

IMPLEMENTASI DATA MINING MENGGUNAKAN ALGORITMA K-MEANS CLUSTERING PENYAKIT PASIEN RAWAT JALAN PADA KLINIK DR. ATIRAH DESA SIOYONG, SULTENG

Rangga Bayu Prasetyo, Yosep Agus Pranoto, Renaldi Primaswara Prasetya

Program Studi Teknik Informatika S1, Fakultas Teknologi Industri
Institut Teknologi Nasional Malang, Jalan Raya Karanglo km 2 Malang, Indonesia
1918048@scholar.itn.ac.id

ABSTRAK

Banyak rumah sakit ataupun klinik yang mulai menggunakan teknologi informasi sebagai pengolahan data pasien. Hal ini tentu saja mempermudah para tenaga medis dalam hal pengolahan data pasien dan pengelempokan jenis penyakit pada salah satu instansi Kesehatan. Oleh karena itu instansi Kesehatan memerlukan aplikasi yang dapat melakukan pengelompokan jenis penyakit. Tujuan dari digunakannya metode K-Means Clustering pada penelitian ini yaitu pengelompokan berdasarkan data yang memiliki karakteristik yang sama akan berada pada satu cluster data yang sama, untuk dilakukannya pencegahan pada jenis penyakit yang masuk pada kriteria atau kategori aman, waspada, dan darurat agar dapat segera ditangani. Hasil pengujian *user* sebanyak 66% dari karyawan Klinik dr. Atirah menyetujui adanya sistem aplikasi data mining clustering penyakit. Berdasarkan dari pengujian yang telah dilakukan dari penelitian ini, bahwa pakar telah memvalidasi jumlah penyakit disetiap clustering yaitu Aman 10 penyakit, Waspada 7 penyakit dan Darurat 1 penyakit. Hasil dari perhitungan manual dan sistem hasilnya sama. Metode K-Means memberikan akurasi nilai cluster rata-rata 88.95% melalui pengujian *Confusion Matrix*.

Kata kunci : Clustering, Data Mining, K-means, Pasien Rawat Jalan

1. PENDAHULUAN

Dalam dunia Kesehatan kemajuan teknologi sangatlah dibutuhkan untuk menunjang kinerja para tenaga medis. Banyak rumah sakit ataupun klinik yang mulai menggunakan teknologi informasi sebagai pengolahan data pasien. Hal ini tentu saja mempermudah para tenaga medis dalam hal pengolahan data pasien dan pengelempokan jenis penyakit pada salah satu instansi Kesehatan dengan membandingkan nilai dari data sebelumnya.

Data yang digunakan untuk penelitian ini bersumber dari Klinik dr. Atira di desa Sioyong Sulawesi Tengah. Data yang akan diteliti berupa pengelompokan beberapa jenis penyakit, usia pada pasien penderitanya, dan gender pasien. Metode yang digunakan adalah *Data Mining* dengan algoritma K-Means Clustering. *Data Mining* merupakan pengambilan data dari data yang sudah dikelompokkan. Telah banyak pengimplementasian *Data Mining* dalam berbagai bidang, salah satunya dalam bidang Kesehatan. Metode dari *Data Mining* yang sering digunakan yaitu klasifikasi, clustering, dan regresi. Tujuan dari penggunaan *Data Mining* adalah sebagai pengambilan pola dan informasi yang digunakan untuk menemukan hubungan antara data dan pengelompokkan kedalam satu atau lebih cluster [1]

Karena meningkatnya jumlah pasien serta permintaan data oleh Puskesmas maka perlunya ada sistem informasi sebagai alat pelaporan kasus sehingga dapat dilakukan pengelompokkan, pemetaan, penanganan dan pencegahan penyakit oleh pihak Puskesmas dan Klinik. Selain itu juga pada Klinik dr.

Atira membutuhkan sistem informasi sebagai pengolahan data secara efisien agar bisa diakses oleh tenaga Kesehatan pada Klinik dr. Atira

Oleh karena itu tujuan digunakannya metode K-Means Clustering yaitu akan mengelompokkan berdasarkan data yang memiliki karakteristik yang sama akan berada pada satu cluster data yang sama, untuk dilakukannya pencegahan pada jenis penyakit yang masuk pada kriteria atau kategori aman, waspada, dan darurat sehingga untuk instansi terkait seperti Klinik dr. Atira dapat melakukan pencegahan dan penanggulangan lebih lanjut karena metode K-Means clustering akan mempermudah dalam clustering penyakit dan melaporkan hasil dari laporan kasus setiap kategori jenis penyakit

2. TINJAUAN PUSTAKA

2.1. Penelitian Terdahulu

Bayu Dwiputra Mudzakir melakukan penelitian pada tahun 2018, *Pengelompokan Data Penjualan Produk Pada PT ADVANTA SEEDS Indonesia Menggunakan Metode K-Means*. Tujuan dari penelitian yang dilakukan untuk pengklasifikasian data agar mengurangi fungsi objek yang ditetapkan dalam perhitungan klasifikasi. Penggunaan algoritma K-Means dapat membantu dalam mengelompokkan data, informasi yang ditampilkan berupa nilai centeroid dari tiap – tiap cluster, untuk mengelompokkan barang berdasarkan tingkat kelarisan barang pada PT.ADVANTA SEEDS INDONESIA [2]

Penelitian yang dilakukan oleh Ari Sulistyawati dan Supriyanto Eko pada tahun 2020 dengan judul *Implementasi Algoritma K-Means Clustering dalam*

Penentuan Siswa Kelas Unggulan. Penelitian ini bertujuan untuk pengembangan sistem informasi yang berbasis *website* yang berfungsi untuk pengelompokan siswa kelas unggulan berdasarkan laporan hasil belajar siswa. Penerapan algoritma *K-Means* dengan mempartisi data ke dalam data dalam bentuk satu atau lebih kelompok, dimana data yang memiliki representasi persamaan dalam satu kelompok dan data yang memiliki perbedaan kelompok yang lain. Pengelompokan data dalam satuan pendidikan dilakukan untuk memudahkan fasilitasi pendidikan siswa berdasarkan perbedaan kemampuannya dalam belajar dan mengikuti pembelajaran. [3]

