



Pengklasterisasian Data Penyakit Hipertensi dengan Menggunakan Metode K-Means

Fitria Rahmadayanti*, Inda Anggraini, Tri Susanti

Teknik Informatika, Institut Teknologi Pagar Alam, Pagar Alam

Jl. Masik Siagim No.75, Karang Dalo, Kec. Dempo Tengah, Kota Pagar Alam, Sumatera Selatan, Indonesia

Email: ^{1*}ria.ria.rr71@gmail.com, ²indaanggraini@gmail.com, ³Trisusanti@gmail.com

Email Penulis Korespondensi: ria.ria.rr71@gmail.com

Submitted: 11/01/2023; Accepted: 31/01/2023; Published: 31/01/2023

Abstrak-Meningkatnya jumlah dan ragam penyakit yang diderita masyarakat akibat perubahan gaya hidup yang dipengaruhi oleh kemajuan zaman. Pendataan penyakit secara berkala akan meningkatkan penumpukkan data. Hal tersebut sering terjadi kesalahan dalam proses pencarian data sehingga membutuhkan waktu yang cukup lama dalam pencarian data tersebut. Penelitian ini berfokus pada pendataan penyakit hipertensi dan bertujuan untuk melakukan klusterisasi terhadap data tersebut. Metode yang digunakan dalam penelitian ini adalah CRISP-DM dengan proses business understanding, data understanding, data preparation, modeling, evaluation dan deployment. Algoritma yang digunakan dalam klusterisasi ini adalah algoritma K-Means. Hasil dari penelitian ini dapat memberikan informasi mengenai hasil 2 cluster, yaitu cluster 0 Normal dan cluster 1 Hipertensi.

Kata Kunci: Penyakit Hipertensi; Data Mining; Klusterisasi; CRISP-DM; K-Means

Abstract-The increasing number and variety of diseases suffered by the community due to lifestyle changes that are influenced by the progress of the times. Periodic disease data collection will increase the accumulation of data. This is often an error in the data search process so that it takes a long time to search the data. This study focuses on data collection on hypertension and aims to cluster the data. The method used in this study is CRISP-DM with a business understanding process, data understanding, data preparation, modeling, evaluation and deployment. The algorithm used in this clustering is the K-Means algorithm. The results of this study resulted in 2 clusters, namely cluster 0 Normal and cluster 1 Hypertension. The results of this study can provide information about the results of 2 clusters, namely cluster 0 Normal and cluster 1 Hypertension

Keywords: Hypertension Disease; Data Mining; Clustering; CRISP-DM; K-Means

1. PENDAHULUAN

Meningkatnya jumlah dan ragam penyakit yang diderita masyarakat akibat perubahan gaya hidup yang dipengaruhi oleh kemajuan zaman[1]. Pendataan penyakit secara berkala akan meningkatkan penumpukkan data. Hal tersebut sering terjadi kesalahan dalam proses pencarian data sehingga membutuhkan waktu yang cukup lama dalam pencarian data tersebut. Penelitian ini berfokus pada pendataan penyakit hipertensi dan bertujuan untuk melakukan klusterisasi terhadap data tersebut. Penyakit hipertensi merupakan salah satu jenis penyakit yang menyerang organ dalam tubuh yang dapat membahayakan hidup seseorang, Hipertensi juga suatu gangguan pada sistem peredaran darah, yang cukup banyak mengganggu kesehatan masyarakat[2][3].

