

PENERAPAN DATA MINING DALAM PENGELOMPOKAN PENYAKIT UNTUK REKOMENDASI MATERI PENYULUHAN KESEHATAN DI KLINIK KELUARGA CIANJUR

Mochamad Nurkhalid Kadafi¹, Alif Finandhita²

^{1,2}Teknik Informatika, Universitas Komputer Indonesia
Jl. Dipati Ukur No. 112-116 Coblong, Kota Bandung, Jawa Barat 40132
E-mail : alif.finandhita@email.unikom.ac.id²

Abstrak

Tujuan dari penelitian ini adalah memberikan rekomendasi materi penyuluhan untuk membantu Koordinator Promkes dan Marketing di Klinik Keluarga dalam menentukan materi penyuluhan kesehatan yang tepat di suatu wilayah. Penyuluhan dilakukan secara terjadwal ke wilayah yang sudah ditentukan. Namun demikian banyak materi yang disampaikan tidak sesuai dengan mayoritas penyakit yang diderita oleh masyarakat di wilayah tersebut. Hal tersebut menyebabkan pelayanan klinik menjadi tidak optimal dikarenakan kurangnya kewaspadaan terhadap suatu penyakit yang ada di wilayahnya. *Data Mining* dengan metode *Clustering* dan algoritma *K-Means* digunakan pada penelitian ini untuk mengetahui pola penyebaran penyakit berdasarkan karakteristiknya dari sumber data yang digunakan, terdiri dari data kunjungan, data induk penyakit, data kategori penyakit, data kecamatan, dan data desa. Hasil evaluasi menunjukkan proses pengelompokan penyakit menghasilkan rata-rata akurasi sebesar 0,263. Hasil tersebut menunjukkan bahwa *data mining* dengan metode *clustering* dapat membantu Koordinator Promkes dan Marketing di Klinik Keluarga dalam menentukan materi penyuluhan yang tepat untuk masyarakat.

Kata kunci : Data Mining, Clustering, K-Means, Pengelompokan Penyakit

Abstract

The purpose of this study is to provide recommendations for counseling materials to assist the Promkes and Marketing Coordinator at the Klinik Keluarga in determining appropriate health counseling materials in an area. Counseling is conducted on a scheduled basis to a predetermined area. However, many of the materials presented are not in accordance with the majority of diseases suffered by the community in the area. This causes clinic services to be not optimal due to lack of awareness of a disease in the region. Data Mining with Clustering method and K-Means algorithm is used in this study to determine the pattern of disease distribution based on its characteristics from the data sources used, consisting of visit data, parent disease data, disease category data, sub-district data, and village data. The evaluation results show that the disease clustering process produces an average accuracy of 0.263. These results show that data mining with the clustering method can help the Promkes and Marketing Coordinator at the Family Clinic in determining the right counseling material for the community.

Keywords : Data Mining, Clustering, K-Means, Disease Clustering

1. PENDAHULUAN

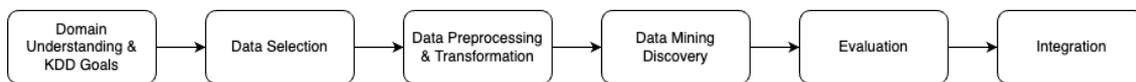
Berdasarkan Riset Kesehatan Dasar (Riskesdas) tahun 2018 terdapat beberapa penyakit dengan tingkat prevalensi yang cukup tinggi di Indonesia khususnya di provinsi Jawa Barat. Beberapa penyakit tersebut adalah *stroke* sebesar 11,44%, hipertensi sebesar 9,67%, ispa sebesar 4,68%, *pneumonia* sebesar 2,57% dan penyakit jantung 1,62% [1]. Klinik Keluarga di bawah naungan PT. Abdi Yakin Anugerah merupakan salah satu penyedia layanan kesehatan yang berlokasi di Kabupaten Cianjur, Provinsi Jawa Barat [2]. Mereka menyediakan pelayanan medis, meliputi pemeriksaan kesehatan, pengobatan, dan penyuluhan kepada pasien dan masyarakat, dengan tujuan mencakup pencegahan penyakit serta peningkatan pengetahuan dan kesadaran dalam bidang kesehatan. Klinik Keluarga juga berperan dalam mengelola rekam medis pasien yang terkena penyakit serta menyelenggarakan kegiatan materi penyuluhan sebagai bagian dari penanggulangan penyakit.

Salah satu tugas yang menjadi tanggung jawab Klinik Keluarga adalah melakukan penyuluhan kesehatan ke berbagai wilayah yang ada di Kabupaten Cianjur secara rutin setiap bulan. Hal tersebut diketahui dari hasil wawancara dengan Koordinator Promkes & *Marketing* Klinik Keluarga. Proses penyampaian materi penyuluhan yang dilakukan dilakukan secara terjadwal, bergantian dengan wilayah yang lain yang sudah ditentukan sebelumnya. Materi yang diberikan sering kali tidak sesuai dengan penyakit yang banyak diderita oleh masyarakat di wilayah yang diberikan penyuluhan, dikarenakan tidak ada pemetaan penyakit yang diderita oleh masyarakat pada kelompok tertentu di suatu wilayah, sehingga wabah penyakit tidak dapat ditangani secara tepat dan optimal oleh klinik. Di sisi lain data rekam medis yang ada saat ini belum dimanfaatkan dengan baik untuk dapat membantu pihak klinik memberikan penyuluhan ke suatu wilayah.

