluster 2: 9 items, cluster 3: 49 items, cluster 4: 59 items, dan kluster 5: 57 items. Model ini juga berpotensi memberikan wawasan baru tentang penggunaan teknologi pembelajaran mesin dalam konteks Kesehatan pedesaan dan dapat menjadi landasan untuk pengembangan kebijakan Kesehatan yang lebih berbasis teknologi dan inklusif. Hasil penelitian ini diharapkan dapat mendorong adopsi teknologi dalam bidang pelayanan Kesehatan serta meningkatkan literasi digital Masyarakat dan tenaga Kesehatan di wilayah pedesaan.

Kata kunci: *K-means clustering, penyakit lambung, Davies Bouldin Index, diagnosis penyakit, kesehatan pedesaan, preprocessing data.*

1. PENDAHULUAN

Perkembangan teknologi merupakan suatu hal yang dapat mempengaruhi segala aspek kehidupan salah satunya dibidang informatika dan bidang informasi. Bidang informatika yang berkembang pesat telah mengubah banyak aspek kehidupan manusia. Saat ini, banyak industry seperti ekonomi, kesehatan, dan Pendidikan bergantung pada teknologi informasi dan komunikasi. Dengan kemajuan dalam analisis data dan kecerdasaan buatan, ada peluang baru untuk memecahkan masalah yang kompleks yang dihadapi masyarakt saat ini. Dalam situasi seperti ini, penggunaan Teknik pengajaran mesin seperti clustering untuk meningkatkan deteksi dan diagnosis penyakit menjadi semakin penting. Terutama di wilayah dengan akses terbatas terhadap layanan kesehatan spesialis[1].

Salah satu masalah utama dalam sistem kesehatan, terutama di wilayah pedesaan seperti Desa Sukamulya, adanya keterbatasan akses terhadap diagnosis penyakit lambung yang tepat dan cepat. Karena gejala yang tidak jelas dan kurangnya sarana diagnostik yang memadai. Penyakit lambung, yang mencangkup berbagai penyakit seperti gastritis, tukak lambung, dan kanker lambung sering kali tidak ditemukan pada tahap awal. Hal ini dapat mempersulit penanganan dan meningkatkan morbiditas serta mortalitas. Penggunaan teknologi pembelajaran mesin, terutama algoritma clustering K-means dapat meningkatkan akurasi dan efesiensi deteksi penyakit lambung. Namun, penerapan teknologi ini di wilayah pedesaan menghadapi sejumlah masalah. Beberapa di antaranya adalah keterbatasan infrastruktur teknologi, kekurangan data representative, dan ketidaksetujuan terhadap adopsi teknologi baru dalam praktik medis konvensional. Selain itu, tenaga kesehatan lokal tidak tahu cara menggunakan dan memahami hasil dari sistem berbasis AI. Ini adalah masalah lain yang muncul karena privasi dan keamanan data pasien, karena data sensitive tentang pasien digunakan dalam analisis. Selain itu, ada kemungkinan bahwa praktisi kesehatan yang terbiasa dengan metode konvensional akan menentang panggabungan sistem AI dengan protocol kesehatan yang sudah ada. Oleh karena itu, Solusi berbasis AI harus diintegrasikan kedalam sistem kesehatan lokal dengan pendekatan yang holistic dan adaptif. Pendekatan ini juga harus mempertimbangkan konteks sosial-budaya dan keterbatasan sumber daya saat ini. Akhirnya, penting untuk mempertimbangkan keberlanjutan penggunaan teknologi ini. Penting untuk mempertimbangkan keberlanjutan penggunaan teknologi ini termasuk, memberikan tenaga kesehatan pelatihan yang berkelanjutan dan memperbarui sistem sesuai dengan kemajuan teknologi dan ilmu pengetahuan[2].

Studi sebelumnya telah menunjukkan bahwa algoritma clustering dapat membantu mendeteksi berbagai kondisi medis dengan lebih baik. Misalnya menggunakan algoritma clustering untuk menganalisis data pasien gastritis, namun masih terbatas pada

populasi urban dengan akses kesehatan yang memadai.[3] Selain itu menggunakan algoritma Fuzzy k-means untuk mengklasifikasikan gambar endoskopi lambung. Mereka menemukan bahwa mereka dapat mendeteksi lesi pra-kanker dengan sensitivitas 92%. Namun, penelitian ini dilakukan fasilitas kesehatan perkotaan yang memiliki akses ke pelaratan diagnostik canggih [4].