Pengelolaan data penyakit pasien selama ini masih dengan manual seperti, pasien datang ke administrasi, pihak administrasi akan mengarahkan pasien ke poli, pihak poli mencatat data pasien, data pasien akan dicatat ke dalam blangko yang diberikan oleh pihak pengolah data penyakit pasien, blangko tersebut dikembalikan ke pengolah data penyakit pasien, Blangko tersebut dikembalikan ke pihak pengelolaan data penyakit pasien, data penyakit pasien diketik menggunakan Microsoft Excel di cetak dan diarsipkan di dalam map odner. Dalam pencarian data penyakit pasien harus membuka arsip dan mencari satu per satu jadi butuh waktu lama dan kemungkinan kesalahan yang terjadi sangat besar dalam pencarian data. Walau sudah memiliki fasilitas komputer, namun belum ada sistem yang dapat menunjang dalam pengelolaan data, khususnya pengelolaan data penyakit pasien. Dengan kata lain, tujuan dari penelitian ini adalah mewujudkan pengelompokan data penyakit pasien sehingga diperoleh hasil klusterisasi Hipertensi.

Menurut Larose data mining merupakan sebuah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan machine learning yang mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terakut dari berbagai database yang besar[4][5]

Clustering merupakan suatu metode pengelompokan berdasarkan ukuran kedekatan(kemiripan)[6][7]. Clustering berbeda dengan classification, bahwa tidak ada variabel target clustering. Pengelompokan tidak mencoba untuk mengklasifikasikan, memperkirakan, atau memprediksi nilai variabel sasaran. Sebaliknya, algoritma klusterisasi berusaha untuk mencari bagian dari seluruh dataset menjadi subkelompok yang relatif homogen atau kelompok, di mana kesamaan data dalam cluster dimaksimalkan, dan kesamaan data yang di luar kluster diminimalkan[8]. Algoritma K-Means adalah salah satu algoritma dalam data mining yang bisa digunakan untuk melakukan pengelompokan/clustering suatu data, Adapun tujuan dari data clustering ini adalah untuk meminimalisasikan objective function yang diset dalam proses clustering, yang pada umumnya berusaha meminimalisasikan variasi di dalam suatu cluster dan memaksimalkan variasi antar cluster [4][6].

Penelitian dilakukan merujuk pada penelitian terdahulu yakni, menghasilkan algoritma KMeans diimplementasikan untuk mendiagnosis gagal ginjal kronis menggunakan kelompok KMeans. Hasil perhitungan kelompok KMeans digunakan pada saat diagnosis, kumpulan data yang digunakan selama pengujian menyediakan dua kelompok, masing-masing terdiri dari pasien dengan gejala dan penyebab dengan diagnosis berat, dan pasien dapat terus menerima pengobatan[9].

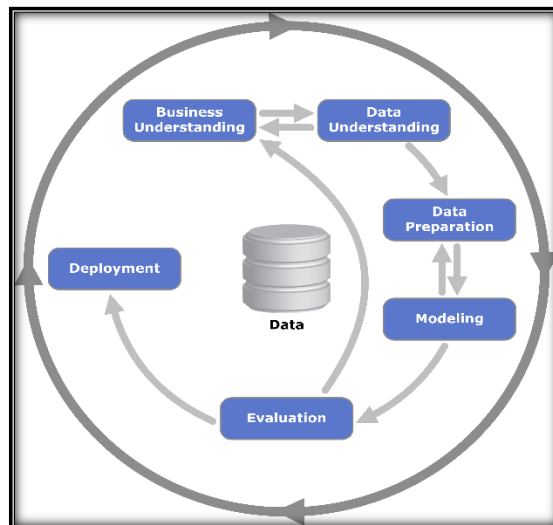
Penelitian dilakukan merujuk pada penelitian terdahulu yang kedua yakni, Penelitian ini menggunakan teknik data mining untuk mencari informasi baru dari dataset rekam medis mata. Algoritma KMeans dipilih untuk tujuan ini. Hasil pengujian untuk penelitian ini menggunakan perbandingan antara perbedaan dalam kelas (BVC) dan variasi dalam kelas (WVC)[10].

Dalam penelitian ini menggunakan metode CRIPS DM dengan tahapan yaitu Business Understanding (Pemahaman Bisnis), Data Understanding (Pemahaman Data), Data Preparation (Pengolahan Data), Modeling (Pemodelan), Evaluation (Evaluasi) dan Deployment (Penyebaran)[11][12][13].