Data mining merupakan suatu metode untuk pengolahan data yang digunakan untuk menemukan pola-pola atau model baru yang sempurna dan dapat dipahami dalam suatu *database* yang besar [3]. *Data mining* adalah proses penambangan data dengan jumlah besar untuk memperoleh pengetahuan baru dengan menyatukan teknik statistik, kecerdasan buatan, *machine learning* dan visualisasi informasi [4]. *Clustering* adalah salah satu metode *data mining* untuk menetapkan data ke kelompok tertentu berdasarkan kemiripannya dengan data lain, dan dapat digunakan untuk menganalisa data di bidang kesehatan [5]. *K-Means* merupakan salah satu algoritma *clustering* dengan performa baik yang dapat digunakan untuk proses pengelompokan data [6][7]. Maka dari itu teknik *data mining* dengan metode *clustering* dan algoritma *K-Means* digunakan pada penelitian ini untuk mendapatkan pengetahuan dari sekumpulan data, termasuk data rekam medis di dalamnya, berupa kelompok pola data penyakit di suatu wilayah. Beberapa penelitian terkait dengan penggunaan teknik *data mining* dengan metode *clustering* di layanan kesehatan sudah pernah dilakukan, namun belum ada yang spesifik menyatakan terkait dengan solusi akhir berupa rekomendasi materi penyuluhan yang disesuaikan dengan hasil pengelompokan penyakit [8][9][10].

2. METODOLOGI

Metodologi penelitian yang digunakan dalam penelitian ini adalah *Knowledge Discovery in Databases* (KDD) [11] yang kemudian disesuaikan berdasarkan penelitian yang dilakukan sehingga menghasilkan sebagaimana yang dijelaskan pada Gambar 1:



Gambar 1. Alur Proses Metodologi Penelitian

Berikut adalah penjelasan dari tahapan-tahapan yang ada dalam metodologi yang digunakan:

- 1) *Domain Understanding & KDD Goals* merupakan tahap awal untuk memahami domain kasus yang diambil dan memecahkan permasalahan yang ada pada Klinik Keluarga. Menentukan secara jelas tujuan dan fungsionalitas pengetahuan yang dapat diberikan kepada Koordinator Promkes dan Marketing, berupa informasi strategis mengenai pengelompokan penyakit sehingga materi penyuluhan bisa secara tepat diberikan kepada masyarakat.
- 2) *Data Selection* merupakan tahap kedua untuk memilih data yang sudah tersedia yang meliputi data kunjungan, data induk penyakit, data kategori penyakit, data kecamatan, dan data desa, lalu mengintegrasikan semua data tersebut menjadi satu *dataset*, sehingga dapat menghasilkan kumpulan data yang berisi karakteristik untuk pengelompokan penyakit.
- 3) *Data Preprocessing & Transformation* merupakan tahap untuk melakukan pembersihan data kosong dan data duplikat dari *dataset* yang sudah dipilih sebelumnya, serta memilih atribut yang akan digunakan untuk proses transformasi. Proses transformasi yang dilakukan adalah mengubah bentuk data dari satu bentuk ke bentuk yang lain yaitu dengan menggunakan *encoding* dan agregasi / *summary* agar siap untuk digunakan dan sesuai untuk dimasukkan pada proses pemodelan yang akan dilakukan pada tahap selanjutnya [4][11]
- 4) *Data Mining Discovery* merupakan tahap penerapan algoritma *K-Means* untuk menemukan pengetahuan dengan melakukan pengelompokan penyakit berdasarkan *dataset* yang sudah ditransformasi pada tahapan sebelumnya dengan membuat model pengelompokan (*clustering*),

sehingga diperoleh pengetahuan berupa kelompok penyakit di suatu wilayah yang dapat dijadikan sebagai bahan acuan rekomendasi materi penyuluhan .

- 5) *Evaluation* merupakan tahapan pengujian terhadap proses data mining dengan metode *clustering* yang sudah dilakukan di tahapan sebelumnya untuk mengetahui apakah proses yang dilakukan dapat menghasilkan kelompok penyakit dan menghasilkan rekomendasi materi penyuluhan atau tidak maka dapat dilakukan evaluasi terhadap model yang telah dibuat.
- 6) *Integration* merupakan tahap terakhir yaitu pembuatan sistem secara keseluruhan, kemudian mengintegrasikan model *clustering* sehingga dapat menghasilkan rekomendasi materi penyuluhan untuk memudahkan koordinator Promkes & *marketing* di Klinik Keluarga dalam mengambil keputusan.

3. HASIL DAN PEMBAHASAN

Hasil dan pembahasan berdasarkan setiap tahapan metodologi penelitian, dapat dilihat pada pembahasan di bawah ini.

