

ANALISIS *DATA MINING* UNTUK KLASTERISASI DATA REKAM MEDIS MENGUNAKAN ALGORITMA K-MEANS PADA RUMAH SAKIT SYLVANI BINJAI

Wahyu Surya Nanda¹, Akim M.H. Pardede², Magdalena Simanjuntak³
^{1,2,3} Fakultas Teknik Informatika, STMIK Kaputama, Binjai, Indonesia

Article Info

Article history:

Received August 16, 2023
 Revised August 16, 2023
 Accepted August 31, 2023

Kata Kunci:

Rekam Medis,
 Clustering,
 K-Mean,

Keywords:

Medical records,
 Clustering,
 K-Means,

ABSTRAK

Rekam medis adalah rekaman riwayat dari pasien yang melakukan pengobatan di rumah sakit maupun klinik. RSUD Sylvani memiliki banyak pasien, setiap bulannya dan membuat data history pasien menumpuk di dalam data rekam medis, Namun tidak ada tindak lanjut manfaat dari data-data yang tersedia. Padahal data-data tersebut memiliki potensi yang besar untuk memberikan informasi baru dan wawasan yang berharga jika digali dengan data mining menggunakan metode k-means clustering. Jumlah data yang uji sebanyak 893 data dan menghasilkan 4 kelompok dari variabel diagnosa penyakit, jenis kelamin dan alamat. Dimana kelompok 1 berjumlah 268 data dengan pusat diagnosa penyakit Hipertension dan jenis kelamin perempuan pada alamat Kebun Lada. Kelompok 2 berjumlah 289 data dengan pusat diagnosa penyakit Asma dan jenis kelamin perempuan pada alamat Pahlawan. Kelompok 3 berjumlah 185 data dengan pusat diagnosa penyakit GERD dan jenis kelamin laki-laki pada alamat Kebun Lada. kelompok 4 berjumlah 151 data dengan pusat diagnosa penyakit Pembesaran Prostat dan jenis kelamin laki-laki pada alamat Kebun Lada.

ABSTRACT

Medical record is a record of the history of patients who take treatment in hospitals or clinics. RSUD Sylvani has many patients, every month and makes patient history data accumulate in the medical record data, but there is no follow-up benefit from the available data. Even though these data have great potential to provide new information and valuable insights if explored with data mining using the k-means clustering method. The amount of data that was tested was 893 data and produced 4 groups from the variables of disease diagnosis, gender and address. Where group 1 totaled 268 data with a diagnosis center for hypertension and female gender at the Pepper Garden address. Group 2 totaled 289 data with a diagnosis center for Asthma and female gender at the Hero's address. Group 3 totaled 185 data with GERD disease diagnosis center and male gender at Kebun Pepper address. group 4 totaling 151 data with a diagnosis center for Prostate Enlargement disease and male sex at Kebun Pepper address.

This is an open access article under the [CC BY](https://creativecommons.org/licenses/by/4.0/) license.



Corresponding Author:

Wahyu Surya Nanda
Fakultas Teknik Informatika, STMIK Kaputama
Binjai, Indonesia
Email: wahyusurya844@gmail.com

1. PENDAHULUAN

Di era digital ini, semua serba efisien dengan penggunaan teknologi informasi. Rekam medis atau biasa disebut dalam bidang kesehatan adalah ICD (*International Clasification Diseases*) merupakan rekaman riwayat dari pasien yang melakukan pengobatan di rumah sakit maupun klinik. Bahasa medis yang dilakukan oleh dokter dalam melakukan diagnosa kemudian memberikan tindakan atas penyakit yang diterima pasien berupa bahasa kedokteran (rekam medis) yang selanjutnya hasil tersebut dikodekan oleh seorang ahli rekam medis menjadi kode-kode ICD. Kode ini adalah bahasa standar yang bisa digunakan oleh semua dokter meskipun bukan dokter spesialis untuk membacanya sesuai dengan aturan-aturan yang berlaku pada kode tersebut (ICD). Selain itu, pemanfaatan rekam medis yang mengikuti teknologi memberikan manfaat kepada dokter dan petugas kesehatan dalam mengakses informasi pasien yang pada akhirnya membantu dalam pengambilan keputusan klinis.