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Metode penelitian yang digunakan sesuai dengan alur pada model CRIPS-DM menyediakan standart proses data mining sebagai strategi pemecahan masalah secara umum dari bisnis atau unit penelitian[14][15].



Gambar 1. Metode CRIPS-DM[16]

Penelitian ini menerapkan metode CRIPS-DM dengan tahapan pada Gambar 1 merupakan Siklus hidup dalam CRIPS-DM (Cross Industry Standard Process for Data Mining) ada enam fase yaitu [9][15][17]:

a) Business Understanding (Pemahaman Bisnis)

Fase pemahaman bisnis ini dilakukan memahami permasalahan dalam topik penelitian ini yakni, membantu pengelompokan data penyakit hipertensi menggunakan algoritma K-Means.

b) Data Understanding (Pemahaman Data)

Fase pemahaman data yakni, dalam penelitian data yang dikumpulkan adalah data penyakit pasien pada UPTD Puskesmas Bumiagung.

Data Preparation (Pengolahan Data)

Fase pengolahan data yaitu melakukan perhitungan yang baik. Pada tahap pengolahan data ini dilakukan proses cleaning data yang digunakan untuk menghilangkan informasi atau data yang tidak dibutuhkan dalam proses pengolahan data[15].

c) Modeling (Pemodelan)

Fase pemodelan yakni, telah dipilih data yang sesuai dengan jenis data yang akan digunakan pada proses mining. Teknik data mining yang digunakan adalah teknik Clustering dan menggunakan Algoritma K-Means. Diharapkan dengan menggunakan teknik dan algoritma tersebut dapat menentukan kelas unggulan dengan baik[15].

d) Evaluation (Evaluasi)

Tahapan evaluasi dalam penelitian ini adalah 1) Mengevaluasi hasil; 2) Meninjau Proses; 3) Menentukan Langkah Selanjutnya[18][17]

e) Deployment (Penyebaran)

Tahapan penyebaran yakni, hasil penelitian ini dapat menjadi acuan untuk menentukan Langkah guna mengatasi permasalahan dalam penumpukkan data penyakit pasien.

f) K-Means

Algoritma K-Means [19] [20] merupakan algoritma yang mempartisi data ke dalam cluster – cluster sehingga data yang memiliki kemiripan berada pada satu cluster yang sama dan data yang memiliki ketidaksamaan berada pada cluster yang lain. Langkah – Langkah melakukan Klastering dengan metode K-Means [15]:

$$D(i, j) = \sqrt{(x_{1i} - x_{1j})^2 + (x_{2i} - x_{2j})^2 + (\dots \dots \dots)^2 + (x_{ki} - x_{kj})^2} \tag{1}$$

dimana[15]:

D(i, j) = Jarak data ke i ke pusat cluster j

Xki = Data ke i pada atribut data ke k

Xkj = Titik pusat ke j pada atribut ke k

Penjelasannya[15] :

- a) Tentukan jumlah cluster.
- b) Inisialisasi pusat cluster (centroid) secara random.
- c) Alokasikan semua data ke cluster terdekat. Demikian juga kedekatan suatu data ke cluster tertentu ditentukan jarak antara data dengan pusat cluster.
- d) Hitung kembali pusat cluster dengan keanggotaan cluster yang sekarang. Pusat cluster adalah rata-rata dari semua data/ objek dalam cluster tertentu.
- e) Tugaskan lagi setiap objek memakai pusat cluster yang baru. Jika pusat cluster tidak berubah lagi maka proses clustering selesai. Atau, kembali ke langkah nomor 3 sampai pusat cluster tidak berubah lagi.