Kemudian pada penelitian yang dilakukan oleh Andy Febrianto dkk. Pada tahun 2021 dengan judul Penerapan Metode *K-Means* untuk Clustering Pengunjung Perpustakaan ITN Malang. Penelitian ini menghasilkan aplikasi yang dapat mengelompokkan pengunjung dan peminjam menggunakan metode *K-Means*, membantu petugas perpustakaan untuk mengevaluasi informasi kunjungan dan peminjaman buku di perpustakaan ITN Malang [4]

2.2. Data Mining

Data mining merupakan serangkaian proses untuk menggali nilai tambah berupa informasi yang selama ini tidak diketahui secara manual dari suatu basis data. *Data mining* digunakan untuk mengambil pola dan informasi yang digunakan untuk menemukan hubungan antara data dan pengelompokkan kedalam satu atau lebih *cluster*, *data mining* juga bagian dari proses penemuan pengetahuan dari basis data *Knowledge Discovery in Database* [1]

Metode data mining melibatkan proses penggalian data dari berbagai sumber, pembersihan data untuk menghilangkan noise atau informasi yang tidak relevan, transformasi data ke dalam format yang lebih sesuai untuk analisis, serta penerapan algoritma dan teknik analisis untuk mengidentifikasi pola yang bermanfaat. Data mining digunakan dalam berbagai bidang seperti bisnis, ilmu pengetahuan, kesehatan, keuangan, dan lainnya untuk mendapatkan wawasan yang dapat membantu perencanaan strategis dan pengambilan keputusan. Tujuan utama dari data mining adalah untuk mengubah data mentah menjadi wawasan yang dapat diaplikasikan dalam berbagai bidang, termasuk bisnis, ilmu pengetahuan, kesehatan, keuangan, dan lainnya. Dalam konteks bisnis, data mining dapat membantu dalam mengidentifikasi tren konsumen, segmentasi pasar, rekomendasi produk, deteksi penipuan, dan masih banyak lagi [5]

2.3. Algoritma K-Means

K-Means Clustering merupakan salah satu algoritma *Machine Learning* khususnya *Unsupervised Learning* yang populer digunakan. Dalam Algoritma ini nilai k yang merupakan jumlah cluster yang akan dibentuk, dimana nilai k harus didefinisikan terlebih dahulu sebelum menggunakan algoritma *K-Means*,

nilai k yang digunakan merujuk pada jumlah *centroid* atau titik pusat dari setiap *cluster*. [6]

2.4. Euclidean Distance

Euclidean Distance salah satu metode perhitungan jarak yang berguna untuk mengukur jarak dari 2 (dua) buah titik dalam *Euclidean space* (meliputi bidang *Euclidean* dua dimensi, tiga dimensi atau bahkan lebih). Proses ini berulang hingga pusat kelompok tidak berubah secara signifikan atau iterasi maksimal tercapai, menghasilkan pengelompokan data yang optimal berdasarkan jarak *Euclidean*. [7]

2.5. Pasien Rawat Jalan

Rawat jalan merupakan pelayanan medis yang diberikan kepada pasien dengan tujuan pengobatan, diagnosis dengan hasil pemeriksaan fisik serta rencana pengobatan atau tindakan lainnya tanpa harus mewajibkan pasien dirawat inap. Pasien rawat jalan adalah individu yang menerima perawatan atau konsultasi medis dari tenaga medis atau profesional kesehatan di luar lingkungan rumah sakit. Pasien ini biasanya mengunjungi fasilitas kesehatan pada waktu yang telah dijadwalkan dan pulang ke rumah pada hari yang sama setelah perawatan atau konsultasi sudah selesai [8]

2.6. Penyakit

Menurut Amrin Madolan penyakit adalah kelainan fisiologis atau psikologis yang menyebabkan gangguan pada fungsi normal tubuh, dicirikan oleh gejala klinis tertentu, dan dapat disebabkan oleh berbagai faktor seperti infeksi, lingkungan, genetik, atau gaya hidup. [9]

2.7. Confusion Matrix

Confusion matrix merupakan pengujian perhitungan metode yang perbandingan hasil klasifikasi oleh model sistem proses klasifikasi terbagi dalam 4 bagian antara lain merupakan *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), *False Negative* (FN). [10]

3. METODE PENELITIAN

3.1. Analisis Sistem

Sistem ini dibuat dikarenakan pada klinik Dr. Atirah dibutuhkan klusterisasi penyakit pasien rawat jalan, dengan menggunakan kriteria pengolahan data yang digunakan meliputi nama pasien, umur pasien yang dibagi menjadi 4 bagian yaitu sesuai rentang umur pasien, penyakit pasien atau diagnosis pasien yang sudah terdaftar pada klinik tersebut, setelah dilakukannya olah data kemudian akan dilakukan pengelompokan atau clustering pada tahap ini yaitu menentukan kelompok clustering atau nilai dari cluster K untuk penelitian ini disebutkan hasil cluster yaitu terdapat Darurat, Waspada, dan Aman cluster ini didapatkan sesuai dengan yang ditentukan oleh pakar dari klinik tersebut, sehingga ketika sudah dilakukannya perhitungan atau pengolahan data, pakar

dr. Atirah akan memvalidasi hasil cluster yang sesuai dengan diagnosa pasien rawat jalan tersebut.