Pencatatan rekam medis adalah wajib bagi dokter-dokter rumah sakit yang melakukan tindakan medis kepada pasien sesuai dengan aturan sehingga tidak ada alasan bagi dokter untuk tidak membuat rekam medik tersebut. Begitu juga dengan pelayanan medis. Pelayanan medis menjadi salah satu hal yang harus ditingkatkan. Dengan meningkatnya pelayanan medis maka ilmu rekam medis sangat penting untuk dipelajari, sedangkan banyak diantara masyarakat masih belum memahami tentang rekam medis [1] [2]. Rekam medis merupakan berkas yang berisi data pasien, data pasien pada rumah sakit mengandung data yang cukup besar karena setiap pasien berkunjung atau mengecek kesehatan pada rumah sakit akan di jadikan rekam medis yang membuat data tersebut menumpuk yang menyebabkan berbagai masalah yang sering terjadi seperti tempat penyimpanan yang besar, hilangnya rekam medis, pengeluaran data yang dibutuhkan, dan lain-lain.

Rumah sakit umum Sylvani merupakan salah satu rumah sakit di Kota Binjai. RSU Sylvani memiliki banyak pasien, setiap bulannya pasien di rumah sakit Sylvani selalu berdatangan membuat data history pasien menumpuk di dalam data rekam medis, dan tidak ada tindak lanjut manfaat dari data data yang tersedia. Padahal data – data tersebut memiliki potensi yang besar untuk memberikan wawasan yang berharga kepada profesional kesehatan, peneliti, dan pembuat kebijakan berbasis bukti. Namun pengelolaan dan analisis data yang besar dan kompleks menjadi tantangan yang signifikan. Salah satu pendekatan yang muncul untuk mengatasi tantangan ini adalah data mining menggunakan metode *k-means clustering*. Selain itu dengan metode *k-means clustering* menggunakan data rekam medis pasien RSU Sylvani Binjai, penulis dapat melihat pola penyebaran penyakit yang ada di Kecamatan Binjai.

Data *mining* merupakan salah satu bagian atau proses utama dari *Knowledge Discovery in Database* (KDD) yang bentuk kegiatannya yaitu mengumpulkan dan menggunakan data masa lalu untuk menemukan keteraturan, pola atau hubungan dalam suatu set data yang lebih besar [3]. Data *mining* merupakan salah satu cabang ilmu komputer yang cukup baru yang banyak digunakan dan dipelajari oleh para ahli ilmu komputer dan programmer. Data *mining*

merupakan sebuah konsep yang diperuntuhkan untuk menemukan pengetahuan atau informasi berharga yang tersembunyi di dalam database besar atau data dalam jumlah yang sangat besar. Data *mining* memiliki beberapa teknik dalam menghasilkan informasi penting dari tumpukan dan teknik-teknik yang berbeda, seperti deskripsi, estimasi, prediksi, klasifikasi, pengklusteran dan asosiasi. Penerapan Data mining banyak sekali digunakan dalam berbagai bidang, seperti bidang pendidikan, pemerintahan, bisnis atau perdagangan dan kesehatan (Rumah Sakit). Teknik pengolahan data dalam penelitian ini yaitu menggunakan teknik data *mining* dengan menerapkan algoritma *K-Means Clustering*.

Penerapan metode *K-Means Clustering* telah banyak dilakukan oleh peneliti sebelumnya, diantaranya tentang klasterisasi data rekam medis pasien dengan menggunakan *K-Means Clustering* di rumah sakit Anwar Medika Bolong Bendo Sidoarjo. Kesimpulan yang diperoleh yaitu menghasilkan 4 klaster yang terdiri dari kecamatan, diagnosa penyakit, usia dan jenis kelamin. Pada klaster 1 menghasilkan banyak pasien terdiri dari pasien jenis kelamin perempuan sebanyak 79 (15%) orang, klaster 2 menghasilkan banyak pasien yang terdiri dari pasien berjenis kelamin laki-laki sebanyak 214 (50%) orang, Pada klaster 3 menghasilkan banyak pasien terdiri dari pasien jenis kelamin perempuan sebanyak 89 (17%) orang, klaster 4 menghasilkan banyak pasien yang terdiri dari pasien berjenis kelamin laki-laki sebanyak 152 (28%) orang. Pengelompokan data rekam medis pasien ini menghasilkan informasi baru mengenai pola pengelompokan penyebaran penyakit disetiap kecamatan berdasarkan data rekam medis pasien dari Rumah Sakit Anwar Medika sebanyak 534 data dengan waktu penyelesaian sebanyak 0,06 detik [4].

Peneliti lain yaitu tentang klasterisasi data rekam medis pasien pengguna layanan BPJS kesehatan menggunakan metode *K-Means*. Kesimpulan yang dihasilkan yaitu bahwa proses *clustering* menggunakan algoritma *K-Means* telah berhasil melakukan pengelompokan data rekam medis pasien BPJS kedalam 3 *cluster* dengan kriteria yang berbeda. Penyakit dengan kode A09.9 (Diare/Disentri) mendominasi pada *cluster* 0, penyebaran jenis penyakit yang lebih beragam pada *cluster* 1 dan penyebaran jenis penyakit dengan kode K30 (Dyspepsia) lebih didominasi pada *cluster* 2 [5].