Sebaliknya, melihat penggunaan algoritma clustering untuk triase pasien di daerah pedesaan india. Penelitian mereka menunjukkan bahwa teknologi ini dapat digunakan dalam situasi dengan sumber daya terbatas, meskipun mereka berpokus pada penyakit lambung secara khsusus.[5]

Tujuan utama dari penelitian ini adalah membuat dan menilai model deteksi penyakit lambung yang dapat digunakan di Desa Sukamulya yang bergantung pada algoriitma clustering K-means. Selain itu, tujuan dari penelitian ini adalah untuk meningkatkan akurasi diagnosis awal, mengoptimalkan alokasi sumber daya kesehatan yang terbatas, dan pada akhirnya meningkatkan hasil kesehatan Masyarakat desa. Penelitian ini dapat membantu mengatasi keterbatasan dalam mendapatkan diagnosis yang tepat di daerah pedesaan dan membantu membangun Solusi teknologi kesehatan yang sesuai dengan konteks lokal. Penelitian ini juga diharapkan dapat membantu mengubah layanan kesehatan pedesaan menjadi digital, mendorong inovasi dan meningkatkan kualitas pelayanan kesehatan secara keseluruhan. Terakhir temuan penelitian ini dapat membantu pembuat kebijakan membuat strategi kesehatan berbasis teknologi yang inklusif dan berkelanjutan untuk daerah-daerah terpencil.

Dalam melaksanakan penelitian ini, kami akan menggunakan metode pendekatan berbasis data, yaitu metode K-Means. Metode K-Means adalah sebuah teknik pengelompokan data yang akan memungkinkan kami untuk mengelompokkan penyakit lambung di puskesmas desa sukamulya ke dalam kelompok-kelompok yang memiliki karakteristik serupa. Data penyakit yang akan kami gunakan melibatkan variabel seperti usia, jenis gejala, dan diagnosis awal. Proses analisis akan dimulai dengan pra-pemrosesan data, termasuk normalisasi data dan pemilihan jumlah kluster yang optimal. Selanjutnya, kami akan menerapkan algoritma K-Means untuk mengelompokkan data penyakit ke dalam kelompok-kelompok yang relevan. Hasil analisis ini akan membantu kami dalam merumuskan rekomendasi strategi yang lebih efektif dan mendukung pengembangan penyakit lambung di puskesmas desa sukamulya secara keseluruhan. Hasil penelitian ini diharapkan memiliki dampak yang signifikan terhadap praktik kesehatan di wilayah pedesaan. Jika berhasil, model deteksi berbasis K-means clustering ini dapat membantu tenaga kesehatan lokal dalam triase dan diagnosis awal penyakit lambung. Ini dapat mengurangi beban pada sistem rujukan dan memungkinkan intervensi dini yang lebih baik untuk meningkatkan prognosis pasien.

2. TINJAUAN PUSTAKA

2.1. Penyakit lambung

Lambung Merupakan organ pencernaan yang berperan penting dalam proses pencernaan makanan. Menurut WHO [1], penyakit lambung dapat diklasifikasikan menjadi beberapa jenis utama, termasuk gastritis, tukak lambung, dan kanker lambung. Setiap jenis penyakit memiliki karakteristik gejala yang berbeda namun sering tumpeng tindih, seperti nyeri ulu hati, mual, muntah, dan rasa terbakar di dada. Diagnosis yang tepat memerlukan analisis terhadap kombinasi gejala yang muncul, Riwayat medis pasien, dan faktor risiko yang ada.

2.2. Algoritma K-means Clustering

K-means clustering merupakan algoritma pembelajaran tanpa pengawasan (unsupervised learning) yang bertujuan mengelompokkan data ke dalam k-means berdasarkan kemiripan karakteristiknya. Prinsip dasar algoritma ini adalah meminimalkan sum of squared distances antara titik data dengan centroid cluster terdekat. Tahapan algoritma K-means meliputi:

a. Inisialisasi K centroid secara acak
b. Perhitungan jarak setiap data ke masing-masing centroid
c. Pengelompokkan data cluster dengan centroid terdekat
d. Pembaruan posisi centroid
e. Pengulangan Langkah 2-4 hingga konvergen

Formula dasar yang digunakan dalam K-means adalah:

a. Euclidean distance:
$d(x,y) = \sqrt{(\Sigma(xi - yi)^2)}$

b. Centroid Update:
$Cj = (1/|Sj|)\Sigma xi,\ xi \in Sj$

Dimana:

- d(x,y), adalah jarak antara data x dan y
- Cj adalah centroid cluster ke-j
- Sj adalah himpunan data dalam cluster ke-j

2.3. Evaluasi performa Clustering

Evaluasi hasil clustering dapat dilakukan menggunakan beberapa metrik, antara lain:

a. Silhouette coefficient
b. Davies-Bouldin index
c. Calinski-Harabasz Index

Masing-masing metrik memiliki karakteristik dan interpretasi yang berbeda dalam menilai kualitas clustering.

2.4. Penelitian Terkait

Beberapa penelitian terdahulu telah menerapkan metode clustering dalam diagnosis penyakit.