3.2. Analisis User

Analisis User yang dibutuhkan pada sistem ini untuk mengetahui kebutuhan pengguna yang bisa mengakses pada sistem yang sedang dibuat diantaranya :

1. Superadmin, sebagai pengelola utama yang berfungsi untuk mengatur semuanya yaitu seperti menambahkan pengguna baru hanya bisa diakses oleh superadmin dan superadmin sendiri berperan penting untuk memvalidasi hasil perhitungan clustering yang sudah dilakukan pada sistem
2. Admin, dalam user admin ini merupakan staff pada klinik tersebut yang hanya dibatasi sebatas menginputkan data pasien baru kemudian diagnosa penyakit baru dan melakukan proses perhitungan clustering tetapi admin tidak bisa melakukan validasi hanya halaman superadmin

3.3. Kebutuhan Fungsional

Kebutuhan fungsional untuk halaman superadmin pada website. Halaman pertama, yaitu Dashboard, bertujuan untuk menampilkan tampilan singkat yang merekap hasil dari seluruh data yang telah diolah. Selanjutnya, Data Pengguna memungkinkan tampilan informasi mengenai data pengguna yang sesuai dengan akses login. Tambah Pengguna memberikan halaman yang berguna untuk keperluan penambahan data pengguna. Data Pasien menyajikan informasi halaman terkait data pasien yang telah diinputkan. Kemudian, Tambah Pasien memberikan halaman yang bermanfaat untuk menambahkan data pasien yang akan diolah. Pada halaman Data Penyakit, data penyakit yang sesuai dengan diagnosa pasien ditampilkan, dan akan dikelompokkan sesuai dengan kriteria yang telah ditentukan. Perhitungan k-means atau clustering, pada halaman ini akan menunjukkan alur perhitungan pengelompokan atau clustering menggunakan metode k-means. Terakhir, Validasi Clustering akan menampilkan hasil dari perhitungan k-means yang berfungsi untuk divalidasi oleh pakar klinik

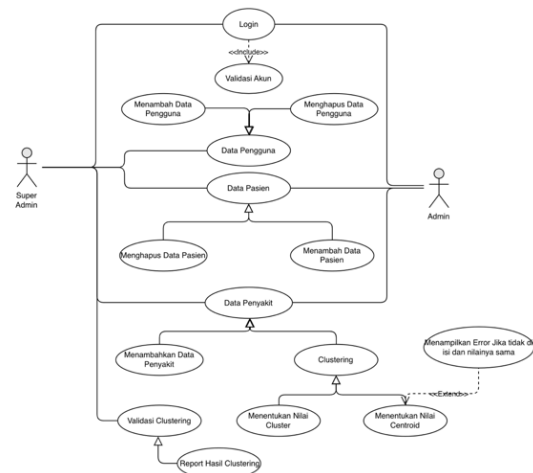
3.4. Kebutuhan Nonfungsional

Kebutuhan nonfungsional adalah kebutuhan yang dibutuhkan sistem ketika sedang berjalan adapun kebutuhan nonfungsional pada sistem yang dibangun adalah sebagai berikut :

1. Sistem yang dikembangkan hanya bisa digunakan ketika adanya koneksi internet
2. Sistem yang dikembangkan hanya diakses oleh staff Klinik dr. Atirah
3. Sistem yang dikembangkan hanya dapat berjalan atau dibuka pada browser seperti Mozilla Firefox, Google Chrome, dan Safari

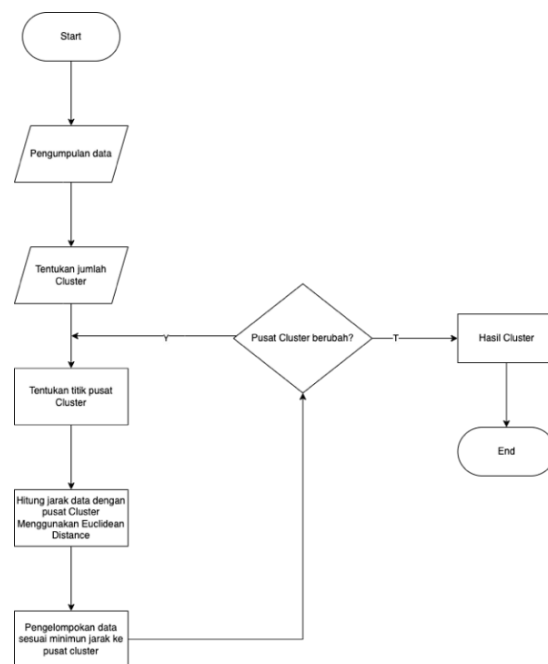
3.5. Use Case Diagram Sistem

Pada gambar 1 merupakan Use Case Diagram aplikasi clustering pada website



Gambar 1. Use case diagram

3.6. Flowchart Metode



Gambar 2. Flowchart metode

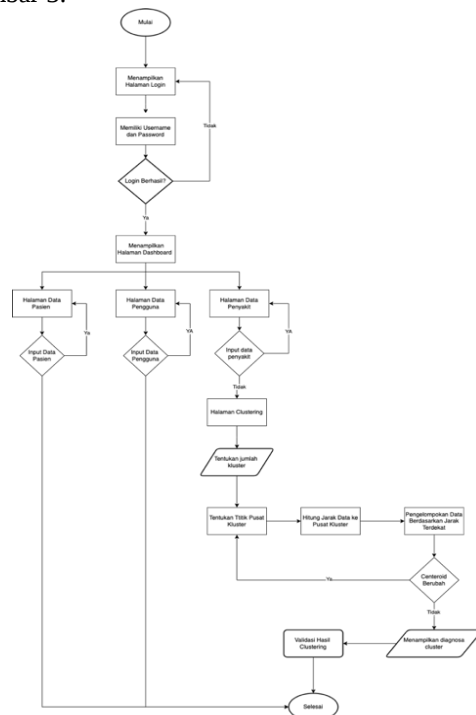
Metode K-means clustering adalah salah satu metode dalam analisis data yang digunakan untuk mengelompokkan data menjadi beberapa kelompok atau klaster berdasarkan kesamaan karakteristik. Pada gambar 2 dibawah ini flowchart memberikan gambaran langkah-langkah dalam metode K-means dengan ketentuan sebagai berikut.

1. Menentukan jumlah cluster yang akan digunakan
2. Menentukan titik pusat cluster atau yang bisa disebut centroid dari data yang sudah ditambahkan
3. Jika sudah maka hitung jarak terdekat tiap cluster menggunakan proses euclidean distance

4. Kemudian pengelompokan data sesuai dengan jarak terdekat pada cluster
5. Jika nilai jarak terdekat atau nilai hasil titik centroid cluster berubah maka akan melakukan pengulangan perhitungan lagi jika tidak berubah maka perhitungan dinyatakan selesai.

