

PENERAPAN ALGORITMA K-MEANS CLUSTERING UNTUK PENGLUMPOKAN PASIEN BERDASARKAN DIAGNOSIS DI RUMAH SAKIT AL-RASYID PALEMBANG

Nurcholis¹, Dafid²

Universitas Multi Data Palembang, Indonesia^{1,2}

Email: nurcholis@mhs.mdp.ac.id¹, Dafid@mdp.ac.id²

Informasi

Abstract

Volume : 3
Nomor : 7
Bulan : Juli
Tahun : 2026
E-ISSN : 3062-9624

The advancement of information technology has compelled hospitals to not only provide accurate and timely medical services but also to manage patient data effectively. Al Rasyid Islamic Hospital Palembang faces challenges in allocating medical personnel efficiently due to underutilization of medical record data. This study aims to develop a patient clustering system based on diagnoses using the K-Means Clustering algorithm. Utilizing outpatient medical record data from 2022 to 2024, the system groups patients based on diagnosis, age, and gender. The development process follows the CRISP-DM methodology, which includes business understanding, data understanding, data preparation, modeling, evaluation, and deployment stages. Clustering results are presented through a web-based dashboard to support hospital management in medical personnel allocation planning. This research is expected to improve the efficiency of healthcare services through data-driven analysis.

Keyword: Data Mining, K-Means Clustering, Diagnosis

Abstrak

Kemajuan teknologi informasi telah mendorong rumah sakit untuk tidak hanya memberikan pelayanan medis yang cepat dan akurat, tetapi juga mengelola data pasien secara efektif. Rumah Sakit Islam Al Rasyid Palembang menghadapi tantangan dalam mendistribusikan tenaga medis secara tepat sasaran akibat kurangnya pemanfaatan data rekam medis. Penelitian ini bertujuan untuk mengembangkan sistem pengelompokan pasien berdasarkan diagnosis menggunakan algoritma K-Means Clustering. Dengan memanfaatkan data pasien rawat jalan dari tahun 2022 hingga 2024, sistem ini mengelompokkan pasien berdasarkan diagnosis, usia, dan jenis kelamin. Proses pengembangan menggunakan metodologi CRISP-DM yang meliputi tahapan pemahaman bisnis, pemahaman data, persiapan data, pemodelan, evaluasi, dan implementasi. Hasil clustering ditampilkan melalui dashboard berbasis web untuk mendukung pengambilan keputusan manajemen rumah sakit dalam perencanaan distribusi tenaga medis. Penelitian ini diharapkan dapat meningkatkan efisiensi pelayanan kesehatan berbasis analisis data.

Kata Kunci: Data Mining, K-Means Clustering, Diagnosis

A. PENDAHULUAN

Kemajuan teknologi informasi telah membawa perubahan signifikan dalam berbagai bidang, termasuk sektor kesehatan. Rumah sakit tidak hanya dituntut memberikan pelayanan medis yang cepat dan tepat, tetapi juga mampu mengelola data pasien secara efektif. Data rekam medis yang terus bertambah setiap hari menjadi sumber informasi penting yang dapat dimanfaatkan untuk mendukung pengambilan keputusan strategis.

Salah satu teknik yang dapat digunakan untuk mengolah data pasien adalah data mining. Data mining merupakan proses penggalian informasi dari kumpulan data dalam jumlah besar untuk menemukan pola tertentu yang berguna dalam proses analisis dan pengambilan keputusan. Dalam bidang kesehatan, penerapan data mining dapat membantu rumah sakit dalam mengidentifikasi pola penyakit, karakteristik pasien, serta kebutuhan pelayanan kesehatan.

Metode yang digunakan dalam penelitian ini adalah algoritma K-Means Clustering. Algoritma ini merupakan metode clustering non-hierarchical yang bertujuan mengelompokkan data berdasarkan tingkat kemiripan karakteristik tertentu. Pengelompokan pasien berdasarkan diagnosis diharapkan mampu membantu rumah sakit dalam menyusun strategi pelayanan yang lebih tepat sasaran, termasuk distribusi tenaga medis dan penyediaan fasilitas kesehatan.

Penelitian sebelumnya menunjukkan bahwa algoritma K-Means efektif digunakan dalam pengelompokan data pasien. Penelitian oleh Herlin Lutfiannisa dkk. (2024) berhasil mengelompokkan data penyakit pasien berdasarkan usia di Puskesmas Tlogomulyo menggunakan metode K-Means dengan hasil dua cluster utama, yaitu penyakit skabies dan ISPA. Selain itu, penelitian oleh Rozy dkk. (2020) menunjukkan bahwa K-Means mampu mengelompokkan penyakit pasien ke dalam kategori akut dan tidak akut secara efektif.

Berdasarkan permasalahan tersebut, penelitian ini mengangkat judul “Penerapan Algoritma K-Means Clustering untuk Pengelompokan Pasien Berdasarkan Diagnosis di Rumah Sakit Islam Ar-Rasyid Palembang”.