"A pattern-recognition-based clustering method for non-invasive diagnosis and classification of

various gastric conditions" Mendemontrasikan metodologi baru non-invasif untuk diagnosis dan klasifikasi akurat berbagai gangguan lambung menggunakan analisis kluster berbasis pengenalan pola dari dataset breathomics. Metode ini mampu membedakan napas pasien dengan ulkus peptic dan disfungsi lambung lainnya dari napas individu sehat dengan sensitivitas dan spesifisitas diagnostic yang tinggi.[6]

"An impored Gabor wavelet transform and rough K-means clustering algorithm for MRI brain tumor image segmentation" Mengusulkan metodologi segmentasi citra MRI otak menggunakan transformasi wavelet gabor yang ditingkatkan dan algoritma Rough K-means. Hasil eksprimen menunjukkan metode yang diusulkan mencapai hasil yang lebih baik dibandingkan dengan metode yang ada dalam hal sensitivitas, spesifisitas, dan akurasi.[7]

A novel Hybrid K-means and GMM machine learning Model for breast cancer detection" Mengusulkan model hybrid K-means dan Gaussian mixture model (GMM) untuk deteksi kanker payudara. Metode ini menunjukkan kinerja yang baik dalam mengklasifikasikan tumor jinak dan ganas, memungkinkan para ahli medis untuk mengenali kanker payudara dengan lebih cepat dan akurasi yeng lebih tinggi.[8]

Machine learning algorithm in healthcare system: A review" oleh (Kushwaha & Kumaresan, 2021) mengkaji berbagai Teknik machine learning yang digunakan dalam sistem kesehatan. Penelitian ini menekankan kemampuan algoritma machine learning, termasuk K-means dalam menganalisis data medis yang kompleks dengan cepat dan akurat. Hasilnya menunjukkan potensi besar machine learning dalam mendukung pengambilan Keputusan kesehatan.[9]

Data-driven versus a domain-led approach to K-means clustering on an open heart failure dataset" Membandingkan pendekatan berbasis domain dan berbasis data dalam eksprimen clustering K-means pada dataset gagal jantung terbuka. Hasil penelitian menunjukkan bahwa pendekatan berbasis data memiliki keungulan data sleksi fitur dengan menghasilkan resiko bias yang dapat diperkenalkan oleh ahli domain. Namun, pengetahuan domain tetap berperan penting dalam interpretasi hasil clustering.[10]

2.5. Preprocessing Data

Kualitas hasil clustering sangat bergantung pada preprocessing data yang dilakukan. Beberapa Teknik preprocessing yang umum digunakan meliputi:

a. Normalisasi Data
   - Min-max Scaling
   - Z-score Standardization

b. Penanganan missing Value
   - Mean imputation
   - Median imputation
   - Mode imputation

c. Feature Selection
   - Principal Component analysis (PCA)
   - Chi-square Test
   - Information Gain

2.6. Optimasi Parameter K-means

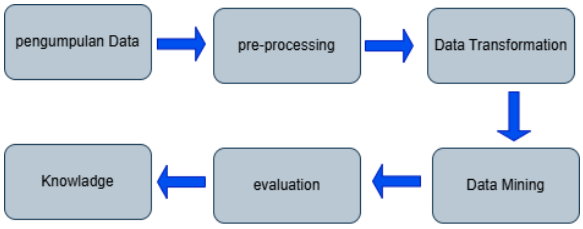
Penentuan jumlah cluster optimal merupakan aspek kritis dalam implementasi K-means. Beberapa metode yang dapat digunakan untuk menentukan nilai K optimal:

a. Elbow Method menganalisis perubahan within-cluster sum of squares (WSS) untuk berbagai nilai k
b. Gap statistic membandingkan total within intra-cluster variation dengan nilai ekspektasinya
c. Cross-validation mengevaluasi cluster menggunakan data training dan testing.

3. METODE PENELITIAN

3.1. Metode Penelitian

Pada penelitian ini menggunakan metode penelitian K-Means Clustering dengan pendekatan kuantitatif yang dimana pada hasilnya akan menampilkan angka statistic dalam bentuk visualisasi grafik. Metode K-Means clustering merupakan suatu bentuk pengelompokkan sebuah nilai pada subjek yang diambil nilai berdekatan kemudian dijadikan sebagai satu kelompok, dan apabila nilai tersebut saling berjauhan maka akan dibuat kelompok baru. Penentuan kelompok tergantung oleh seseorang yang akan disaring menjadi beberapa kelompok, umumnya sebuah angka statistik dinilai akurat dengan membuat menjadi tiga kelompok. Pada metode penelitian ini menggunakan tahapan KDD (Knowledge Discovery Database). KDD merupakan sebuah Langkah praktis dalam melakukan tahapan data mining untuk mencari, mengidentifikasi pola pada data, dimana pola yang ditemukan bersifat sah dan bermanfaat dan mudah dimengerti. Dalam tahapan KDD terdapat beberapa tahapan diantaranya pembersihan data, integrasi data, pemilihan data, seleksi data, transformasi data. Berikut adalah gambar mengenai tahapan – tahapan dalam KDD (Knowledge Discovery Database):



Gambar 1. Bagan metode penellitian

3.2. Sumber Data

Data sekunder dari puskesmas sukamulya digunakan dalam penelitian ini. Data tersebut berasal dari rekam medis pasien yang berisi informasi tentang gejala awal dan diagnosis penyakit lambung. Pemilihan sumber data ini didasarkan pada relevansi dan ketersediaan data yang diperlukan untuk analisis,

yang dilakukan menggunakan algoritma K-means clustering. Pengumpulan data dengan survey atau observasi peneliti mendapatkan data sekunder dengan izin dari pihak puskesmas untuk melihat data yeng memiliki Riwayat penyakit lambung.