3.7. Flowchart Sistem

Flowchart sistem merupakan gambaran awal untuk berjalan nya sistem pada perancangan aplikasi website pada Klinik dr. Atirah. Proses pada flowchart tersebut merupakan proses dari mulai login kemudian meminta hak akses pada sistem sesuai dengan *role* dan menampilkan halaman – halaman yang sudah diberikan sesuai dengan *role* yang sudah dibuat sesuai dengan kebutuhan fungsional kemudian terdapat halaman perhitungan sesuai dengan metode hingga report yang akan mengeluarkan hasil *print out* perhitungan k-means clustering flowchart sistem pada gambar 3.



Gambar 3. Flowchart sistem

3.8. Pengelompokan Penyakit

Pada pengelompokan penyakit ini kriteria cluster yang sudah ditentukan oleh pakar dr. Atirah pada penelitian ini menggunakan 3 kriteria cluster:

1. Aman

Pada kriteria aman menunjukkan kondisi jumlah penyakit yang di derita oleh banyaknya pasien yang datang rendah atau tidak menunjukkan adanya kejadian luar biasa (KLB). Sehingga tidak perlu adanya intervensi dan sosialisasi ke masyarakat oleh pihak terkait.

2. Waspada

Pada kriteria waspada kondisi jumlah penyakit yang di derita oleh banyaknya pasien yang datang cukup banyak atau sedang dan tidak menunjukkan kejadian luar biasa (KLB). Sehingga perlu adanya sosialisasi atau pencegahan.

3. Darurat

Pada kriteria darurat menunjukkan kondisi jumlah penyakit yang di derita oleh banyaknya pasien yang datang tinggi dan menunjukkan adanya kejadian luar biasa (KLB). Sehingga perlu adanya sosialisasi dan penanganan lebih lanjut dari pihak terkait.

Penetapan kriteria diatas digunakan untuk skrining dan laporan dari klinik ke puskesmas atau pihak terkait sebagai inovasi penanganan cepat tanggap dalam mengatasi peningkatan kasus kejadian luar biasa (KLB)

3.9. Hasil Pengolahan Data Penyakit

Setelah dilakukan pengolahan data penyakit pasien didapatkan 18 jenis penyakit dan dikelompokkan menjadi data penyakit yang akan digunakan untuk perhitungan K-Means Clustering. Pada tabel 1 dibawah ini merupakan hasil pengolahan dari seluruh data pasien, yang sudah di inputkan sesuai dengan diagnosa pasien dan rentang umur pasien kemudian dikelompokkan.

Tabel 1. Data penyakit

No	Jenis Penyakit	anak - anak 1- 11 thn	remaja 12- 25 tahun	dewasa 26- 45 tahun	lansia 46 tahun keatas	laki - laki	Perempuan
1.	ISPA	9	75	42	41	70	97
2.	GEA	0	28	8	8	16	28
3.	GERD	0	21	7	3	10	21
4.	VERTIGO	0	9	1	2	4	8
5.	HT	1	31	2	6	16	24
6.	DERMATITIS	0	18	10	2	15	15
7.	COUGH	1	13	1	0	6	9
8.	CC	0	11	4	6	8	13
9.	TTH	0	4	2	0	1	5
10.	CEPHALGIA	1	7	3	0	6	5
11.	DYSPEPSIA	0	17	4	1	9	13
12.	FARINGITIS	0	20	3	3	10	16
13.	GASTRITIS	1	17	3	5	11	15

No	Jenis Penyakit	anak - anak 1- 11 thn	remaja 12- 25 tahun	dewasa 26- 45 tahun	lansia 46 tahun keatas	laki - laki	Perempuan
14.	TFA	0	14	1	4	8	11
15.	TONSILITIS	0	9	6	3	6	12
16.	VULNUS	0	16	5	4	11	14
17.	URTICARIA	0	9	3	0	4	8
18.	MYALGIA	0	10	5	2	7	10

3.10. Menghitung Jarak Ke Cluster menggunakan Euclidean distance

Untuk menghitung jarak ke tiap cluster digunakannya euclidean distance yang langkah awalnya merupakan menentukan jumlah cluster terlebih dahulu pada penelitian ini, hanya menggunakan 3 cluster yang sudah ditentukan oleh pihak studi kasus, kemudian memilih titik pusat atau centroid awal secara acak pada tabel 1 data penyakit, pada penelitian ini titik pusat awal dipilih 3 data acak pada tabel data penyakit

Tabel 2. Data centroid awal

C1	HT	1	31	2	6	16	24
C2	FARINGITIS	0	20	3	3	10	16
C3	VULNUS	0	16	5	4	11	14

Setelah mendapatkan titik pusat atau *centroid* awal, langkah selanjutnya merupakan proses perhitungan dengan *euclidean distance* untuk menemukan jarak terdekat tiap cluster

1. Menghitung jarak data *centroid* pertama C1

$$d(P_1, C_1) = (9 - 1)^2 + (75 - 31)^2 + (42 - 2)^2 + (41 - 6)^2 + (70 - 16)^2 + (97 - 24)^2 = 13.070$$

$$= \sqrt{13.070}$$

$$= 114,324101$$

2. Menghitung jarak data *centroid* kedua C2

$$d(P_1, C_2) = (9 - 0)^2 + (75 - 20)^2 + (42 - 3)^2 + (41 - 3)^2 + (70 - 10)^2 + (97 - 16)^2 = 16.232$$

$$= \sqrt{16.232}$$

$$= 127,404866$$

3. Menghitung jarak data *centroid* ketiga C3

$$d(P_1, C_3) = (9 - 0)^2 + (75 - 16)^2 + (42 - 5)^2 + (41 - 4)^2 + (70 - 11)^2 + (97 - 14)^2 = 16.670$$

$$= \sqrt{16.670}$$

$$= 129,112354$$

Sesudah dilakukan nya perhitungan seperti cara diatas kemudian didapatkannya kesimpulan jarak terdekat dari data ke 1 merupakan golongan cluster 1 karena nilai minimum pada C1, C2, dan C3 nilai minimumnya berada pada C1, ulangi perhitungan di semua data penyakit pada tabel 1 Data Penyakit setelah selesai kemudian cari nilai optimal *centroid* kemudian dihitung lagi hingga hasil cluster dinyatakan berhenti ketika cluster tidak berubah posisi, hasil semua *cluster* bisa dilihat pada tabel 3 hasil perhitungan tiap *cluster*

