



Implementasi Data Mining Dalam Pengelompokan Data Penyakit Pasien Menggunakan Metode K-Means Clustering

Gusna Mayang Sari^{1*}, Irfan AP²

^{1,2}Informatika, Fakultas Teknik, Universitas Sulawesi Barat
ipalaloi@gmail.com

Abstract

This study aims to identify the pattern and distribution of diseases at the Mambi Health Center using the K-Means Clustering method. Patient data includes age, gender, address, and type of disease suffered. This method was chosen because of its ability to group data based on similar characteristics, making it easier to identify dominant diseases. The results reveal the age groups susceptible to certain diseases, the geographical distribution of diseases with high prevalence, and the types of diseases often encountered at the Puskesmas. These insights are expected to help Puskesmas in allocating resources effectively, designing more appropriate prevention programs, and improving the quality of health services. This study also provides recommendations for disease management strategies based on the identified groups, so as to support efforts to improve public health.

Keywords: K-Means Clustering, patient data, healthcare

Abstrak

Penelitian ini bertujuan mengidentifikasi pola dan distribusi penyakit di Puskesmas Mambi menggunakan metode K-Means Clustering. Data pasien meliputi usia, jenis kelamin, alamat, serta jenis penyakit yang diderita. Metode ini dipilih karena kemampuannya mengelompokkan data berdasarkan karakteristik serupa, sehingga memudahkan identifikasi penyakit dominan. Hasil penelitian mengungkap kelompok usia rentan terhadap penyakit tertentu, distribusi geografis penyakit dengan prevalensi tinggi, dan jenis penyakit yang sering dijumpai di Puskesmas. Wawasan ini diharapkan membantu Puskesmas dalam mengalokasikan sumber daya secara efektif, merancang program pencegahan yang lebih tepat, serta meningkatkan kualitas pelayanan kesehatan. Penelitian ini juga memberikan rekomendasi strategi pengelolaan penyakit berdasarkan kelompok yang teridentifikasi, sehingga dapat mendukung upaya peningkatan kesehatan masyarakat.

Kata kunci: K-Means Clustering, data pasien, kesehatan.

1. Pendahuluan

Puskesmas Mambi, yang terletak di Kecamatan Mambi, Kabupaten Mamasa, Sulawesi Barat, merupakan fasilitas kesehatan masyarakat yang melayani puluhan pasien setiap hari. Sistem pendaftaran dan pencatatan data pasien yang masih manual menyebabkan keterbatasan dalam memperoleh informasi mengenai penyebaran penyakit di wilayah tersebut. Hal ini berdampak pada pengambilan kebijakan yang kurang optimal dalam pencegahan dan pengobatan penyakit.

Penelitian ini menggunakan algoritma K-Means Clustering untuk mengelompokkan data penyakit pasien berdasarkan nama, umur, alamat, jenis kelamin, dan diagnosis penyakit. Teknik ini diharapkan dapat membantu Puskesmas Mambi dalam mengetahui penyakit yang paling sering dialami oleh pasien, serta meningkatkan efektivitas pelayanan kesehatan dan penyuluhan di wilayah tersebut. Penelitian ini berjudul "Implementasi Data Mining dalam Pengelompokan

Data Penyakit Pasien di Puskesmas Mambi Kabupaten Mamasa Menggunakan Metode K-Means Clustering.

2. Metode Penelitian

2.1. Data Mining

Merupakan sebuah langkah dalam proses Knowledge Discovery in Database (KDD) yang terdiri dari penerapan analisis data dan penemuan algoritma yang menghasilkan enumerasi tertentu terhadap pola pada data. Dan juga mengartikan data mining sebagai sebuah proses ekstraksi informasi baru dari sejumlah besar data yang dapat berguna dalam proses pengambilan keputusan. Proses penambahan pengetahuan dari sejumlah besar data spasial dikenal sebagai spatial data mining. (Haryanto, Vatesia and Faurina, 2021) Istilah Data Mining memiliki beberapa pandangan, seperti knowledge discovery ataupun pattern recognition. Istilah knowledge discovery atau penemuan pengetahuan tepat digunakan karena tujuan utama dari data mining memang untuk mendapatkan

pengetahuan yang masih tersembunyi di dalam bongkahan data. Sedangkan istilah untuk pattern recognition atau pengenalan pola tepat untuk digunakan untuk memenuhi pola yang tersembunyi di dalam bongkahan data. (Zulfa, Auliya and Zaenal. 2021).