3.3. Populasi

Penelitian ini melibatkan semua pasien puskesmas desa sukamulya dengan Riwayat kunjungan terkait keluhan lambung. Populasi ini terdiri dari orang-orang dari berbagai kelompok dan jenis kelamin yang memiliki gejala atau diagnosis penyakit lambung dalam sistem rekam medis puskesmas. Karakteristik populasi meliputi:

- Pasien terdaftar di puskesmas sukamulya
- Memiliki keluhan atau gejala yang berkaitan dengan penyakit lambung
- Usia bervariasi dari anak-anak hingga lansia
- Mencangkup kedua jenis kelamin (laki-laki dan Perempuan).

3.4. Teknik Pengumpulan Data

Pada teknik pengumpulan data menggunakan observasi sebagai pengumpulan data yaitu mengumpulkan data dengan observasi melalui unit bisnis center smk wahidin sebagai pemilik data. Adapun untuk observasi pengumpulan data dilakukan pada 1 oktober 2024, data tersebut dikumpulkan dan dikaji kecocokannya untuk penelitian. Kemudian setelah dikaji menentukan analisis pada data.

3.5. Teknik Analisis Data

Setelah data didapatkan tahapan selanjutnya proses pengolahan dan Teknik Analisa data. Pada tahap ini penulis mengelola data menggunakan Teknik analisis *K-means Clustering*. *K-means Clustering* adalah Teknik untuk mengelompokkan data berdasarkan kemiripan dari suatu data. Pada tahap ini penulis melakukan 9 kali percobaan dalam pencarian jumlah K yang optimal berdasarkan nilai DBI. Adapun rumus dalam pencarian DBI sebagai berikut:

$$DBI = \frac{1}{K}\sum_{i=1}^{K} \max_{j \neq i} (R_{i,j})$$

Gambar 2. Rumus pencarian DBI

Rumus ini menjelaskan:

a. K adalah jumlah *cluster* yang dihasilkan oleh *K-means*
b. R1,j adalah rata rata jarak antara semua pasangan titik data dalam cluster i ke pusat cluster j

4. HASIL DAN PEMBAHASAN

4.1. Data Selection

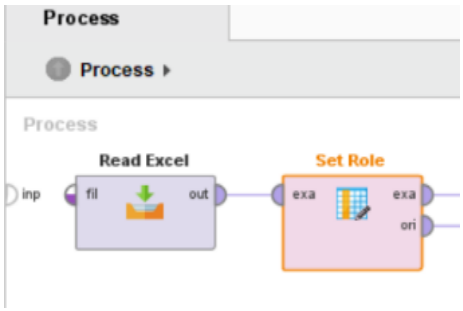
Tahapan pertama dalam proses KDD yaitu data selection. Pada tahapan ini penlusi melakukan seleksi data yang akan digunakan dalam proses pengelompokkan. Data yang akan diproses yaitu data penyakit lambung dari bulan September sampai November 2024. Tahapan seleksi data bertujuan untuk memilih data yang akan digunakan dalam proses pengelompokkan melalui aplikasi rapidminer, tahapan seleksi data ini dilakukan di Microsoft excel. Data yang digunakan dalam penelitian ini yaitu data pasien dengan Riwayat penyakit lambung dengan jumlah 309 data dan 8 atribut diantaranya USIA, JENIS KELAMIN, NYERI ULU HATI, MUNTAH, DIARE, PANAS DALAM, dan DIAGNOSIS AWAL.

4.2. Preprocessing

Setelah melakukan tahapan data selection maka tahapan selanjutnya adalah melakukan preprocessing. Tujuan dari preprocessing yaitu untuk penghapusan atribut data yang tidak diperlukan, data yang null atau missing, mengilangkan data yang tidak konsisten, serta menambahkan ID kedalam dataset yang akan digunakan. Pada tahapan preprocessing ini penulis menggunakan 2 jenis operator yaitu operator set role dab operator select attributes. Adapun Langkah-langkah yang dilakukan penulis dalam menggunakan 2 operator tersebut yaitu sebagai berikut

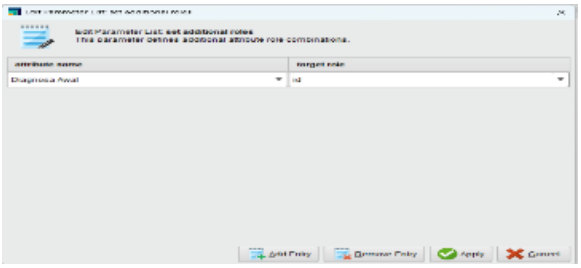
a. Operator set role

Langkah yang pertama pada tahapan preprocessing adalah menggunakan operator set role. Operator set role digunakan untuk mengubah peran dari satu ataau lebih atribut. Tampilan operator set role dapat dilihat pada gambar 3.