Tabel 3. Hasil perhitungan cluster

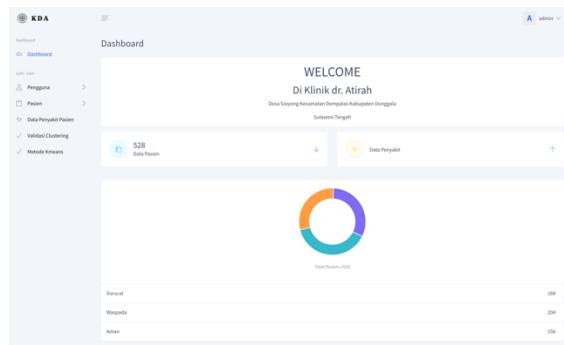
Jenis Penyakit	C1 (Darurat)	C2 (Wasapada)	C3 (Aman)	Jarak Terdekat
ISPA	0	122,3566216	138,2162798	0
GEA	110,5079183	12,35198075	28,71828686	12,35198075
GERD	122,6458316	4,035556255	16,83270626	4,035556255
VERTIGO	141,1240589	19,50824002	3,367491648	3,367491648
HT	114,3241007	11,80798761	27,62136854	11,80798761
DERMATITIS	124,9959999	7,801098824	14,91777463	7,801098824
COUGH	138,2389236	16,07126984	3,916631206	3,916631206
CC	133,2141134	13,21254815	6,044832504	6,044832504
TTH	147,0646117	25,95050233	9,345587194	9,345587194
CEPHALGIA	143,0034965	21,8958574	5,842944463	5,842944463
DYSPEPSIA	131,4001522	9,196272537	8,315046602	8,315046602
FARINGITIS	127,4048665	5,182387756	12,48759384	5,182387756
GASTRITIS	128,2575534	6,824326236	10,63672882	6,824326236
TFA	134,5065054	12,78391847	5,435071297	5,435071297
TONSILITIS	136,0073527	15,95529469	4,352011029	4,352011029
VULNUS	129,1123542	7,708992893	9,409569597	7,708992893
URTICARIA	141,1240589	19,50824002	3,246536616	3,246536616
MYALGIA	136,8722032	15,92841127	2,39582971	2,39582971

4. HASIL DAN PEMBAHASAN

4.1. Halaman Dashboard

Halaman dashboard merupakan halaman utama yang menampung keseluruhan data singkat yang berguna untuk mengetahui total keseluruhan data yang sudah diolah kemudian terdapat menu menu sesuai hak akses yang sudah ditentukan oleh peneliti. Pada gambar 1 merupakan tampilan halaman utama

dashboard yang berfungsi menampung keseluruhan hasil dari pengolahan data dan berisikan menu menu yang akan diakses



Gambar 1. Halaman dashboard

4.2. Halaman Data Penyakit

Pada Gambar 2 Tampilan Data Penyakit merupakan halaman data penyakit yang berguna untuk ketika pasien masuk dan mempunyai kriteria yang sudah di tentukan maka akan masuk kedalam tabel kriteria data penyakit digunakan untuk perhitungan K-Means

Data Penyakit Pasien								
Tambah Data Penyakit								
Tampilkan Tampilkan Data								
#	Jenis Penyakit	Anak	Remaja	Dewasa	Lansia	Laki-Laki	Perempuan	Action
1	ISPA	9	75	42	41	70	97	Edit
2	GEA	0	28	8	8	16	28	Edit
3	GERD	0	21	7	3	10	21	Edit
4	VERTIGO	0	9	1	2	4	8	Edit
5	HT	1	31	2	6	16	24	Edit
6	DERMATITIS	0	18	10	2	15	15	Edit
7	COUGH	1	13	1	0	6	9	Edit
8	CC	0	11	4	6	8	13	Edit
9	TTN	0	4	2	0	1	5	Edit
10	CEPHALGIA	1	7	3	0	6	5	Edit
11	DYSPEPSIA	0	17	4	1	9	13	Edit
12	FARINGITIS	0	20	3	3	10	16	Edit
13	GASTRITIS	1	17	3	5	11	15	Edit
14	TIA	0	14	1	4	8	11	Edit
15	TONSILITIS	0	9	6	3	6	12	Edit
16	VULNUS	0	16	5	4	11	14	Edit
17	URTICARIA	0	9	3	0	4	8	Edit
18	MYALGIA	0	10	5	2	7	10	Edit

Gambar 2. Tampilan data penyakit

4.3. Halaman Menentukan centroid

Pada halaman menentukan centroid berguna untuk memilih titik pusat atau centroid dari data penyakit yang sudah di inputkan kemudian data digunakan untuk memproses perhitungan K-Means menggunakan euclidean distance

Centroid		
#	Nama Centroid	Tentukan Centroid
1	1	5. HT --> [Anak: 1] Remaja: 31 Dewasa: 2 Lansia: 6 Laki-laki: 16 Perempuan: 24
2	2	12. FARINGITIS --> [Anak: 0] Remaja: 20 Dewasa: 3 Lansia: 3 Laki-laki: 10 Perempuan: 16
3	3	16. VULNUS --> [Anak: 0] Remaja: 16 Dewasa: 5 Lansia: 4 Laki-laki: 11 Perempuan: 14

Gambar 3. Tampilan menentukan centroid

4.4. Halaman Proses Perhitungan K-means

Pada gambar 4 merupakan hasil dari perhitungan dari sistem yang menunjukkan perhitungan yang berawal dari iterasi ke-1 dan mendapat nilai atau jarak

terdekat dari hasil nya kemudian menentukan nilai centeroid dengan cara menjumlahkann nilai data yang ada di jumlah cluster Pada halaman ini berisikan perhitungan K-Means clustering yang diambil dari data penyakit kemudian di hitung dengan titik centroid yang sudah di pilih.