2.2. K-Means Clustering

K-Means merupakan algoritma Clustering. Tujuan nya yaitu untuk membagi data menjadi beberapa kelompok. Algoritma ini menerima masukan yang berupa data. Pada algoritma ini, computer mengelompokkan data-data yang menjadi masukannya tanpa mengetahui terlebih dahulu target kelasnya. Masukan yang di terima yaitu data atau objek dan k buah kelompok (cluster) yang di inginka. Algoritma ini akan mengelompokkan data atau objek ke dalam k buah kelompok tersebut. Pada setiap cluster terdapat titik pusat (centroid) yang merepresentasikan cluster tersebut.

Algoritma K-means clustering dilakukang dengan proses sebagai berikut:

Langkah 1: Pilih K buah titik centroid secara acak

Langkah 2: Kelompokkan data sehingga terbentuk K buah cluster dengan titik centroid dari setiap cluster merupakan titik centroid yang telah dipilih sebelumnya

Langkah 3: Perbaharui nilai titik centroid

Langkah 4: Ulangi langkah 2 dan 3 sampai nilai dari titik centroid tidak lagi berubah Langkah. (May and Abrianto. 2018).

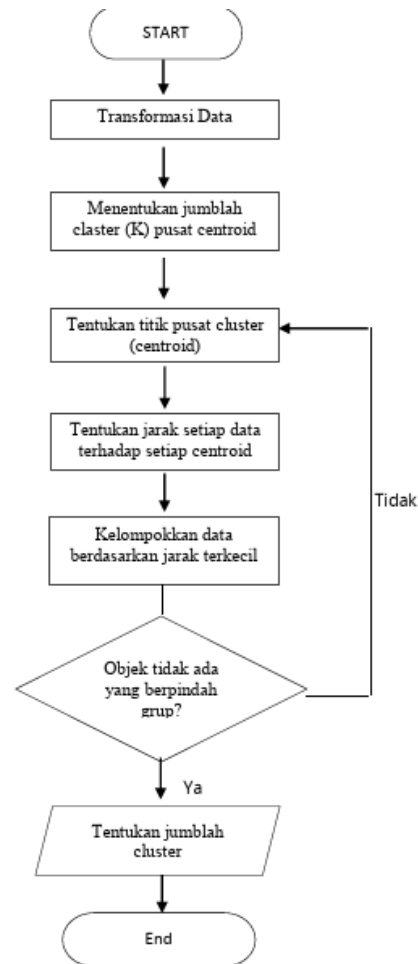
3. Hasil dan Pembahasan

Penelitian ini menganalisis jumlah kunjungan pasien yang besar di Puskesmas Mambi, di mana proses pendaftaran dan pencatatan data pasien masih dilakukan secara manual. Setiap data kunjungan pasien dicatat secara tertulis dalam map, termasuk hasil pemeriksaan, yang juga dimasukkan ke dalam rekam medis secara manual hingga seluruh proses layanan selesai. Cara kerja ini mengakibatkan keterbatasan informasi mengenai penyebaran penyakit yang sering dialami oleh pasien, terutama di berbagai wilayah di sekitar Puskesmas Mambi. Keterbatasan ini berdampak pada kurang optimalnya pengambilan keputusan dalam hal kebijakan kesehatan, antisipasi pengobatan, serta pencegahan penyakit di masyarakat.

Untuk mengatasi masalah tersebut dan mendapatkan informasi yang lebih efektif tentang pola penyakit di wilayah tersebut, penulis menggunakan Teknik Data Mining dengan algoritma K-Means Clustering. Metode ini digunakan untuk mengelompokkan data pasien berdasarkan beberapa variabel, seperti nama, jenis kelamin, umur, alamat, dan diagnosis penyakit. Hasil pengelompokan ini diharapkan dapat memberikan gambaran yang lebih jelas bagi petugas Puskesmas

dalam mengetahui jenis penyakit yang paling sering dialami oleh pasien. Selain itu, hasil ini juga diharapkan dapat memberikan manfaat dalam penyuluhan kesehatan, pengobatan yang lebih tepat, serta peningkatan mutu layanan kesehatan bagi masyarakat di wilayah Puskesmas Mambi.

3.1. Perancangan Sistem



Gambar 1. Alur K-Means

3.2. Tahapan Data Mining

Tahapan pada data mining diperlukan agar mempermudah proses pengambilan data yang akan dijadikan informasi yang bermanfaat dari tambang data yang besar. Berikut ini adalah tahapan pada data mining dalam membangun sistem data mining dalam menentukan pengelompokkan data pasien berdasarkan jenis penyakit dengan metode k-means clustering.