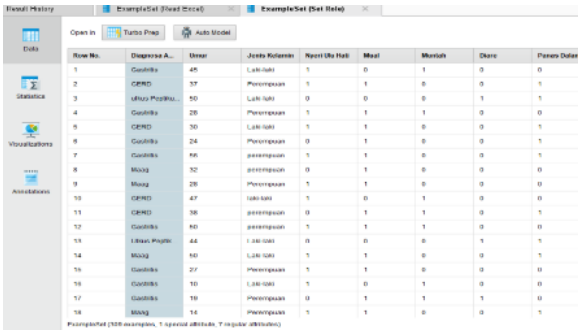
Gambar 3 operator set role

Pada operator set role terdapat parameter yang harus disesuaikan terlebih dahulu. Tampilan proses menggunkaan parameter set role untuk mengubah atribut menjadi id dapat dilihat pada gambar 4.



Gambar 4. Parameter Set Role

Atribut data yang menjadi target role id yaitu diagnose awal. Hasil yang diperoleh dari proses operator ini dapat dilihat pada gambar 5



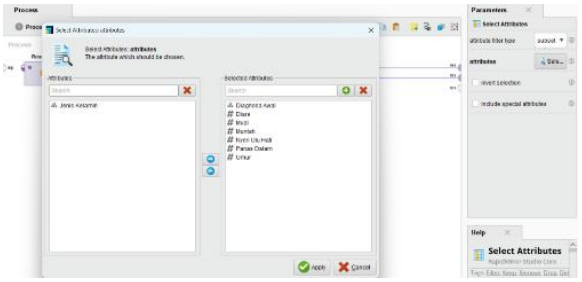
Gambar 5. Exampleset

b. Operator select attributes

Langkah kedua yang dilakukan penulis dalam tahapan preprocessing yaitu menggunakan operator select attributes. Operator select attributes adalah operator yang berfungsi untuk memilih subset atribut yang akan digunakan dari exampleset dan menghapus atribut lainnya yang tidak digunakan. Tampilan operator select attributes dapat dilihat pada gambar 6.

Gambar 6. Select Atribut

Pada operator select atribut terdapat parameter yang harus disesuaikan terlebih dahulu. Tampilan proses menggunakan parameter select atribut untuk menghilangkan atribut yang tidak digunkaan dapat dilihat pada gambar 7



Gambar 7. Parameter Select Atribut

Atribut yang tidak gunakan dalam penelitian ini adalah atribut data jenis kelamin, maka setelah dilakukan proses select attributes maka dataset yang semula berjumlah 8 atribut menjadi 7 atribut yang akan digunakan yaitu Diagnosa awal, usia, nyeri ulu hati, mual, muntah, diare, dan panas dalam.

4.3. Transformation

Pada tahapan transformation, yang biasanya dilakukan adalah mengubah tipe atribut data non-numerik menjadi numerik agar sesuai dengan kebutuhan algoritma K-means. Namun, dalam penelitian ini, seluruh atribut data yang akan digunakan dalam proses clustering sudah dalam bentuk numerik sejak tahap preprocessing. Oleh karena itu, tidak diperlukan transformasi menggunakan operator nominal to numerical. Hal ini menguntungkan karena:

a. Mempercepat proses pengolahan data karena menghilangkan satu tahap transformasi
b. Mengurangi resiko perubahan informasi yang mungkin terjadi dalam proses transformasi
c. Data sudah siap digunakan untuk proses clustering dengan algoritma K-means.

4.4. Data Mining

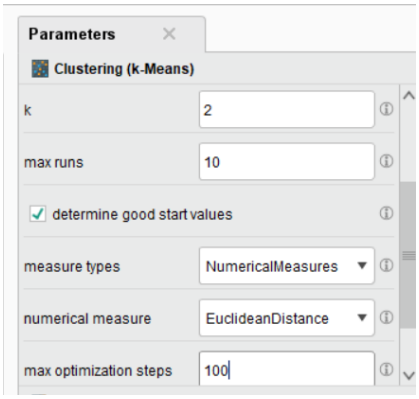
Setelah melakukan tahapan transformasi maka tahapan selanjutnya adalah melakukan penerapan pada data mining. Pada tahapan penerapan data mining penulis menggunakan tools rapidminer. Dengan metode algoritma K-means. Pada tahapan data mining ini penulis membagi menjadi dua tahapan yaitu tahapan pertama penulis melakukan proses clustering menggunakan K-means dengan operator yang digunakan yaitu operator clustering (k-means) dan tahapan kedua penulis melakukan pengujian dan mengevaluasi hasil clustering menggunakan operator cluster distance performance dengan metode yang digunakan evaluasi David bouldin index (DBI).

a. Tahapan Clustering K-means

Tahapan pertama yaitu penulis melakukan proses clustering K-means dilakukan 9 kali percobaan dimulai dari K-2 sampai K-10. Pemodelan data mining pada pengelompokkan data pengiriman paket menggunakan algoritma k-means dapat dilohat pada gambar 8.