Jenis Penyakit	Anak	Remaja	Dewasa	Lansia	Laki-laki	Perempuan	1	2	3	JT	Cluster
ISPA	9	75	42	41	70	97	114.32431009622	127.404864006	128.112129417284	114.32431009622	1
GEA	0	28	8	8	16	28	8.12403840436	17.146421519462	19.748417658131	8.12403840436	1
GERD	0	21	7	3	10	21	13.41467064999	6.480740584079	8.944271369992	6.480740584079	2
VERTIGO	0	9	1	2	4	8	30.0331483362	15.033296378373	12.409673840991	12.409673840991	3
HT	1	31	2	6	16	24	0	15.231540211728	19.07879402839	0	1
DERMATITIS	0	18	10	2	15	15	18.220867156289	8.944271369992	7.071067813865	7.071067813865	3
COUGH	1	13	1	0	6	9	26.191801707418	11.31170849885	9.591643046254	9.591643046254	3
CC	0	11	4	6	8	13	24.389915022962	10.198039227186	6.204553203368	6.204553203368	3
TTN	0	4	2	0	1	5	36.7095208217	21.63107052794	18.79026033887	18.79026033887	3
CEPHALGIA	1	7	3	0	6	5	32.773939216348	17.77638834631	14.422105101856	14.422105101856	3
DYSPEPSIA	0	17	4	1	9	13	19.899748742132	4.898974855664	4	4	3
FARINGITIS	0	20	3	3	10	16	15.231540211728	0	5.099019513028	0	2
GASTRITIS	1	17	3	5	11	15	17.430309774653	4	2.828471247462	2.828471247462	3
TIA	0	14	1	4	8	11	22.97620586152	8.366002053408	6.164414002869	6.164414002869	3
TONSILITIS	0	9	6	3	6	12	27.450904514502	12.721922061356	8.944271369992	8.944271369992	3
VULNUS	0	16	5	4	11	14	19.07879402839	5.099019513028	0	0	3
URTICARIA	0	9	3	0	4	8	30.364452901378	15.16575888103	12.409673840991	12.409673840991	3
MYALGIA	0	10	5	2	7	10	27.270363339372	12.247448713916	8.4852813742386	8.4852813742386	3

Hasil Iterasi Centroid 1

#	Anak	Remaja	Dewasa	Lansia	Laki-laki	Perempuan
1	3.333333333333	44.666666666667	17.333333333333	18.333333333333	34	40.666666666667
2	0	10.5	5	3	10	16.5
3	0.207092070923	11.846153846154	3.6023076923077	2.2307692307692	7.3846153846154	10.435846153845

Gambar 4. Perhitungan k-means iterasi 1

4.5. Halaman Perhitungan K-means Iterasi 2

Pada gambar 5 merupakan hasil dari perhitungan dari sistem yang menunjukkan perhitungan yang berawal dari iterasi ke-1 dan mendapat nilai atau jarak terdekat dari hasil nya kemudian menentukan nilai centeroid dengan cara menjumlahkann nilai data yang ada di jumlah cluster

Jenis Penyakit	Anak	Remaja	Dewasa	Lansia	Laki-laki	Perempuan	1	2	3	JT	Cluster
ISPA	9	75	42	41	70	97	74.80586852263	125.0599895601	135.71588423892	74.80586852263	1
GEA	0	28	8	8	16	28	35.72452577517	14.71588993562	20.254333848828	14.71588993562	2
GERD	0	21	7	3	10	21	46.07170569987	3.2403703492039	14.453418218474	3.2403703492039	2
VERTIGO	0	9	1	2	4	8	66.72980940212	17.190113437671	5.80960213492128	5.80960213492128	3
HT	1	31	2	6	16	24	39.52495954984	11.98213145411	25.29787710312	11.98213145411	2
DERMATITIS	0	18	10	2	15	15	51.044442770379	8.336660002665	12.449186661812	8.336660002665	2
COUGH	1	13	1	0	6	9	63.63087994613	11.729510217746	4.3214667400579	4.3214667400579	3
CC	0	11	4	6	8	13	59.11708574201	11.59741204743	4.597400905743	4.597400905743	3
TTN	0	4	2	0	1	5	72.83983456271	23.52680712037	11.305943021967	11.305943021967	3
CEPHALGIA	1	7	3	0	6	5	68.88213388144	15.462024966041	7.936216516235	7.936216516235	3
DYSPEPSIA	0	17	4	1	9	13	56.76220590577	6.9641941385921	6.0432867204315	6.0432867204315	3
FARINGITIS	0	20	3	3	10	16	52.626567012827	3.2403703492039	10.170731880364	3.2403703492039	2
GASTRITIS	1	17	3	5	11	15	53.711316178625	5.873870962254	8.221779790076	5.873870962254	2
TIA	0	14	1	4	8	11	59.95792279025	10.93165545338	3.940388932287	3.940388932287	3
TONSILITIS	0	9	6	3	6	12	62.044786091192	13.83525203098	4.231484743846	4.231484743846	3
VULNUS	0	16	5	4	11	14	54.09511454925	6.510202405026	6.8338779438999	6.510202405026	2
URTICARIA	0	9	3	0	4	8	66.78986127234	17.07371078964	5.446479322879	5.446479322879	3
MYALGIA	0	10	5	2	7	10	62.72318593808	13.87430923512	2.3982241050954	2.3982241050954	3

Hasil Iterasi Centroid 2

#	Anak	Remaja	Dewasa	Lansia	Laki-laki	Perempuan
1	9	75	42	41	70	97
2	0.28571428571429	21.571428571429	5.4285714285714	4.4285714285714	12.714285714286	19
3	0.2	10.3	3	1.8	5.9	9.4

Gambar 5 perhitungan k-means iterasi 2

4.6. Halaman Perhitungan K-Means Iterasi 3

Tampilan Proses Perhitungan K-Means pada gambar 6 merupakan hasil dari perhitungan dari sistem yang menunjukkan perhitungan yang berawal dari iterasi ke-3 dan mendapat nilai atau jarak terdekat dari hasil nya kemudian menentukan nilai centeroid dengan

cara menjumlahkann nilai data yang ada di jumlah cluster jika nilai cluster yang di dapat sama dan centroid tidak berubah, perhitungan dinyatakan berhenti