3.1 Analisis Pemahaman Tujuan

Penelitian ini bertujuan untuk membantu pihak Koordinator Promkes & *Marketing* di Klinik Keluarga dalam menentukan materi penyuluhan yang sesuai dengan keadaan masyarakat dan diharapkan dapat membantu mencegah terjadinya penyakit di suatu wilayah yang menjadi area kerjanya. Hasil akhir dari penelitian yang dilakukan adalah terciptanya sebuah sistem yang dibuat dengan menggunakan teknik data mining *clustering*, dapat mengelompokkan penyakit dan memberikan rekomendasi terkait dengan materi penyuluhan. Dengan demikian kegiatan penyuluhan yang dilakukan dapat tepat sasaran dan membantu dalam proses pencegahan penyakit di suatu wilayah.

3.2 Analisis Pengumpulan Data

Tahapan ini merupakan kegiatan pengumpulan data untuk memilih dan membuat *dataset*, yaitu memperoleh data yang sudah tersedia lalu meng-integrasikan data tersebut. Data yang digunakan dalam penelitian ini adalah data kunjungan, data kategori penyakit, data induk penyakit, data kecamatan dan data desa sebanyak 5385 baris *record* yang bersumber dari *database* hasil *join* dengan periode data Januari tahun 2022. Proses analisis data ini dilakukan dengan cara menggunakan data file Excel yang memiliki format *.xlsx yang di ekspor melalui *database*. Berikut adalah hasil integrasi data yang digunakan yang dapat dilihat pada Tabel 1.

Tabel 1. Data Hasil Integrasi

No	id	id_pendaftaran	tanggal_kunjungan	...	id_kota	nama_kecamatan	...	nama_desa	...	deleted
1	341544	51733	01/01/22	...	3203	Pacet Cianjur	...	Cipendawa	...	0000-00-00:00:00
2	341545	66947	01/01/22	...	3203	Sukaresmi	...	Sukaresmi	...	0000-00-00:00:00
3	341546	2475	01/01/22	...	3203	Pacet Cianjur	...	Ciputri	...	0000-00-00:00:00
4	341548	15062	01/01/22	...	3203	Pacet Cianjur	...	Cipendawa	...	0000-00-00:00:00
5	341549	48194	01/01/22	...	3203	Cipanas	...	Ciloto	...	0000-00-00:00:00
6	341550	4703	01/01/22	...	3203	Pacet Cianjur	...	Cipendawa	...	0000-00-00:00:00
7	341551	73572	01/01/22	...	3203	Mande	...	Kutawaringin	...	0000-00-00:00:00
8	341552	51070	01/01/22	...	3203	Pacet Cianjur	...	Cipendawa	...	0000-00-00:00:00
9	341553	96286	01/01/22	...	3203	Cipanas	...	Cipanas	...	0000-00-00:00:00
10	341554	47824	01/01/22	...	3203	Cipanas	...	Cipanas	...	0000-00-00:00:00
...
5385	348779	35411	31/01/22	...	3203	Cipanas	...	Gadog	...	0000-00-00:00:00

3.3 Analisis Persiapan Data

Pada tahap ini dilakukan analisis persiapan data yang bertujuan untuk membuat data hasil integrasi menjadi lebih siap untuk digunakan dalam proses pemodelan data dengan cara melakukan *data preprocessing & transformation*.

3.3.1 Data Preprocessing

Proses *cleaning data* dilakukan pada tahap ini. Proses yang dilakukan adalah menghapus data yang bernilai *NULL* (data yang tidak memiliki nilai) dan menghapus data duplikat agar data yang digunakan dapat akurat [12].

3.3.2 Data Transformation

Pada tahap ini dilakukan proses *encoding* atribut *nama_penyakit*, *nama_kecamatan* dan *nama_desa* menjadi nilai numerik, kemudian dilakukan proses *summary* atau agregasi data dengan menambahkan atribut baru yaitu atribut *jumlah_pasien* untuk kebutuhan dampam proses *data mining* [13]. Hasil transformasi dapat dilihat pada Tabel 2.

Tabel 2. Hasil *Data Preprocessing & Transformation*

No	nama_penyakit	kode	nama_kecamatan	nama_desa	penyakit_encode	kecamatan_encode	desa_encode	jumlah_pasien
1	Dyspepsia	K30	Pacet Cianjur	Cipendawa	167	13	21	61
2	Acute tonsillitis	J03	Sukaesmi	Sukaesmi	43	16	65	2
3	Dyspepsia	K30	Pacet Cianjur	Ciputri	167	13	22	17
4	Acute nasopharyngitis [common cold]	J00	Pacet Cianjur	Cipendawa	35	13	21	75
5	Chronic obstructive pulmonary disease with acute exacerbatio	J44.1	Cipanas	Ciloto	119	6	17	2
...				
176 9	Bitten or stung by nonvenomous insect and other nonvenomo	W57	Sukaesmi	Cibadak	88	16	7	1

3.4 Analisis Pemodelan Data

Tahapan analisis pemodelan data dilakukan dengan menerapkan metode *clustering* untuk menghasilkan kelompok penyakit berdasarkan atribut terpilih yang telah melewati *data preprocessing & transformation*. Dalam *machine learning*, *clustering* adalah salah satu contoh dari *unsupervised learning* [14]. Pengelompokan penyakit dilakukan dengan menggunakan algoritma *K-Means*. *K-Means* adalah metode *clustering* yang termasuk dalam pendekatan *partitioning*, di mana data set dipartisi untuk mengelompokkan *cluster* dengan karakteristik serupa ke dalam satu kelompok, sementara yang berbeda dikelompokkan secara terpisah. Algoritma *K-Means* adalah model *centroid* yang menggunakan titik tengah (*centroid*) untuk membentuk *cluster*. Obyek data akan masuk ke dalam *cluster* yang memiliki jarak terpendek dengan *centroid*. Tujuan dari algoritma *K-Means* adalah meminimalkan kesalahan ganda dengan nilai awal *centroid* diatur secara acak pada iterasi pertama [15][16]. Berikut adalah langkah-langkah yang dilakukan dalam algoritma *K-Means* :