Gambar 8 Model penerapan K-means

Pada proses clustering terdapat parameter yang harus disesuaikan terlebih dahulu. Tampilan parameter clustering dapat dilihat pada gambar 9



Gambar 9. Parameter K-means

Untuk penjelasan lebih jelas terhadap parameter yang digunakan yaitu sebagai berikut:

1. K adalah jumlah cluster yang digunakan. Jumlah cluster yang digunakan oleh penulis yaitu dari (2-10 cluster).
2. Max rumus adalah jumlah maxsimum run K-means (default). Jumlah yang digunakan penulis yaitu 10
3. Measure types adalah jenis pengukuran numeric. Jenis pengukuran numeric yang digunakan penulis yaitu Numerical Measures.
4. Numerical Measure adalah perhitungan hjarak dari 2 buah titik. Perhitungan jarak yang digunakan penulis yaitu Euclidean Distance.
5. Max optimization steps adalah jumlah maksimal iterasi (default jumlah maksimal iterasi yang digunakan penulis yaitu 100.

b. Evaluasi David Bouldin Index (DBI)

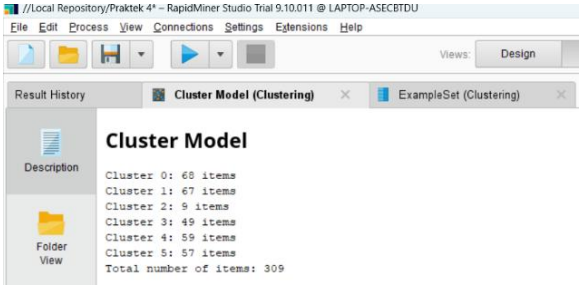
Tahapan yang kedua yaitu melakukan pengujian hasil clustering dengan menerapkan nilai DBI pada parameter Cluster Distance Performance. Pada pengujian DBI, cluster yang memiliki nilai DBI terkecil atau mendekati 0 dijadikan sebagai cluster yang terbaik. Untuk mencari cluster terbaik penulis melakukan percobaan dari cluster 2 sampai dengan 10 rekapitulasi. Jumlah K cluster yang dihasilkan dari nilai Davies Bouldin Index dapat dilihat dari tabel 1 berikut:

Tabel 1 Hasil Clustering K2-10

Clustering	Jumlah anggota cluster	DBI
2	Cluster0:157 itms	
	Cluster1:152 itms	
3	Cluster0:58 itms	
	Cluster1:133 itms	
	Cluster2:118 itms	
4	Cluster0:65 itms	
	Cluster1:58 itms	
	Cluster2:94 itms	
	Cluster3:92 itms	
5	Cluster0:66 itms	
	Cluster1:63 itms	
	Cluster2:35 itms	
	Cluster3:80 itms	
	Cluster4:65 itms	
6	Cluster0:68 itms	
	Cluster1:67 itms	
	Cluster2:9 itms	
	Cluster3:49 itms	
	Cluster4:59 itms	
	Cluster5:57 itms	
7	Cluster0:51 itms	
	Cluster1:59 itms	
	Cluster2:47 itms	
	Cluster3:45 itms	
	Cluster4:9 itms	
	Cluster5:57 itms	
	Cluster6:41 itms	
8	Cluster0:44 itms	
	Cluster1:59 itms	
	Cluster2:30 itms	
	Cluster3:47 itms	
	Cluster4:54 itms	
	Cluster5:9 itms	
	Cluster6:34 itms	
	Cluster7:32 itms	
9	Cluster0:42 itms	
	Clluster1:34 itms	
	Cluster2:59 itms	
	Cluster3:42 itms	
	Cluster4:7 itms	
	Cluster5:28 itms	
	Cluster6:33 itms	
	Cluster7:41 itms	
	Cluster8:23 itms	
9	Cluster0:30 itms	
	Cluster1:26 itms	
	Cluster2:52 itms	
	Cluster3:11 itms	
	Cluster4:50 itms	
	Cluster5:42 itms	
	Cluster6:23 itms	
	Cluster7:9 itms	
	Cluster8:35 itms	
	Cluster9:31 itms	

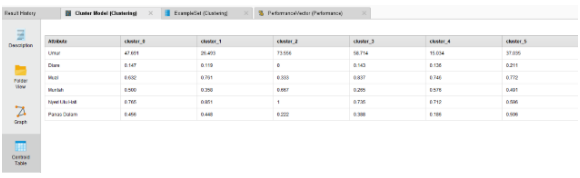
4.5. Evaluation

Pada tahapan evaluation merupakan tahapan dalam menghasilkan pola – pola guna memperkirakan tujuan yang diharapkan. Berdasarkan tabel 4.2 menunjukkan bahwa cluster data penyakit lambung di puskesmas desa sukamulya menggunakan perhitungan davies Bouldin index nilai yang paling mendekati angka 0 dengan percobaan cluster 2 sampai cluster 10 menghasilkan nilai K terbaik pada cluster 6 yaitu 0.488 dengan jumlah anggota cluster 0: 68 items, cluster 1: 67 items, cluster 2: 9 items, Cluster 3: 49 items, Cluster 4: 59 items, Cluster 5: 57 items.