Iterasi 2

Jenis Penyakit	Anak	Remaja	Dewasa	Lansia	Laki-laki	Perempuan	1	2	5	JT	Cluster
ISPA	9	75	42	41	70	97	74.90586852263	125.0059985661	135.7155843382	74.90586852263	1
GEA	0	28	8	8	16	28	35.71430257757	14.71338880552	26.25431384828	34.71338880552	2
GERD	0	21	7	3	10	21	48.07170898807	3.24037040209	14.49345234474	3.24037040209	2
VERTIGO	0	9	1	2	4	8	66.72968945212	17.28111347671	5.809023349228	5.809023349228	3
HT	1	31	2	6	16	24	39.52495894884	13.9823145411	25.256787713312	13.9823145411	2
DERMATITIS	0	18	10	2	15	15	51.044642770379	6.33666000365	12.44913666102	6.33666000365	2
COUGH	1	13	1	0	6	9	63.63879094813	13.72530217746	4.3214067400579	4.3214067400579	3
CC	0	11	4	6	8	13	59.11758276201	11.58741304743	4.597406807543	4.597406807543	3
TTH	0	4	2	0	1	5	72.90853456271	23.526380712037	11.90943822967	11.90943822967	3
CEPHALGIA	1	7	3	0	6	9	68.88153883144	18.8022406041	7.596139161025	7.596139161025	3
DYSEPISA	0	17	4	1	9	13	56.79223089377	6.964194138921	6.043238720433	6.043238720433	3
FARINGITIS	0	20	3	3	10	16	32.428587812827	3.24037040209	35.170713880364	3.24037040209	2
GASTRITIS	1	17	3	5	11	15	53.71316878025	5.873670862354	8.231775796079	5.873670862354	2
TIA	0	14	1	4	8	11	58.957392278925	10.91620554538	3.9493838932287	3.9493838932287	3
TONSILITIS	0	9	6	3	6	12	62.044788811502	13.83835263089	4.2314684736846	4.2314684736846	3
VULNUS	0	16	5	4	11	14	54.8913454925	6.510024053206	6.831877943899	6.510024053206	2
URITICARIA	0	9	3	0	4	8	66.79986127234	17.873371078964	5.6484789232879	5.6484789232879	3
MYALGIA	0	10	5	2	7	10	62.72378938088	13.87443025512	2.396241300054	2.396241300054	3

Hasil Iterasi Centroid 2

#	Anak	Remaja	Dewasa	Lansia	Laki-laki	Perempuan
1	9	75	42	41	70	97
2	0.2857143871429	21.57143871429	5.428571438714	4.428571438714	12.71438714386	19
3	0.2	10.3	3	1.8	5.9	9.4

Gambar 6. Halaman k-means iterasi 3

4.7. Halaman Hasil Cluster yang didapatkan

Untuk proses penentuan clustering pada pasien rawat jalan melalui diagnosa pasien yaitu dibagi menjadi 3 cluster yang digolongkan menjadi 3 bagian yaitu aman, wasapada, dan darurat, pengambilan data cluster sendiri diambil dari nilai minimum pada masing masing hasil cluster terakhir dan akan ditampilkan pada pasien yang berdiagnosa dengan ketentuan cluster tersebut, tampilan cluster bisa dilihat pada gambar 7 tampilan hasil clustering

Pasien > Simulasi

Data Pasien

Search

Link Lupa

#	Nama	Tanggal Lahir	Umur	Jenis Kelamin	Diagnosa	Dalam Cluster
1	TOKA	01-July-1970	53 Tahun	Laki-Laki	ISPA	Darurat
2	ANWAR LARUSENG	01-January-1970	53 Tahun	Perempuan	TIA	Aman
3	ANWAR LARUSENG	04-February-1995	28 Tahun	Laki-Laki	GEA	Waspada
4	RAMU	12-September-1996	26 Tahun	Laki-Laki	ISPA	Darurat
5	YUSMAN SURMATA	23-March-1957	64 Tahun	Laki-Laki	ISPA	Darurat
6	MUHAMMAD SUTAMIRIN	19-November-1959	63 Tahun	Laki-Laki	DERMATITIS	Waspada
7	KADEK JURADI	05-August-1960	61 Tahun	Laki-Laki	ISPA	Darurat
8	ARSAD SEGA GADI	31-December-1960	60 Tahun	Laki-Laki	ISPA	Darurat
9	SALUJI	02-August-1961	59 Tahun	Laki-Laki	ISPA	Darurat
10	ARMADE	04-August-1964	57 Tahun	Laki-Laki	TIA	Aman

Gambar 7. Halaman Hasil Clustering

4.8. Tabel Hasil Clustering

Didapatkan hasil perhitungan clustering seperti pada tabel 4 hasil clustering

Tabel 4. Hasil clustering penyakit

Cluster	Keterangan	Jenis Penyakit
1.	Darurat	ISPA
2.	Waspada	GEA, GERD, HT, DERMATITIS, FARINGITIS, GASTRITIS,, VULNUS
3.	Aman	VERTIGO, COUGH, CC, TTH, CEPHALGIA,

Cluster	Keterangan	Jenis Penyakit
		DYSEPISA, TFA, TONSILITIS, URITICARIA, MYALGIA

4.9. Pengujian Confusion Matrix

Terdapat pengujian confusion Matrix yang berguna untuk mengetahui tingkat keakuratan hasil dari perhitungan k-means yang berada pada sistem pengujian confusion matrix merupakan pengujian yang efektif ketika digunakan untuk melakukan pengujian dari pengklasifikasian suatu sistem dengan cara membandingkan data aktual dan data hasil prediksi sistem. Pengujian matrix confusion menggunakan 3 label yaitu Darurat, Waspada dan Aman yang merupakan.