3.4.1 Menentukan jumlah *k cluster* menggunakan *Elbow Method*

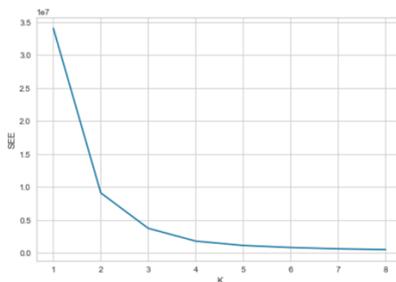
Pada tahap ini dilakukan perhitungan menggunakan *elbow method* dengan cara menghitung nilai dari *Sum of Square Error* (SSE). *Sum of Square Error* (SSE) merupakan rumus yang digunakan untuk mengukur perbedaan antara data yang diperoleh dengan model perkiraan yang telah dilakukan sebelumnya [17]. SSE sering digunakan sebagai acuan penelitian terkait dalam menentukan optimal *cluster* [18]. Berikut adalah rumus SSE yang digunakan:

$$SSE = \sum_{k=1}^K \sum_{X_i \in S_k} ||X_i - C_k||^2 \quad [13] \quad (1)$$

Setelah dilakukan proses perhitungan menggunakan *elbow method* dengan menghitung nilai SSE kemudian dipilih jumlah *k cluster* adalah *k=3*, berikut grafik *elbow* dan nilai SSE yang dihasilkan pada Tabel 3 dan Gambar 2.

Tabel 3. Hasil Perhitungan *Sum of Square Error*

Nilai K	<i>Sum of Square Error</i>
1	34721855,41
2	9101374,527
3	3723883,674
4	1798410,892
5	1788971,584
6	950408,9875
7	760033,606
8	591788,412



Gambar 2. Grafik *Elbow Method*

3.4.2 Memilih nilai awal centroid sebanyak k cluster

Tahap ini dilakukan untuk memilih nilai awal *centroid* secara acak dalam *dataset* sebanyak jumlah k *cluster* berdasarkan hasil *Elbow Method* dengan menghitung nilai SSE [19]. Nilai awal *centroid* dapat dilihat pada Tabel 4.

Tabel 4. Nilai Awal *Centroid*

centroid	penyakit_encode	kecamatan_encode	desa_encode	jumlah_pasien
1	475	16	16	1
2	43	16	65	2
3	44	8	68	1

3.4.3 Menghitung masing-masing jarak setiap data terhadap centroid

Pada tahap ini dilakukan proses perhitungan jarak terhadap setiap data awal *centroid* yang telah ditentukan menggunakan persamaan *Euclidean Distance* [20]. Persamaanya adalah sebagai berikut:

$$d(x, y) = \sqrt{\sum_{t=1}^n (x_t - y_t)^2} \tag{2}$$

Berikut adalah contoh perhitungan jarak data terhadap *centroid* menggunakan persamaan (2) pada iterasi pertama yang telah dilakukan:

Jarak data 1 ke centroid 1

$$d(1,1) = \sqrt{(167 - 475)^2 + (13 - 16)^2 + (21 - 16)^2 + (61 - 1)^2} = 313,8439102$$

Jarak data 1 ke centroid 2

$$d(1,2) = \sqrt{(167 - 43)^2 + (13 - 16)^2 + (21 - 65)^2 + (61 - 2)^2} = 144,2289846$$

Jarak data 1 ke centroid 3

$$d(1,3) = \sqrt{(167 - 44)^2 + (13 - 8)^2 + (21 - 68)^2 + (61 - 1)^2} = 144,7860491$$

3.4.4 Menentukan kelompok setiap data berdasarkan jarak terpendek

Seperti contoh perhitungan data 1 di tahap sebelumnya pada iterasi pertama, data 1 masuk ke dalam kelompok 2 karena nilai jarak ke pusat *centroid* 3 merupakan jarak terpendek diantara jarak terhadap titik pusat *centroid* yang lainnya. Hal ini berlaku untuk seluruh data yang ada ke setiap pusat *centroid*.