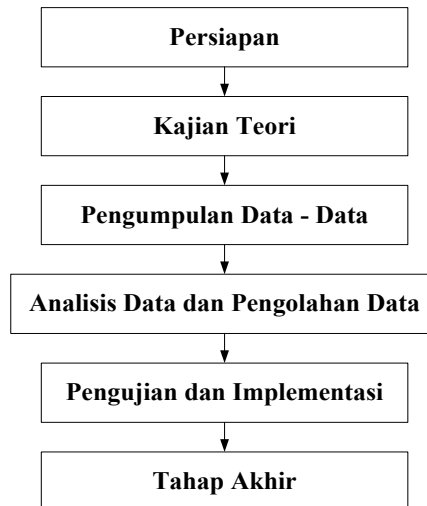
Peneliti lain yaitu tentang penerapan data *mining* untuk pengelompokan data rekam medis pasien berdasarkan jenis penyakit dengan Algoritma *K-Means*. Kesimpulan yang dihasilkan yaitu dengan adanya data *mining* metode Algoritma *K-Means*, dapat membantu mengelompokan data rekam medis pasien poli klinik PT. Inecda berdasarkan wilayah, jenis kelamin dan umur. Jumlah pasien berdasarkan umur yang pertama adalah dewasa dengan jumlah 4912 pasien, yang kedua kategori anak-anak dengan jumlah 1262 pasien dan yang ketiga dengan kategori balita dengan jumlah 144 pasien. jumlah penyakit dengan pasien terbanyak adalah pertama ISPA dengan jumlah 1985 pasien dikarenakan lingkungan perumahan PT. Inecda yang merupakan perkebunan kelapa sawit dan PKS (Pabrik Kelapa Sawit) dan juga penyakit terbanyak dengan penyakit lain-lain seperti jatuh dari motor, kolestrol, kontrol kehamilan cek tensi dengan jumlah pasien 2142 pasien [6].

Peneliti lain yaitu tentang pemanfaatan rekam medis dalam menentukan *cluster* penyakit melalui data *mining* di RS Aisiyiah Siti Fatimah Tulangan Sidoarjo. Dengan kesimpulan proses *clustering software* RapidMiner dengan 5 *cluster* mendapatkan hasil C₀ didominasi diagnosa I11.0, C₁ didominasi diagnosa E11.8, C₂ didominasi diagnosa A15.3, C₃ didominasi diagnosa E05.9, C₄ didominasi diagnosa E11.8. Uji validitas cluster menggunakan metode Davies

Bouldin Index (DBI) dengan 5 *cluster* mendapatkan nilai 0.469 [7].

2. METODE

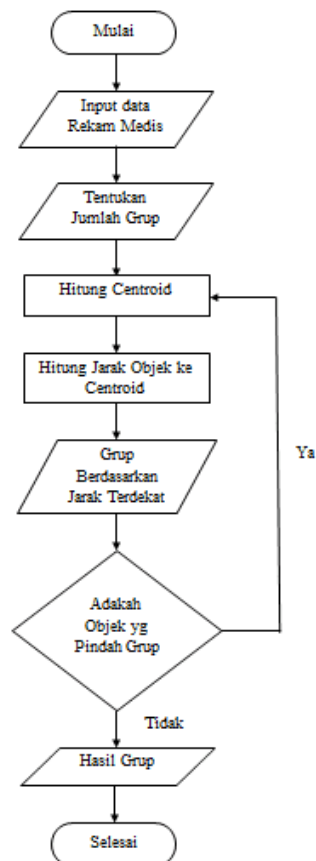
Dalam menyelesaikan sebuah masalah dalam penelitian, tentunya peneliti harus memiliki cara atau sebuah metode yang akan diterapkan dalam menyelesaikan masalah agar penelitian yang dilakukan dapat terselesaikan dengan baik dan sesuai dengan hasil yang diharapkan. Metode penelitian dilakukan untuk mencari sesuatu secara sistematis dengan menggunakan metode ilmiah serta sumber yang berlaku. Atas dasar metodologi penelitian yang digunakan pada penelitian ini, dapat dibuat suatu alur kegiatan metode kerja penelitian sebagai berikut:



Gambar 1. Alur Kerja Penelitian

2.1 Flowchart

Dalam penggunaan metode *clustering*, proses awal yang dilakukan untuk pembentukan *cluster* adalah mentransformasikan data ke dalam bentuk *numeric* dengan kode-kode yang telah ditentukan, lalu tentukan jumlah group (K), hitung *centroid*, hitung jarak objek ke *centroid* dan kemudian groupkan berdasarkan jarak terdekat, jika tidak ada objek yang pindah atau group maka iterasi selesai. Proses atau alur kerja sistem untuk pengklasteran rekam medis pasien pada RSUD Sylvani dengan menggunakan metode *clustering* dapat dilihat dalam bentuk *flowchart* seperti pada gambar dibawah ini:



Gambar 2. Flowchart Metode Clustering

3. HASIL DAN PEMBAHASAN

Dalam pembahasan antarmuka ini akan di jelaskan mengenai hasil perancangan program yang menggunakan GUI (*Graphical User Interface*) yang ada di *Software* MATLAB R2014a, dan dapat dilihat sebagai berikut :

1. Menu Home

Menu Utama / *Menu Home* disini menampilkan *interface* awal yang berisi *home*, *cluster*, dan *exit*. Ketika di klik *cluster* maka akan langsung masuk ke halaman *cluster*. Adapun tampilan halaman *menu home* sebagai berikut :



Gambar 3. Tampilan Menu Home

2. Cluster

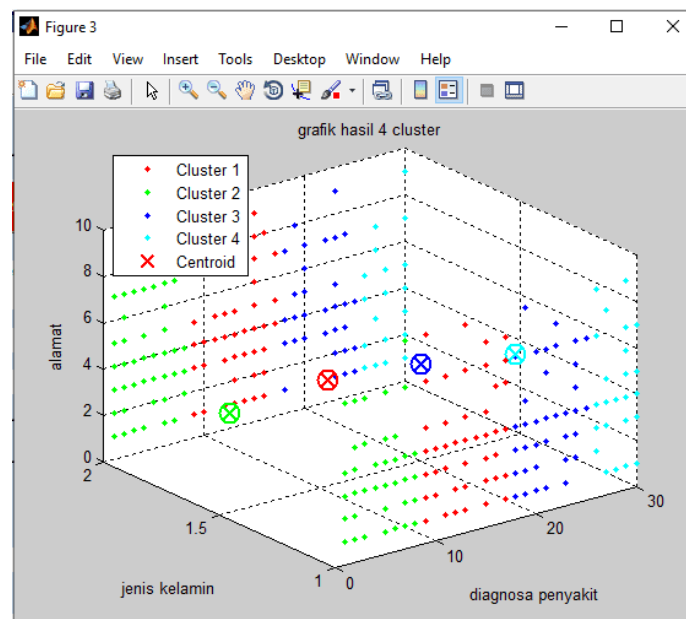
Menu cluster, disini akan terlihat keseluruhan proses data *mining* sampai pada kemunculan grafik dan keterangan *centroid* sebagai hasil dari perhitungan dengan metode *clustering* menggunakan algoritma *k-means*. Untuk lebih jelasnya sebagai berikut :

The screenshot shows a web application interface for clustering. It includes a 'Data' section with 'IMPORT DATA' and 'RESET DATA' buttons, a 'Jumlah Cluster' dropdown set to 4, and a 'PROSES CLUSTERING' button. On the right, a table shows the number of members in each of the 4 clusters. Below, a 'Hasil Centroid' table displays the calculated centroid coordinates for each cluster across three dimensions: Diagnosis Penyakit (X), Jenis Kelamin (Y), and Alamat / Kelurahan (Z).

Cluster	Centroid X	Centroid Y	Centroid Z
Cluster 1, Centroid	13.5037	1.6157	3.7351
Cluster 2, Centroid	3.4913	1.609	3.4879
Cluster 3, Centroid	21.7189	1.573	3.6649
Cluster 4, Centroid	28.6623	1.4636	3.7682

Gambar 4. Tampilan Menu Cluster

Berdasarkan data yang diolah, kemudian akan muncul hasil *clustering* pada masing masing *centroid* dan akan juga muncul grafik hasil *clustering*. Adapun hasil grafik *clustering* dapat dilihat pada gambar berikut :



Gambar 5. Tampilan Grafik hasil 4 Cluster

4. KESIMPULAN

Diketahui hasil dari proses 2 *cluster* menunjukkan pembagian data ke dalam dua kelompok berdasarkan jarak data terhadap *centroid* / pusat *cluster*. Dimana *cluster* 1 berjumlah 346 data dan *cluster* 2 berjumlah 547 data. Dari data tersebut dapat diketahui pada *cluster* 1 berjumlah 346 data dengan pusat diagnosa penyakit *Myalgia* (Nyeri Otot) dan jenis kelamin laki-laki pada alamat Kebun Lada dan pada *cluster* 2 berjumlah 547 data dengan pusat diagnosa penyakit