Cleaning Data: Sebelum memasuki tahap data mining, kita harus membersihkan data terlebih dahulu. Proses ini melibatkan langkah-langkah seperti menghapus entri kosong, menyingkirkan data ganda, memeriksa konsistensi data, dan memperbaiki kesalahan. Dalam hal ini, kita punya data pasien dari puskesmas Mambi tahun 2023. Awalnya, kita punya 1.200 entri data pasien. Setelah proses pembersihan data, di mana kita menghilangkan entri kosong dan hanya memilih data

yang relevan untuk analisis, setelah pembersihan data selesai maka data bersih yang di peroleh berjumlah 1.050 entri data pasien yang siap digunakan. Data yang sudah bersih ini kemudian bisa diambil sebagai sampel untuk melakukan perhitungan manual pada tahap mining, dengan jumlah pasien sebanyak 1.050.

Transformasi Data: Proses transformasi data mengubah informasi dalam data dari bentuk karakter menjadi bentuk numerik. Misalnya, dalam penelitian ini, data yang telah dipilih untuk analisis akan diubah menjadi angka agar dapat digunakan dalam algoritma k-means clustering. Ada empat kriteria yang digunakan dalam penelitian ini: jenis kelamin, usia, alamat dan diagnosa penyakit. Berikut adalah penjelasan tentang empat kriteria tersebut yang telah diubah menjadi format numeric.

3.3. Tahapan Mining

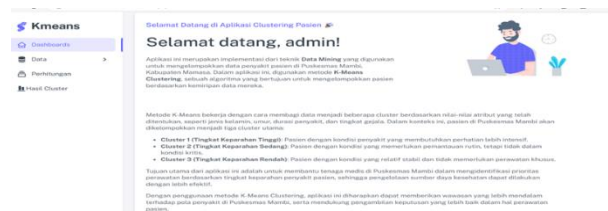
Saat kita mencari informasi baru dari data yang besar, atau dalam konteks ini, data yang digunakan dalam penelitian, dengan menggunakan algoritma clustering seperti k-means. Berikut adalah tahapan dalam menganalisis algoritma k-means:

Menentukan Jumlah Cluster: Sebelum kita mulai proses pengelompokkan menggunakan algoritma K-means, langkah pertama yang harus dilakukan adalah menentukan jumlah cluster yang akan digunakan. Dalam konteks pengelompokkan data pasien berdasarkan jenis penyakit, kita memilih menggunakan 3 cluster.

Menentukan titik cluster: Dalam tahap ini, kita akan menentukan nilai pusat cluster (centroid) dari data yang digunakan. Centroid akan diperoleh berdasarkan beberapa kriteria, yaitu jenis kelamin, alamat desa, usia, dan diagnosa penyakit. Untuk memulai, centroid akan diinisialisasi secara acak dari data pasien yang akan digunakan. Berikut adalah nilai centroid yang diperoleh secara acak dari data pasien: Jenis Kelamin, Umur, Alamat, Diagnose Penyakit.

Dengan inisialisasi centroid yang dipilih secara acak, kita dapat melanjutkan proses pengelompokkan data menggunakan algoritma K-means.

3.3. Hasil Proses Sistem



Gambar 2. Tampilan Home

Tampilan halaman data pasien setelah pengguna berhasil mengakses sistem.

ID	NAMA	JENIS KELAMIN	UMUR	ALAMAT	PENYAKIT	KAMI
1	Ardiansa	1	1	1	4	1
2	Hanifah	1	2	1	2	1
3	Iti Sembah	1	1	1	1	1
4	Abdullah	2	2	1	1	1
5	Ridwan kamil	2	1	2	2	1
6	Nur Fadiah	1	2	2	5	1
7	Jumrahah	1	1	2	5	1
8	Waniyaka	2	1	6	5	1

Gambar 3. Form Data Pasien

Tampilan halaman perhitungan, Dimana sistem melakukan perhitungan secara otomatis.

NAMA	JENIS KELAMIN	UMUR	ALAMAT	PENYAKIT	C1	C2	C3	CLUSTER
Ardiansa	1	1	1	4	3.92	0	2.24	C2
Hanifah	1	2	1	2	2	2.24	4	C1
Iti Sembah	1	1	1	1	1.41	3	5.1	C1
Abdullah	2	2	1	1	1.41	5.32	5.1	C1
Ridwan kamil	2	1	2	2	1	2.45	4.36	C1
Nur Fadiah	1	2	2	5	4.24	1.73	1.41	C3
Jumrahah	1	1	2	5	4.92	1.41	1.73	C3
Waniyaka	2	1	6	5	4.92	1.41	1.73	C3

Gambar 4. Form Perhitungan

Pada gambar berikut menunjukkan hasil klaster, yang menampilkan data berdasarkan hasil pengelompokkan klaster.