3.4.5 Memperbaharui nilai centroid baru

Setelah dilakukan perhitungan jarak data ke setiap pusat *centroid*, selanjutnya adalah menghitung pusat *centroid* baru berdasarkan anggota setiap *cluster* yang sudah didapatkan dengan persamaan berikut:

$$V_{ij} = \frac{1}{N_i} \sum_{k=0}^{N_i} X_{kj} \tag{3}$$

Berikut adalah contoh perhitungan dengan persamaan (3) untuk memperbaharui nilai *centroid* baru yang akan digunakan pada iterasi selanjutnya.

$$\begin{aligned} &\text{Atribut penyakit_encode / Jumlah Cluster 1} \\ &= \frac{304+285+437+453+\dots+370}{668} = 365,8922156 \end{aligned}$$

$$\begin{aligned} &\text{Atribut kecamatan_encode / Jumlah Cluster 1} \\ &= \frac{13+8+13+8+\dots+6}{668} = 9,902694611 \end{aligned}$$

$$\begin{aligned} &\text{Atribut desa_encode / Jumlah Cluster 1} \\ &= \frac{21+4+13+43+\dots+18}{668} = 28,16766467 \end{aligned}$$

$$\begin{aligned} &\text{Atribut jumlah_pasien / Jumlah Cluster 1} \\ &= \frac{19+1+2+1+\dots+1}{668} = 2,142215569 \end{aligned}$$

Berikut adalah hasil perhitungan *centroid* baru menggunakan persamaan (3) pada Tabel 5.

Tabel 5. Centroid Baru

centroid	penyakit_encode	kecamatan_encode	desa_encode	jumlah_pasien
1	365,8922156	9,902694611	28,16766467	2,142215569
2	78,3036566	10,7217806	22,65500795	4,298887122
3	179,8644068	8,737288136	38,51059322	2,343220339

3.4.6 Menghitung jarak data terhadap centroid baru

Pada tahapan ini dilakukan proses mengulang perhitungan jarak data terhadap *centroid* baru yang terbentuk hingga tidak ada lagi perubahan posisi pada nilai *centroid*. Perhitungan jarak data terhadap *centroid* pada penelitian ini dihentikan hingga iterasi terakhir atau iterasi ke-12. Berikut adalah hasil pengelompokan penyakit dari iterasi terakhir pada Tabel 6.

Tabel 6. Hasil Pengelompokan Iterasi Terakhir

nama_penyakit	kode	nama_kecamatan	nama_desa	Kelompok
Dyspepsia	K30	Pacet Cianjur	Cipendawa	3
Acute tonsillitis	J03	Sukaresmi	Sukaresmi	2
Dyspepsia	K30	Pacet Cianjur	Ciputri	3
Acute nasopharyngitis [common cold]	J00	Pacet Cianjur	Cipendawa	2
Chronic obstructive pulmonary disease with acute exacerbatio	J44.1	Cipanas	Ciloto	2
...
Bitten or stung by nonvenomous insect and other nonvenomo	W57	Sukaresmi	Cibadak	2

Proses pengelompokan menghasilkan kelompok 1 sebanyak 422 anggota, kelompok 2 sebanyak 27 anggota dan kelompok 3 sebanyak 27 anggota. Kemudian berdasarkan hasil pengelompokan di tabel 6 dilakukan rekapitulasi hasil kelompok dan didapatkan karakteristik yang dapat dilihat pada Tabel 7.

Tabel 7. Karakteristik Hasil Pengelompokan

Kelompok	Karakteristik	Tingkat Prioritas
1	Jumlah penyakit yang muncul sebanyak 174 penyakit	Biasa
	Total pasien yang terjangkau sebanyak 1190 pasien	
	Wilayah yang sering muncul berasal dari kecamatan Pacet Cianjur di desa Ciherang sebanyak 57 anggota dan desa Cipendawa sebanyak 45 anggota, kemudian dari kecamatan Cipanas di desa Sindanglaya sebanyak 33 anggota dan desa Cipanas sebanyak 33 anggota	
2	Jumlah penyakit yang muncul sebanyak 135 penyakit	Tinggi
	Total pasien yang terjangkau sebanyak 2210 pasien	
	Wilayah yang sering muncul berasal dari kecamatan Pacet Cianjur di desa Ciherang sebanyak 56 anggota dan desa Cipendawa sebanyak 49 anggota, kemudian dari kecamatan Cipanas di desa Cipanas sebanyak 54 anggota dan desa Sindanglaya sebanyak 41 anggota	
3	Jumlah penyakit yang muncul sebanyak 166 penyakit	Sedang
	Total pasien yang terjangkau sebanyak 1841 pasien	
	Wilayah yang paling sering muncul berasal dari kecamatan Pacet Cianjur di desa Ciherang sebanyak 64 anggota dan desa Cipendawa sebanyak 55 anggota, kemudian dari kecamatan Cipanas di desa Cipanas sebanyak 51 anggota dan desa Cimacan sebanyak 45 anggota	

Berikut adalah data induk kategori penyakit yang akan digunakan untuk menentukan materi penyuluhan pada Tabel 8.

Tabel 8. Data Induk Penyakit

kode	nama
A00-A09	Penyakit infeksi usus
A15-A19	Tuberkulosis
A20-A28	Penyakit bakteri zoonosis tertentu
A30-A49	Penyakit bakteri lainnya
A50-A64	Infeksi dengan cara penularan yang sebagian besar melalui hubungan seksual
...	...
Z77-Z99	Orang dengan potensi bahaya kesehatan yang berkaitan dengan riwayat keluarga dan pribadi serta kondisi tertentu yang memengaruhi status kesehatan

Setelah mengetahui hasil karakteristik pengelompokan, kemudian ditinjau kode penyakit yang diambil untuk dicocokkan dengan kode yang pada data induk penyakit pada Tabel 8. Rekomendasi materi penyuluhan yang diberikan berdasarkan hasil karakteristik yang sudah dibuat dapat dilihat pada Tabel 9.