B. METODE PENELITIAN

Penelitian ini menggunakan metode CRISP-DM (Cross Industry Standard Process for Data Mining) yang terdiri dari enam tahapan utama, yaitu:

2.1 Fase Pemahaman Bisnis (*Business Understanding Phase*)

Tahap ini dilakukan untuk memahami permasalahan yang terjadi di Rumah Sakit Islam

Ar-Rasyid Palembang terkait pengelolaan data pasien dan distribusi tenaga medis. Tujuan penelitian ditetapkan untuk membangun sistem clustering pasien berdasarkan diagnosis penyakit.

2. 2 Fase Pemahaman Data (*Data Understanding Phase*)

Data yang digunakan berupa data pasien rawat jalan periode 2022–2024 dengan atribut usia, jenis kelamin, dan diagnosis penyakit. Tahap ini dilakukan untuk memahami struktur dan kualitas data sebelum diproses lebih lanjut.

2. 3 Fase Pengolahan Data (*Data Preparation Phase*)

Pada tahap ini dilakukan preprocessing data yang meliputi:

- Data cleaning untuk menghapus data kosong dan duplikat.
- Transformasi teks diagnosis menggunakan metode TF-IDF.
- Normalisasi data menggunakan metode Min-Max Scaling agar seluruh atribut memiliki rentang nilai yang sama.

Normalisasi data dilakukan menggunakan rumus berikut:

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}}$$

2. 4 Fase Pemodelan Data (*Modelling Phase*)

Tahap modeling dilakukan menggunakan algoritma K-Means Clustering. Penentuan jumlah cluster optimal dilakukan menggunakan metode Elbow dengan menghitung nilai Within Cluster Sum of Squares (WCSS).

Rumus WCSS:

$$WCSS = \sum_{i=1}^k \sum_{x \in C_i} \|x - \mu_i\|^2$$

Perhitungan jarak menggunakan Euclidean Distance:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

2. 5 Fase Evaluasi (*Evaluation Phase*)

valuasi hasil clustering dilakukan menggunakan Silhouette Coefficient dan Davies-Bouldin Index untuk mengetahui kualitas cluster yang terbentuk.

Rumus Silhouette Coefficient:

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))}$$

C. HASIL DAN PEMBAHASAN

Hasil implementasi algoritma K-Means menunjukkan bahwa data pasien dapat dikelompokkan berdasarkan karakteristik diagnosis, usia, dan jenis kelamin. Proses clustering menghasilkan beberapa kelompok pasien dengan karakteristik penyakit yang berbeda.

Cluster yang terbentuk memberikan informasi mengenai pola distribusi penyakit pasien sehingga pihak rumah sakit dapat:

1. Menentukan prioritas pelayanan kesehatan.
2. Mengatur distribusi tenaga medis secara lebih efektif.
3. Menyusun strategi penanganan penyakit berdasarkan kelompok pasien.

Penerapan metode TF-IDF membantu proses transformasi diagnosis penyakit ke bentuk numerik sehingga dapat diproses oleh algoritma K-Means. Selain itu, normalisasi data menggunakan Min-Max Scaling membuat proses clustering menjadi lebih optimal karena setiap atribut memiliki skala yang seimbang.

Hasil evaluasi menggunakan Silhouette Coefficient menunjukkan bahwa cluster yang terbentuk memiliki kualitas yang cukup baik dan dapat digunakan sebagai dasar pengambilan keputusan dalam pelayanan kesehatan.

Dashboard berbasis web yang dikembangkan mampu menampilkan hasil clustering secara visual dan interaktif sehingga mempermudah pihak rumah sakit dalam memahami pola pengelompokan pasien.

D. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa:

1. Algoritma K-Means Clustering berhasil diterapkan untuk mengelompokkan pasien berdasarkan diagnosis penyakit di Rumah Sakit Islam Ar-Rasyid Palembang.
2. Metode CRISP-DM membantu proses penelitian mulai dari pemahaman bisnis hingga implementasi sistem.
3. Penggunaan TF-IDF dan normalisasi Min-Max Scaling mampu meningkatkan kualitas data sebelum proses clustering.
4. Hasil clustering dapat membantu rumah sakit dalam perencanaan alokasi tenaga medis dan strategi pelayanan kesehatan.
5. Dashboard berbasis web memudahkan visualisasi hasil clustering sehingga informasi lebih mudah dipahami oleh pihak rumah sakit.

E. DAFTAR PUSTAKA

- Ali. (2019). Klasterisasi Data Rekam Medis Pasien Menggunakan Metode K-Means Clustering di Rumah Sakit Anwar Medika Balong Bendo Sidoarjo.
- Anindya Khrisna Wardhani. (2016). Penerapan Metode K-Means Clustering untuk

Pengelompokan Data Penyakit Pasien di Puskesmas Kajen Pekalongan.

Bayu Prasetyo, Agus Pranoto, & Primaswara Prasetya. (2023). Implementasi Data Mining Menggunakan Algoritma K-Means Clustering Penyakit Pasien Rawat Jalan.

Herlin Lutfiannisa, Maimunah, & Sukmasetya. (2024). Pengelompokan Data Penyakit Berdasarkan Usia Menggunakan K-Means Clustering.

Prastiwi & Mulyati. (2023). Penerapan Algoritma K-Means Clustering Untuk Klasterisasi Penyakit Pasien.

Rozy, Fitri Boy, & Affandi. (2020). Implementasi Data Mining K-Means Clustering Dalam Pengelompokan Data Penyakit Pasien.