ID	NAMA	JK	UP	D	DP	CLUSTER	TINDAK KEPERAWATAN
1	Ardiansa	1	1	1	4	C2	Sedang (C2)
2	Hanifah	1	2	1	2	C1	Tinggi (C1)
3	Iti Sembah	1	1	1	1	C1	Tinggi (C1)
4	Abdullah	2	2	1	1	C1	Tinggi (C1)
5	Ridwan kamil	2	1	2	2	C1	Tinggi (C1)
6	Nur Fadiah	1	2	2	5	C3	Rendah (C3)
7	Jumrahah	1	1	2	5	C3	Rendah (C3)
8	Waniyaka	2	1	6	5	C3	Rendah (C3)

Gambar 5. Form hasil kluster

Klasifikasi hasil analisa juga dapat ditampilkan dalam bentuk visual grafik.



Gambar 6. Form tampilan grafik penyakit

4. Kesimpulan

Penelitian di Puskesmas Mambi menggunakan metode k-means clustering berhasil mengelompokkan data pasien berdasarkan jenis kelamin, umur, alamat, dan diagnosa penyakit. Hasilnya menunjukkan bahwa hipertensi mendominasi di wilayah Mambi, disusul oleh ISPA, dyspepsia, DM, dan myalgia. Temuan ini

memberikan gambaran distribusi penyakit yang jelas, membantu Puskesmas dalam merencanakan strategi kesehatan yang lebih tepat, seperti memperkuat program pencegahan hipertensi, edukasi diabetes, dan penanganan ISPA. Hasil penelitian juga mendukung alokasi sumber daya yang lebih efisien, sehingga pelayanan kesehatan dapat lebih efektif dan sesuai kebutuhan tiap wilayah.

Daftar Rujukan

- [1]. Laode muhammad alghazali,s.kep.,ns. 2023. Puskesmas mambi.
- [2]. Damanik, y.f.s.y. *Et al.* (2021) 'penerapan data mining untuk pengelompokan penyebaran covid-19 di sumatera utara menggunakan algoritma k- means', *jurnal ilmu komputer dan informatika*,1(2),pp.109–132.availableat: <https://doi.org/10.54082/jiki.13>.
- [3]. Ali, a. (2019) 'klasterisasi data rekam medis pasien menggunakan metode k- means clustering di rumah sakit anwar medika balong bendo sidoarjo', *matrik : jurnal manajemen, teknik informatika dan rekayasa komputer*, 19(1), pp. 186–195. Available at: <https://doi.org/10.30812/matrik.v19i1.529>.
- [4]. Harjanto, t.d., vatresia, a. And faurina, r. (2021) *analisis penetapan skala prioritas penanganan balita stunting menggunakan metode dbscan clustering*, *jurnal rekursif*. Available at: <http://ejournal.unib.ac.id/index.php/rekursif/30>.
- [5]. Deni, r. And setiawan, p. (2018) *penerapan metode k-means clustering untuk analisis potensi produksi komoditi buah pada kabupaten lumajang*, *jurnal mahasiswa teknikinformatika*.
- [6]. Maulia, s., serasi ginting, b. And sihombing, a. (2021) 'implementasi data mining pengelompokan jenis penyakit pasien menggunakan metode clustering (studi kasus : puskesmas sambirejo)', *jik*, 5(1).
- [7]. May, p. And abrianto, c. (2018) *penerapan metode k-means clustering untuk pengelompokan pasien penyakit liver*, *jurnal mahasiswa teknik informatika*. Available at: <http://archive.Ics.uci.edu/ml/>.
- [8]. Ordila, r. *Et al.* (2020). 'penerapan data mining untuk pengelompokan data rekam medis pasien berdasarkan jenis penyakit dengan algoritma
- [9]. Rahmawati, l., widya sihwi, s. And suryani, e. (2016) 'analisa clustering menggunakan metode k-means dan hierarchical clustering (studi kasus : dokumen skripsi jurusan kimia, fmipa, universitas sebelas maret)', *jurnal teknologi & informasi itsmart*, 3(2), p. 66. Available at: <https://doi.org/10.20961/its.v3i2.654>.
- [10]. Adani, n.f. *Et al.* (no date) 'implementasi data mining untuk pengelompokan data penjualan berdasarkan pola pembelian menggunakan algoritma k- means clustering pada toko syihan stmik triguna dharma ** program
- [11]. Clustering (studi kasus : poli klinik pt.inecda)', *jurnal ilmu komputer*, 9(2),pp.148–153.availableat: <https://doi.org/10.33060/jik/2020/vol9.iss2.181>.