3. HASIL DAN PEMBAHASAN

Penelitian ini menggunakan algoritma K-Means dalam penerapan metode data mining. Adapun untuk mengetahui hasil klasterisasi mengenai data penyakit hipertensi. Dalam penelitian ini mempunyai beberapa tahapan adalah sebagai berikut:

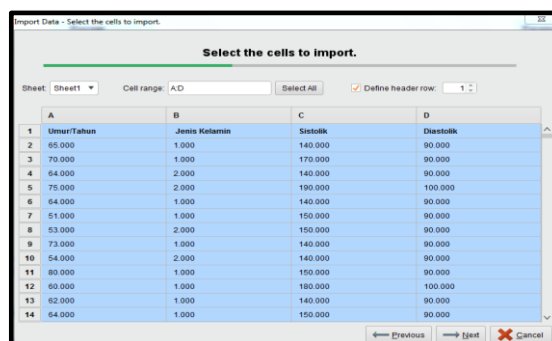
3.1. Dataset

Pada penelitian ini dataset yang digunakan mempunyai beberapa atribut yaitu umur/tahun, jenis kelamin, sistolik dan diastolik. Berikut tabel dataset yang digunakan.

Tabel 1. Dataset

No	Umur/Tahun	Jenis Kelamin	Sistolik	Diastolik
1	65	1	140	90
2	70	1	170	90
3	64	2	140	90
4	75	2	190	100
5	64	1	140	90
..
..
..

Setelah mendapatkan dataset Langkah selanjutnya adalah proses Import data dengan software Rapid Minner.



Gambar 2. Import Data

3.2 Tampilan Performance Rapid Miner

Gambar di bawah ini merupakan tampilan dataset yang telah di Importkan pada Rapid Minner, kemudian dataset akan dikelola dengan Algoritma K-Means dengan jumlah klaster yang terdiri dari dua klaster.

```

PerformanceVector

PerformanceVector:
Avg. within centroid distance: -159.206
Avg. within centroid distance_cluster_0: -134.700
Avg. within centroid distance_cluster_1: -231.275
Davies Bouldin: -0.718
    
```

Gambar 3. Halaman Performance Rapid Miner

3.3. Hasil Clustering Data Dengan K-Means

Halaman ini merupakan halaman hasil clustering dari proses pengImportan data sebelumnya ketika kita melakukan run as pada halaman proses clustering maka tampil halaman hasil clustering seperti pada gambar di bawah ini, dimana jumlah klaster yang terbentuk menjadi dua klaster yaitu cluster 0 dan cluster 1. Cluster 0 kategori normal dan cluster 1 kategori hipertensi.

Row No.	id	cluster	Umur/Tahun	Jenis Kela...	Sistolik	Diastolik
1	1	cluster_0	65	1	140	90
2	2	cluster_1	70	1	170	90
3	3	cluster_0	64	2	140	90
4	4	cluster_1	75	2	190	100
5	5	cluster_0	64	1	140	90
6	6	cluster_0	51	1	150	90
7	7	cluster_0	53	2	150	90
8	8	cluster_0	73	1	140	90
9	9	cluster_0	54	2	140	90
10	10	cluster_0	80	1	150	90
11	11	cluster_1	60	1	180	100
12	12	cluster_0	62	1	140	90
13	13	cluster_0	64	1	150	90
14	14	cluster_0	51	2	140	90
15	15	cluster_0	61	1	140	90
16	16	cluster_0	63	2	150	90

Gambar 4. Hasil Clustering Data Dengan K-Means

Aturan yang digunakan untuk mengelompokkan/cluster data berdasarkan hasil perhitungan jarak adalah :
 1) Jika $C0 < C1$ maka cluster 0; 2) Jika $C0 > C1$ maka cluster 1. Di dapatkan pola jika umur lebih dari sama dengan 50 dan sistolik lebih dari 140 dan diastolic lebih dari 100 maka cluster 1 atau hipertensi. Jika tidak maka normal.