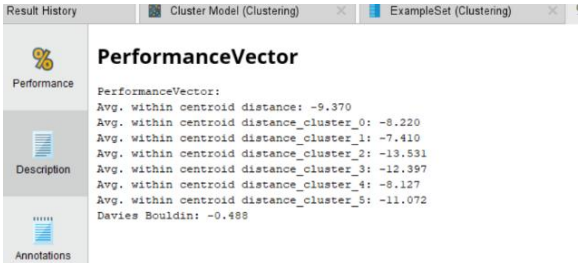


Gambar 10. cluster model

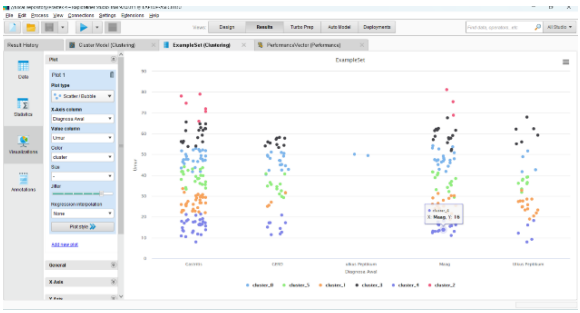
Berikut merupakan jarak antar cendroid cluster dari data penyakit lambung di puskesmas Desa sukamulya menggunakan algoritma K-means bisa dilihat pada gambar 11, hasil dari deskripsi performanceVector dapat dilihat pada gambar 12, dan plot grafik dapat dilihat pada gambar 13.



Gambar 11. Jarak Centroid



Gambar 12. Performance Vector



Gambar 13. Plot Grafik

4.6. Hasil clustering

Hasil clustering dengan nilai DBI terbaik pada cluster yaitu, terdapat 6 cluster dengan nilai DBI -0.488, nilai paling rendah dianggap paling baik, semakin rendah nilai semakin baik untuk pembagian cluster. Cluster 0: 68 items dengan diagnosis gastritis sebanyak 26 anggota, GERD 16 anggota, Maag 20 anggota dan ulkus peptikum 6 anggota, dan memiliki usia rata-rata 40-50 tahun. Cluster 1: 67 items, dengan diagnosis gastritis 32 anggota, GERD 4 anggota, maag 16 anggota, dan ulkus peptikum 15 anggota, dan memiliki usia rata-rata 20-30 tahun. Cluster 2:9 items, dengan diagnosis gastritis 6 anggota dan maag 3 anggota, dan memiliki usia rata-rata 70-80 tahun. Cluster 3: 49 items, dengan gastritis 17 anggota, GERD 10 anggota, maag 16 anggota dan ulkus peptikum 6 anggota, dan memiliki usia rata-rata 50-70 tahun. Cluster 4: 59 items, dengan diagnosis gastritis 30 anggota, GERD 9 anggota, maag 15 anggota, dan ulkus peptikum 5 anggota, dan memiliki usia rata-rata 10-20 tahun. Cluster 5: 57 items, dengan diagnosis gastritis 22 anggota, GERD 13 anggota, maag 13 anggota dan ulkus peptikum 9 anggota, dan memiliki usia rata-rata 30-40 tahun. total ada 309 items.

4.7. Hasil Evaluasi DBI

Evaluasi setelah dilakukan proses clustering dengan menggunakan cluster 2 sampai 10 dapat dilihat pada nilai DBI pada tabel berikut:

Tabel 2. Hasil DBI

K	DBI
2	-0.520
3	-0.554
4	-0.521
5	-0.544
6	-0.488
7	-0.530
8	-0.545
9	-0.571
10	-0.557

Dari tabel diatas hasil DBI dapat dilihat bahwa K=0 dibagi 6 cluster dengan Nilai DBI 0.488, nilai DBI paling rendah dianggap baik, semakin rendah nilai semakin baik untuk pembagian cluster.

5. KESIMPULAN DAN SARAN

5.1. Kesimpulan

Hasil pengelompokkan data diagnosis penyakit lambung di puskesmas desa sukamulya menggunakan algoritma K-means menghasilkan 6 cluster, yaitu cluster0: 68 anggota dengan memiliki usia rata-rata 40-50 tahun, cluster 1: 67 anggota dengan memiliki usia rata-rata 20-30 tahun, cluster 2: 9 anggota dengan memiliki usia rata-rata 70-80 tahun, cluster 4: 59 anggota dengan memiliki usia rata-rata 10-20 tahun, dan cluster 5: 57 anggota dengan memiliki usia rata-rata 30-40 tahun. Dari hasil analisis, ditemukan bahwa Gastritis menjadi frekuensi paling tinggi sebanyak 133 kasus, penyakit maag menempati posisi kedua sebanyak 83 kasus, sementara GERD dan ulkus peptikum mencatat frekuensi terendah dengan masing-masing 52 dan 41 kasus.Nilai K yang mendekati Optimum adalah k=6 dengan hasil nilai DBI -0.488, nilai DBI paling rendah dianggap paling baik. semakin rendah nilai semakin baik untuk pembagian cluster.