Dengue Hemorrhagic Fever (Demam Berdarah) dan jenis kelamin perempuan pada alamat Pahlawan. Kemudian diketahui hasil dari proses 3 *cluster* menunjukkan pembagian data ke dalam tiga kelompok berdasarkan jarak data terhadap *centroid* / pusat *cluster*. Dimana *cluster* 1 berjumlah 277 data, *cluster* 2 berjumlah 319 data dan *cluster* 3 berjumlah 297 data. Dari data tersebut diketahui pada *cluster* 1 berjumlah 277 data dengan pusat diagnosa penyakit *Hydronephrosis* dan jenis kelamin laki-laki pada alamat Kebun Lada. Sedangkan pada *cluster* 2 berjumlah 319 data dengan pusat diagnosa penyakit *Asthma* (Asma) dan jenis kelamin laki-laki pada alamat Pahlawan. Dan pada *cluster* 3 berjumlah 297 data dengan pusat diagnosa penyakit *Conghestive Heart Failure* (Gagal Jantung) dan jenis kelamin laki-laki pada alamat Kebun Lada. Berikutnya diketahui hasil dari proses 4 *cluster* menunjukkan pembagian data ke dalam empat kelompok berdasarkan jarak data terhadap *centroid* / pusat *cluster*. Dimana pada *cluster* 1 berjumlah 268 data, *cluster* 2 berjumlah 289 data, *cluster* 3 berjumlah 185 data dan *cluster* 4 berjumlah 151 data. Pada data tersebut *cluster* 1 berjumlah 268 data dengan pusat diagnosa penyakit *Hipertension* (Tekanan Darah Tinggi) dan jenis kelamin perempuan pada alamat Kebun Lada. Pada *cluster* 2 berjumlah 289 data dengan pusat diagnosa penyakit *Asthma* (Asma) dan jenis kelamin perempuan pada alamat Pahlawan. Pada *cluster* 3 berjumlah 185 data dengan pusat diagnosa penyakit *Gastro Esophageal Reflux Dases* (GERD) dan jenis kelamin laki-laki pada alamat Kebun Lada dan pada *cluster* 4 berjumlah 151 data dengan pusat diagnosa penyakit *Enlarged Prostat* (Pembesaran Prostat) dan jenis kelamin laki-laki pada alamat Kebun Lada.

REFERENSI

- [1] S. M. Putri, U. Hayati, and R. Dzulkarnaen, "Perancangan Arsitektur Electronic Medical Record (EMR) Menggunakan Metode Enterprise Architecture Planning (EAP) Arsitektur Enterprise," *Journal of Information Technology*, vol. 2, no. 1, pp. 25–30, Mar. 2020, doi : 10.47292/joint.v2i1.21
- [2] A. W. Prastiwi, "Pengembangan Sistem Informasi Rekam Medis dalam Pelayanan Kesehatan (Studi Kasus: Praktik Dokter dr. M. Akbar Arifin dan Bidan Umi M. Arifin Kulon Progo)." Universitas Teknologi Yogyakarta, 2021
- [3] Erwin Widiaworo, *Inovasi Pembelajaran: berbasis life skill & entrepreneurship*, Yogyakarta: Ar-Ruzz Media, 2017
- [4] Amir Tjolleng, *Pengantar Pemrograman Matlab*, Jakarta: PT. Alex Media Komputindo, 2017
- [5] Jeri Wanda, *Klasterisasi Rekam Medis Pasien Pengguna Layanan BPJS Kesehatan Menggunakan Metode K-Means*, *Jurnal Informasi dan Teknologi*, vol. 2, no. 4, pp. 119-125, doi: 10.37034/jidt.v2i4.73
- [6] Rian Ordila, *Penerapan Data Mining Untuk Pengelompokan Data Rekam Medis Pasien Berdasarkan Jenis Penyakit Dengan Algoritma Clustering (Studi Kasus : Poli Klinik PT.Inecda)*, *Jurnal Ilmu Komputer STMIK Hang Tuah Pekanbaru*, vol. 9, no. 2, pp. 148-153, Okt. 2020, doi: 10.33060/JIK/2020/Vol9.Iss2.181
- [7] Brema Arya Yudano, 2022, pemanfaatan rekam medis dalam menentukan cluster penyakit melalui data mining di RS Aisiyiah Siti Fatimah Tulangan Sidoarjo, *Nursing, Medical and Science Journal*, vol. 2, no. 1, pp. 14-23, Dec. 2022, doi : 10.35584/carejournal.v2i1.98