4. KESIMPULAN

Penerapan Metode K-Means untuk pengelompokan data penyakit Hipertensi menyimpulkan bahwa algoritma ini mampu membuat pengelompokan data penyakit hipertensi. Dengan kata lain, melalui aplikasi rapidminner sama dengan hasil yang di terapkan dengan jumlah cluster yang terdiri dari dua cluster yang mana cluster di mulai dari cluster 0, dan cluster 1. Di dapatkan pola jika umur lebih dari sama dengan 50 dan sistolik lebih dari 140 dan diastolic lebih dari 100 maka cluster 1 atau hipertensi. Jika tidak maka normal. Obyek yang diteliti dengan metode K-Means dalam mengkalsifikasi data penyakit hipertensi adalah belum pernah diteliti oleh peneliti lain sebelumnya. Penelitian ini memiliki kelemahan yaitu hanya menggunakan satu metode saja dalam penelitian. Karenanya untuk penelitian lebih lanjut diperlukan melakukan dengan gabungan dua atau lebih metode data mining, sehingga menghasilkan hasil data mining yang lebih berguna dan lebih unggul. Juga diperlukan melakukan penelitian pada kumpulan obyek data besar lainnya, sehingga memberikan pengetahuan yang luas terhadap implikasi dari data mining.

REFERENCES

[1] B. S. Santoso, M. Rahmah, T. Setiasari, and S. Puji, "Perkembangan dan masa depan telemedika di indonesia," Res. Gate, vol. 2, no. 100, p. 8, 2015, [Online]. Available: https://www.researchgate.net/profile/Budi_Santoso28/publication/281497363_PERKEMBANGAN_DAN_MASA_DEPAN_TELEMEDIKA_DI_INDONESIA/links/55eb90cf08ae3e1218469f85/PERKEMBANGAN-DAN-MASA-



DEPAN-TELEMEDIKA-DI-INDONESIA.pdf.