1. Cluster Darurat

True Positive : Label aktual adalah DARURAT dan hasil clustering juga DARURAT

True Negative : Label aktual bukan DARURAT dan hasil clustering juga bukan DARURAT

False Positive : Label aktual bukan DARURAT tetapi hasil clustering adalah DARURAT

False Negative : Label aktual adalah DARURAT tetapi hasil clustering bukan DARURAT

Tabel 5. Pengujian confusion matrix cluster darurat

True Positive	True Negative	False Positive	False Negative
168	316	0	44

2. Cluster Waspada

True Positive : Label aktual adalah WASPADA dan hasil clustering juga WASPADA

True Negative : Label aktual bukan WASPADA dan hasil clustering juga bukan WASPADA

False Positive : Label aktual bukan WASPADA tetapi hasil clustering adalah WASPADA

False Negative : Label aktual adalah WASPADA tetapi hasil clustering bukan WASPADA

Tabel 6. Pengujian confusion matrix cluster waspada

True Positive	True Negative	False Positive	False Negative
175	312	29	12

3. Cluster Aman

True Positive : Label aktual adalah AMAN dan hasil clustering juga AMAN

True Negative : Label aktual bukan AMAN dan hasil clustering juga bukan AMAN

False Positive : Label aktual bukan AMAN tetapi hasil clustering adalah AMAN

False Negative : Label aktual adalah AMAN tetapi hasil clustering bukan AMAN

Tabel 7. Pengujian confusion matrix cluster waspada

True Positive	True Negative	False Positive	False Negative
83	355	73	17

Jika sudah mendapatkan nilai sesuai dengan *true positive*, *true negative*, *false positive*, dan *false negative* selanjutnya mengetahui tingkat akurasi,

1. Accuracy

Menghitung nilai akurasi dari setiap label confusion matrix dengan menggunakan cara perhitungan akurasi

$$\frac{(True\ Positive + True\ Negative)}{(TP + TN + FP + FN)} \quad (1)$$

True Positive = 168
 True Negative = 316
 False Positive = 0
 False Negative = 44
 Accuracy = $(168 + 316) / (168 + 316 + 0 + 44)$
 = $484 / 528 = 0.916 * 100 = 91.67\%$

Tabel 8. Hasil akurasi tiap cluster

C1 = Darurat	C2 = Waspada	C3 = Aman	Rata - Rata
91.67%	92.23%	82.95%	88.95%

5. KESIMPULAN DAN SARAN

Kesimpulan dari penelitian ini adalah bahwa sistem berhasil melakukan klasterisasi pada data pasien berdasarkan diagnosa dan rentan usia, menciptakan klaster sesuai kriteria yang telah ditentukan. Selain itu, sistem juga dapat mengelompokkan diagnosa pasien ke dalam klaster yang sesuai. Hasil klasterisasi ini ditampilkan dalam laporan perhitungan bersama dengan jenis penyakit yang telah dikelompokkan. Sistem ini juga telah diuji secara fungsional dan berjalan dengan baik di berbagai browser seperti Mozilla Firefox, Google Chrome, dan Safari. Hasil perhitungan manual dan hasil yang diperoleh melalui sistem menunjukkan kesamaan. Hasil survei menunjukkan bahwa sekitar 66% pengguna menyetujui penggunaan sistem aplikasi data mining clustering penyakit ini. Metode K-Means berhasil memberikan akurasi nilai cluster rata-rata sebesar 88.95% melalui pengujian menggunakan confusion matrix. Sebagai saran untuk penelitian selanjutnya, ada beberapa hal yang bisa dipertimbangkan, termasuk kemungkinan pengembangan lebih lanjut menjadi sistem yang lebih kompleks. Selain itu, fokus dapat diberikan pada pengembangan aplikasi berbasis mobile agar lebih mudah diakses oleh pengguna. Selain metode K-Means, eksplorasi terhadap metode lain seperti K-Medoids dan algoritma C4.5 juga bisa menjadi fokus penelitian selanjutnya.

DAFTAR PUSTAKA

- [1] P. Alkhairi dan A. P. Windarto, "Penerapan K-Means Cluster pada Daerah Potensi Pertanian

Karet Produktif di Sumatera Utara," *Seminar Nasional Teknologi Komputer & Sains*, hlm. 762–767, 2019.

- [2] B. D. Mudzakkir, "Pengelompokan Data Penjualan Produk Pada Pt Advanta Seeds Indonesia Menggunakan Metode K-Means," *Jurnal Mahasiswa Teknik Informatika*, vol. 2, no. 2, hlm. 34–40, 2018.
- [3] A. Sulistiyawati dan E. Supriyanto, "Implementasi Algoritma K-means Clustering dalam Penentuan Siswa Kelas Unggulan," *Jurnal Tekno Kompak*, vol. 15, no. 2, hlm. 25, 2021, doi: 10.33365/jtk.v15i2.1162.
- [4] A. Febriyanto, S. Achmadi, dan A. P. Sasmito, "Penerapan Metode K-Means Untuk Clustering Pengunjung Perpustakaan Itn Malang," *Jurnal Mahasiswa Teknik Informatika*, vol. 5, no. 1, hlm. 61–70, 2021.
- [5] N. J., "Penerapan Data Mining Clustering Dalam Mengelompokkan Buku Dengan Metode K-Means," *Indonesian Journal of Computer Science*, vol. 10, no. 1, 2021, doi: 10.33022/ijcs.v10i1.3008.
- [6] G. D. Nursyafitri, "K-Means Clustering, Salah Satu Contoh Teknik Analisis Data Populer," *dqlab*, 2022. <https://dqlab.id/k-means-clustering-salah-satu-contoh-teknik-analisis-data-populer> (diakses 10 Juli 2023).
- [7] M. Nishom, "Perbandingan Akurasi Euclidean Distance, Minkowski Distance, dan Manhattan Distance pada Algoritma K-Means Clustering berbasis Chi-Square," *Jurnal Informatika: Jurnal Pengembangan IT*, vol. 4, no. 1, hlm. 20–24, 2019, doi: 10.30591/jpit.v4i1.1253.
- [8] I. B. Wiyata, "Apa Itu Rekam Medis? Ini Penjelasan Lengkapnya," *LUMINA MAGAZINE Institut Ilmu Kesehatan Bhakti Wiyata*, 2022.
- [9] A. Madolan, "Pengertian Penyakit menurut WHO, KBBI, Kemenkes RI, dan Para Ahli," *Mitra Kesmas*, 2023. <https://www.mitrakesmas.com/2023/03/pengertian-penyakit-menurut-who-kbbi.html> (diakses 10 Juli 2023).
- [10] F. Hamdi Bahar, N. Indah Sari, dan D. Armin Lawi, "Klasifikasi Suara Kucing dan Anjing Menggunakan LSTM-GRU dan ANN-BP," *Konferensi Nasional Ilmu Komputer (KONIK) 2021*, hlm. 202–207, 2021, [Daring]. Tersedia pada: <https://www.kaggle.com/mmoraux/audio-cats-and-dogs>.