5.2. Saran

Untuk pengembangan penelitian selanjutnya, disarankan untuk memperkaya dataset dengan menambah jumlah sampel dan melengkapi variabel-variabel yang relevan. Hal ini akan meningkatkan akurasi analisis clustering dan memberikan gambaran yang lebih komprehensif tentang pola penyakit lambung di masyarakat. Penggunaan aplikasi analisis yang lebih modern dan canggih juga direkomendasikan untuk meningkatkan presisi hasil clustering dan mempermudah proses analisis data.

DAFTAR PUSTAKA

[1] Huerta, E. A., Khan, A., Davis, E., Bushell, C., Gropp, W. D., Katz, D. S., Kindratenko, V., Koric, S., Kramer, W. T. C., McGinty, B., McHenry, K., & Saxton, A. (2020). Convergence of artificial intelligence and high performance computing on NSF-supported cyberinfrastructure. *Journal of Big Data*, *7*(1), 1–12. https://doi.org/10.1186/S40537-020-00361-2/FIGURES/5

[2] Jasinska-Piadlo, A., Bond, R., Biglarbeigi, P., Brisk, R., Campbell, P., Browne, F., & McEneaney, D. (2023). Data-driven versus a domain-led approach to k-means clustering on an open heart failure dataset. *International Journal of Data Science and Analytics*, *15*(1), 49–66. https://doi.org/10.1007/S41060-022-00346-

9/TABLES/8

[3] Jebarani, P. E., Umadevi, N., Dang, H., & Pomplun, M. (2021). A Novel Hybrid K-Means and GMM Machine Learning Model for Breast Cancer Detection. *IEEE Access*, *9*, 146153–146162. https://doi.org/10.1109/ACCESS.2021.3123425

[4] Kumar, D. M., Satyanarayana, D., & Prasad, M. N. G. (2021). An improved Gabor wavelet transform and rough K-means clustering algorithm for MRI brain tumor image segmentation. *Multimedia Tools and Applications*, *80*(5), 6939–6957. https://doi.org/10.1007/S11042-020-09635-6/METRICS

[5] Kushwaha, P. K., & Kumaresan, M. (2021). Machine learning algorithm in healthcare system: A Review. *Proceedings of International Conference on Technological Advancements and Innovations, ICTAI 2021*, 478–481. https://doi.org/10.1109/ICTAI53825.2021.9673220

[6] Lee, S. A., Cho, H. C., & Cho, H. C. (2021). A Novel Approach for Increased Convolutional Neural Network Performance in Gastric-Cancer Classification Using Endoscopic Images. *IEEE Access*, *9*, 51847–51854. https://doi.org/10.1109/ACCESS.2021.3069747

[7] Maity, A., Bhattacharya, S., Mahato, A. C., Chaudhuri, S., & Pradhan, M. (2023). A pattern-recognition-based clustering method for non-invasive diagnosis and classification of various gastric conditions. *Https://Doi.Org/10.1177/14690667231174350*, *29*(3), 192–199. https://doi.org/10.1177/14690667231174350

[8] Mollura, D. J., Culp, M. P., Pollack, E., Battino, G., Scheel, J. R., Mango, V. L., Elahi, A., Schweitzer, A., & Dako, F. (2020). Artificial intelligence in low- And middle-income countries: Innovating global health radiology. *Radiology*, *297*(3), 513–520. https://doi.org/10.1148/RADIOL.2020201434/ASSET/IMAGES/LARGE/RADIOL.2020201434.FIG5.JPEG

[9] Qiao, D., Liu, H., Zhang, X., Lei, L., Sun, N., Yang, C., Li, G., Guo, M., Zhang, Y., Zhang, K., & Liu, Z. (2022). Exploring the potential of thyroid hormones to predict clinical improvements in depressive patients: A machine learning analysis of the real-world based study. *Journal of Affective Disorders*, *299*, 159–165. https://doi.org/10.1016/J.JAD.2021.11.055

[10] Si, Y. T., Xiong, X. S., Wang, J. T., Yuan, Q., Li, Y. T., Tang, J. W., Li, Y. N., Zhang, X. Y., Li, Z. K., Lai, J. X., Umar, Z., Yang, W. X., Li, F., Wang, L., & Gu, B. (2024). Identification of chronic non-atrophic gastritis and intestinal metaplasia stages in the Correa's cascade through machine learning analyses of SERS spectral signature of non-invasively-collected human gastric fluid samples. *Biosensors and Bioelectronics*, *262*, 116530. https://doi.org/10.1016/J.BIOS.2024.116530