Tabel 9. Rekomendasi Materi Penyuluhan

No	Materi	Detail Penyakit	Wilayah		Total Pasien
			Kecamatan	Desa	
1	Infeksi saluran pernapasan atas akut	<i>Acute nasopharyngitis [common cold]</i>	Pacet	Ciherang, Cipendawa, Gadog	432
		<i>Acute pharyngitis</i>	Cianjur		
		<i>Acute upper respiratory infection, unspecified</i>			
2	Bentuk-bentuk penyakit jantung lainnya	<i>Congestive heart failure</i>	Cipanas	Cipanas, Sindanglaya, Cimacan	27
3	Dermatitis dan eksim	<i>Allergic contact dermatitis</i>	Pacet Cianjur	Ciputri, Ciherang, Gadog	27

3.5 Evaluation

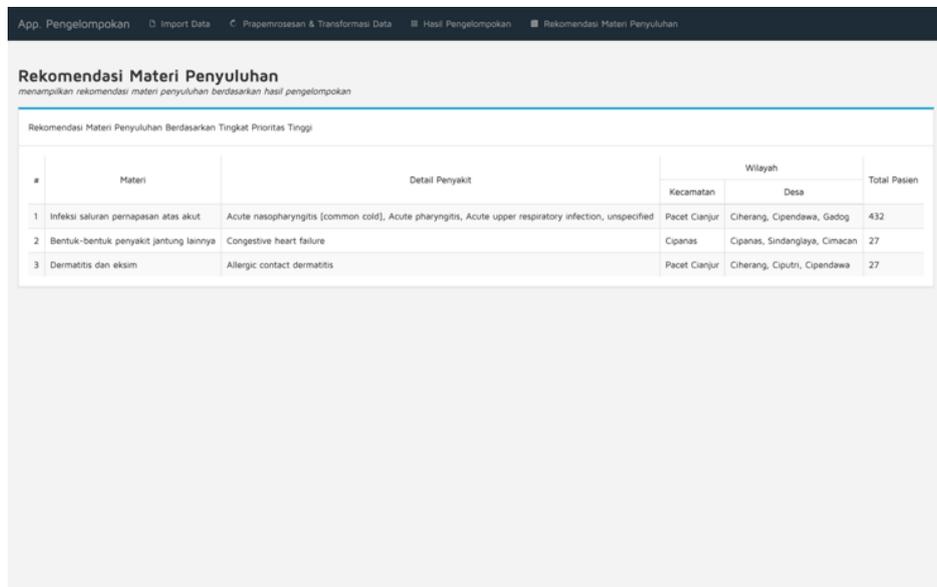
Tahap selanjutnya adalah melakukan evaluasi terhadap pengelompokan yang telah dilakukan dengan menggunakan metrik *Davies Bouldin Index (DBI)*. DBI merupakan salah satu metode yang dipakai untuk mengevaluasi *cluster* yang diperkenalkan oleh David L.Davies dan Donald W.Bouldin [21]. DBI adalah salah satu cara untuk mengukur kualitas cluster pada *clustering* [22]. Cluster dikatakan optimal jika memiliki nilai kohesi yang rendah dengan nilai separasi yang tinggi [23]. Berikut adalah hasil evaluasi model yang telah dilakukan:

$$DBI = \frac{(0,19748131+0,296284057+0,296284057)}{3} = 0,263349808$$

Hasil DBI menghasilkan nilai akhir mendekati 0 yang berarti bahwa hasil *clustering* ini mempunyai kualitas yang baik dan tingkat kemiripan yang lumayan tinggi pada satu kelompoknya.

3.6 Integration

Pada tahapan ini dilakukan implementasi terhadap model yang sudah dibuat dalam bentuk perangkat lunak. Perangkat lunak yang dibangun diharapkan dapat mudah dipahami dan digunakan oleh pengguna. Tampilan antarmuka pada perangkat lunak yang dibangun dapat dilihat pada Gambar 3.



Gambar 3. Tampilan Antarmuka Aplikasi

3.6 Pengujian

Pengujian dilakukan melalui beberapa jenis pengujian sesuai dengan kebutuhan. Berikut adalah pengujian yang telah dilakukan.

3.6.1 Pengujian Fungsionalitas

Pengujian ini dilakukan menggunakan *black box testing* yang berfokus pada persyaratan fungsional perangkat lunak yang dibangun [24]. Pengujian dilakukan dengan menguji setiap proses untuk kemungkinan kesalahan yang terjadi. Berdasarkan hasil pengujian dengan kasus sampel uji yang sudah ditentukan dapat ditarik kesimpulan bahwa seluruh fungsi berjalan dengan baik dan sesuai dengan yang diharapkan.

3.6.2 Pengujian Pengelompokan

Pengujian pengelompokan penyakit dilakukan berdasarkan skenario yang dirancang untuk memastikan bahwa pengelompokan yang dilakukan sudah sesuai dan dapat mengetahui pola penyebaran penyakit. Data yang digunakan menghasilkan kelompok penyakit yaitu kelompok 1 sebanyak 422 anggota, kelompok 2 sebanyak 27 anggota dan kelompok 3 sebanyak 27 anggota. Hasil pengujian yang dilakukan sudah sesuai dengan hasil perhitungan yang dilakukan secara manual.