- [2] I. Print, N. Azwanti, E. Elisa, U. P. Batam, and J. R. S. Kuning, “InfoTekJar : Jurnal Nasional Informatika dan Teknologi Jaringan Analisis Pola Penyakit Hipertensi Menggunakan Algoritma C4 . 5,” vol. 2, 2019.
- [3] F. Risiko et al., “J Urnal,” vol. 5, no. 2, 2015.
- [4] P. M. Silitonga Irene Sri, “Klusterisasi Pola Penyebaran Penyakit Pasien Berdasarkan Usia Pasien Dengan Menggunakan K-Means Clustering,” J. TIMES, vol. VI, no. Vol 6, No 2 (2017), pp. 22–25, 2017, [Online]. Available: <http://ejournal.stmik-time.ac.id/index.php/jurnalTIMES/article/view/584>.
- [5] M. G. Sadewo, A. P. Windarto, and D. Hartama, “Penerapan Datamining Pada Populasi Daging Ayam Ras Pedaging Di Indonesia Berdasarkan Provinsi Menggunakan K-Means Clustering,” InfoTekJar (Jurnal Nas. Inform. dan Teknol. Jaringan), vol. 2, no. 1, pp. 60–67, 2017, doi: 10.30743/infotekjar.v2i1.164.
- [6] D. Tyas and P. Hapsari, “Penerapan algoritma k-means pada kualitas gizi bayi di indonesia,” Semin. Nas. Teknol. Inf. dan Multimед. 2017, pp. 79–84, 2017.
- [7] A. De Wibowo, M. Sidik, and M. Artiyasa, “Menerapkan K-Means Clustering untuk Segmentasi,” vol. 2, no. 3, pp. 57–61, 2020.
- [8] T. Kesehatan et al., “Dan Kota Di Jawa Tengah.”
- [9] H. Hervisari, S. Musdalifah, and I. W. Sudarsana, “Implementasi Algoritma K-Means Untuk Diagnosa Penyakit Gagal Ginjal Kronis,” J. Ilm. Mat. Dan Terap., vol. 11, no. 1, pp. 27–35, 2016, doi: 10.22487/2540766x.2014.v11.i1.7468.
- [10] A. M. Siregar, “Penerapan Algoritma K-Means Untuk Pengelompokan Daerah Rawan Bencana Di Indonesia,” Intern. (Information Syst. Journal), vol. 1, no. 2, pp. 1–10, 2019, doi: 10.32627/internal.v1i2.42.
- [11] M. Rafi Muttaqin, T. Iman Hermanto, M. Agus Sunandar, P. Studi Teknik Informatika, and S. Tinggi Teknologi Wastukancana, “Penerapan K-Means Clustering Dan Cross-Industry Standard Process for Data Mining (Crisp-Dm) Untuk Mengelompokan Penjualan Kue,” vol. 19, no. 1, pp. 38–53, 2022, [Online]. Available: <https://journal.unpak.ac.id/index.php/komputasi>.
- [12] G. Gunawan, “Data Mining Using Crisp-Dm Process Framework on Official Statistics: a Case Study of East Java Province,” J. Ekon. dan Pembang., vol. 29, no. 2, pp. 183–198, 2021, doi: 10.14203/jep.29.2.2021.183-198.
- [13] D. Astuti, “Penentuan Strategi Promosi Usaha Mikro Kecil Dan Menengah (UMKM) Menggunakan Metode CRISP-DM dengan Algoritma K-Means Clustering,” J. Informatics, Inf. Syst. Softw. Eng. Appl., vol. 1, no. 2, pp. 60–72, 2019, doi: 10.20895/inista.v1i2.71.
- [14] I. P. Astuti, “Prediksi Ketepatan Waktu Kelulusan Dengan Algoritma Data Mining C4.5,” Fountain Informatics J., vol. 2, no. 2, p. 5, 2017, doi: 10.21111/fij.v2i2.1067.
- [15] C. Satria and A. Anggrawan, “Aplikasi K-Means berbasis Web untuk Klasifikasi Kelas Unggulan,” MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput., vol. 21, no. 1, pp. 111–124, 2021, doi: 10.30812/matrik.v21i1.1473.
- [16] Y. Suhandi, I. Kurniati, and S. Norma, “Penerapan Metode Crisp-DM Dengan Algoritma K-Means Clustering Untuk Segmentasi Mahasiswa Berdasarkan Kualitas Akademik,” J. Teknol. Inform. dan Komput., vol. 6, no. 2, pp. 12–20, 2020, doi: 10.37012/jtik.v6i2.299.
- [17] A. P. Fadillah, “Penerapan Metode CRISP-DM untuk Prediksi Kelulusan Studi Mahasiswa Menempuh Mata Kuliah (Studi Kasus Universitas XYZ),” J. Tek. Inform. dan Sist. Inf., vol. 1, no. 3, pp. 260–270, 2015, doi: 10.28932/jutisi.v1i3.406.
- [18] A. Rohman and A. Rufiyanto, “Komparasi Model Decision Tree Untuk Prediksi Kelulusan Mahasiswa Di Universitas Pandanaran,” Neo Tek., vol. 6, no. 1, pp. 8–12, 2020, doi: 10.37760/neoteknika.v6i1.1427.
- [19] mohamad jajuli nurul rohmawati, sofi defiyanti, “Implementasi Algoritma K-Means Dalam Pengklasteran Mahasiswa Pelamar Beasiswa,” Jitter 2015, vol. I, no. 2, pp. 62–68, 2015.
- [20] H. Priyatman, F. Sajid, and D. Haldivany, “Klasterisasi Menggunakan Algoritma K-Means Clustering untuk Memprediksi Waktu Kelulusan Mahasiswa,” J. Edukasi dan Penelit. Inform., vol. 5, no. 1, p. 62, 2019, doi: 10.26418/jp.v5i1.29611.