3.6.3 Pengujian Rekomendasi Materi Penyuluhan

Pengujian rekomendasi materi penyuluhan dilakukan berdasarkan data uji hasil pengelompokan penyakit. Data yang digunakan selanjutnya akan dilakukan rekapitulasi sehingga akan menghasilkan karakteristik kelompok penyakit dan rekomendasi materi penyuluhan. Hasil pengujian yang dilakukan memberikan rekomendasi materi sesuai dengan hasil perhitungan yang dilakukan secara manual.

3.6.3 Pengujian Data Hasil Pengelompokan

Pengujian dilakukan terhadap data hasil pengelompokan untuk mengetahui nilai akurasi yang didapat pada pembangunan model *data mining*. Pada pengujian menghasilkan nilai akurasi rata-rata sebesar 0.263 dari 1769 baris *record* yang diambil. Hasil pengujian yang dilakukan memberikan rekomendasi materi sesuai dengan hasil perhitungan yang dilakukan secara manual.

3.6.4 Pengujian terhadap Pengguna

Pengujian ini dilakukan dengan cara melakukan kegiatan wawancara untuk mengetahui kesesuaian antara kebutuhan dengan penggunaan perangkat lunak yang dibangun [25]. Pihak yang di wawancarai yaitu Koordinator Promkes & *Marketing* di Klinik Keluarga. Berdasarkan hasil wawancara yang dilakukan terhadap pengguna dan dapat disimpulkan bahwa aplikasi yang dibangun dapat dipahami dengan baik dan membantu pihak Promkes & *Marketing* untuk menentukan materi penyuluhan yang tepat dengan kondisi masyarakat yang akan diberikan materi penyuluhan kesehatan tanpa mengalami masalah dan kendala.

4. PENUTUP

Berdasarkan hasil penelitian yang telah dilakukan mulai dari perancangan, implementasi hingga pengujian dapat disimpulkan bahwa aplikasi pengelompokan penyakit ini dapat menghasilkan kelompok penyakit sehingga dapat memberikan rekomendasi materi penyuluhan terhadap Koordinator Promkes & *Marketing* di Klinik Keluarga sehingga materi penyuluhan yang dilakukan dapat sesuai dengan kondisi masyarakat dan dapat ditangani dengan tepat oleh pihak klinik keluarga.

Adapun saran untuk pengembangan selanjutnya adalah di harapkan rekomendasi materi yang telah dihasilkan dapat disimpan ke dalam database agar proses pengelompokan tidak dilakukan berulang kali.

DAFTAR PUSTAKA

- [1] Riskesdas, *Laporan Riskesdas Provinsi Jawa Barat*. 2018. [Online]. Available: <https://litbang.kemkes.go.id>
- [2] M. R. Maulana and R. Lubis, "Sistem Informasi Manajemen Persediaan Obat Di Gudang Apotek Keluarga Cianjur," *Komputa J. Ilm. Komput. dan Inform.*, vol. 10, no. 2, pp. 53–60, 2021, doi: 10.34010/komputa.v10i2.6804.
- [3] K. P. Sinaga and M. S. Yang, "Unsupervised K-means clustering algorithm," *IEEE Access*, vol. 8, pp. 80716–80727, 2020, doi: 10.1109/ACCESS.2020.2988796.
- [4] J. Ha, M. Kambe, and J. Pe, *Data Mining: Concepts and Techniques*. 2011. doi: 10.1016/C2009-0-61819-5.
- [5] S. Saeed, A. Shaikh, M. A. Memon, and S. M. R. Naqvi, "Impact of Data Mining Techniques to Analyze Health Care Data," *J. Med. Imaging Heal. Informatics*, vol. 8, no. 4, pp. 682–690, 2018, doi: 10.1166/jmih.2018.2385.
- [6] T. M. Ghazal *et al.*, "Performances of k-means clustering algorithm with different distance metrics," *Intell. Autom. Soft Comput.*, vol. 30, no. 2, pp. 735–742, 2021, doi: 10.32604/iasc.2021.019067.
- [7] M. Ahmed, R. Seraj, and S. M. S. Islam, "The k-means algorithm: A comprehensive survey and performance evaluation," *Electron.*, vol. 9, no. 8, pp. 1–12, 2020, doi: 10.3390/electronics9081295.
- [8] F. S. Agoestina, H. S. Tambunan, and R. A. Nasution, "Pemanfaatan Algoritma K-Means untuk Pengelompokan Pasien Penyakit Infeksi Saluran Pernafasan Akut (ISPA)," *SATESI J. Sains Teknol. dan Sist. Inf.*, vol. 2, no. 1, pp. 60–69, 2022, doi: 10.54259/satesi.v2i1.804.
- [9] Okta Jaya Harmaja, Hadirat Halawa, Wicarda Sandi Hulu, and Seriani Loi, "Implementasi Algoritma K-Means Clustering Untuk Pengelompokan Penyakit Pasien Pada Puskesmas Pulo Brayan," *J. Sains dan Teknol.*, vol. Volume 5 N, no. 1, pp. 150–157, 2023.
- [10] 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.
- [11] O. Maimon and L. Rokach, *Data Mining and Knowledge Discovery*. 2021. doi: 10.1007/978-981-

- 15-8983-6_42.
- [12] A. Nugraha, O. Nurdiawan, and G. Dwilestari, "Penerapan Data Mining Metode K-Means Clustering Untuk Analisa Penjualan Pada Toko Yana Sport," *JATI (Jurnal Mhs. Tek. Inform., vol. 6, no. 2, pp. 849–855, 2022, doi: 10.36040/jati.v6i2.5755.*
- [13] D. Ariyanto, "Data Mining Menggunakan Algoritma K-Means untuk Klasifikasi Penyakit Infeksi Saluran Pernafasan Akut," *J. Sistim Inf. dan Teknol., vol. 4, pp. 13–18, 2022, doi: 10.37034/jsisfotek.v4i1.117.*
- [14] D. Dharmayanti, A. M. Bachtiar, and A. C. Prasetyo, "Penerapan Metode Clustering Untuk Membentuk Kelompok Belajar Menggunakan Di Smpn 19 Bandung," *Komputa J. Ilm. Komput. dan Inform., vol. 6, no. 2, pp. 49–56, 2017, doi: 10.34010/komputa.v6i2.2477.*
- [15] B. Jumadi Dehotman Sitompul, O. Salim Sitompul, and P. Sihombing, "Enhancement Clustering Evaluation Result of Davies-Bouldin Index with Determining Initial Centroid of K-Means Algorithm," *J. Phys. Conf. Ser., vol. 1235, no. 1, 2019, doi: 10.1088/1742-6596/1235/1/012015.*
- [16] S. Hajar, A. A. Novany, A. P. Windarto, A. Wanto, and E. Irawan, "Penerapan K-Means Clustering pada ekspor minyak kelapa sawit menurut negara tujuan," *Semin. Nas. Teknol. Komput. Sains 2020, pp. 314–318, 2020.*
- [17] R. Nainggolan, R. Perangin-Angin, E. Simarmata, and A. F. Tarigan, "Improved the Performance of the K-Means Cluster Using the Sum of Squared Error (SSE) optimized by using the Elbow Method," *J. Phys. Conf. Ser., vol. 1361, no. 1, 2019, doi: 10.1088/1742-6596/1361/1/012015.*
- [18] A. T. Rahman, W. Wiranto, and R. Anggrainingsih, "Coal Trade Data Clustering Using K-Means (Case Study Pt. Global Bangkit Utama)," *ITSMART J. Teknol. dan Inf., vol. 6, no. 1, pp. 24–31, 2017.*
- [19] D. Marutho, S. Hendra Handaka, E. Wijaya, and Muljono, "The Determination of Cluster Number at k-Mean Using Elbow Method and Purity Evaluation on Headline News," *Proc. - 2018 Int. Semin. Appl. Technol. Inf. Commun. Creat. Technol. Hum. Life, iSemantic 2018, pp. 533–538, 2018, doi: 10.1109/ISEMANTIC.2018.8549751.*
- [20] T. Juninda, Mustasim, and E. Andri, "Penerapan Algoritma K-Medoids untuk Pengelompokan Penyakit di Pekanbaru Riau," *Semin. Nas. Teknol. Informasi, Komun. dan Ind., vol. 11, no. 1, pp. 42–49, 2019.*
- [21] M. Herviany, S. Putri Delima, T. Nurhidayah, and K. Kasini, "Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Daerah Rawan Tanah Longsor Pada Provinsi Jawa Barat," *MALCOM Indones. J. Mach. Learn. Comput. Sci., vol. 1, no. 1, pp. 34–40, 2021, doi: 10.57152/malcom.v1i1.60.*
- [22] A. K. Singh, S. Mittal, P. Malhotra, and Y. V. Srivastava, "Clustering Evaluation by Davies-Bouldin Index(DBI) in Cereal data using K-Means," *Proc. 4th Int. Conf. Comput. Methodol. Commun. ICCMC 2020, no. Iccmc, pp. 306–310, 2020, doi: 10.1109/ICCMC48092.2020.ICCMC-00057.*
- [23] I. W. Septiani, A. C. Fauzan, and M. M. Huda, "Implementasi Algoritma K-Medoids Dengan Evaluasi Davies-Bouldin-Index Untuk Klasterisasi Harapan Hidup Pasca Operasi Pada Pasien Penderita Kanker Paru-Paru," *J. Sist. Komput. dan Inform., vol. 3, no. 4, p. 556, 2022, doi: 10.30865/json.v3i4.4055.*
- [24] T. Suryani, A. Faisol, and N. Vendyansyah, "Sistem Informasi Geografis Pemetaan Kerusakan Jalan Di Kabupaten Malang Menggunakan Metode K-Means," *JATI (Jurnal Mhs. Tek. Inform., vol. 5, no. 1, pp. 380–388, 2021, doi: 10.36040/jati.v5i1.3259.*
- [25] A. Finandhita and O. M. Wibowo, "Visualisasi Data Harga Komoditas Pangan (Studi Kasus : Website Dinas Tanaman Pangan Dan Hortikultura Provinsi Jawa Barat)," *Komputa J. Ilm. Komput. dan Inform., vol. 7, no. 2, pp. 59–68, 2018, doi: 10.34010/komputa.v7i2.3038.*