

Clustering Pasien Rawat Inap Di RS USU Menggunakan Algoritma K-Means

Clustering of Inpatients at USU Hospital Using the K-Means Algorithm

Ilka Zufria ¹, Isna Damaiani Iskandar ²

^{1,2}Universitas Islam Negeri Sumatera Utara

E-mail: ¹ ilkazufria@uinsu.ac.id, ² isnaiskandar601@gmail.com

Abstrak

Penelitian ini bertujuan untuk mengidentifikasi pola perbedaan karakteristik pasien rawat inap di Rumah Sakit Universitas Sumatera Utara (RS USU) berdasarkan kelas Badan Penyelenggara Jaminan Sosial (BPJS) menggunakan pendekatan algoritma k-means. Data yang digunakan dalam penelitian ini melibatkan jumlah pasien dengan kelas BPJS yang berbeda selama periode tertentu. Metode analisis data menggunakan algoritma k-means untuk membentuk kelompok pasien yang serupa berdasarkan variabel yang relevan, seperti diagnosa, dokter yang bertanggung jawab, dan penanggung. Hasil analisis cluster diharapkan dapat memberikan pemahaman yang lebih mendalam tentang pola-pola karakteristik pasien rawat inap dengan kelas BPJS yang berbeda di RS USU. Temuan dari penelitian ini dapat memberikan informasi berharga bagi pihak rumah sakit dan Badan Penyelenggara Jaminan Sosial dalam pengelolaan sumber daya, perencanaan layanan kesehatan, serta perbaikan kualitas pelayanan. Selain itu, penelitian ini juga dapat memberikan dasar untuk pengembangan strategi personalisasi perawatan pasien yang lebih efektif sesuai dengan kebutuhan kelompok pasien tertentu. Dengan menggunakan pendekatan ini, diharapkan penelitian ini dapat memberikan kontribusi positif terhadap pengelolaan sumber daya kesehatan secara efisien, meningkatkan kualitas pelayanan kesehatan, dan memberikan landasan untuk penelitian lebih lanjut dalam konteks analisis cluster pada pasien rawat inap dengan kelompok asuransi kesehatan yang berbeda.

Kata kunci: *Clustering, K-Means, Rapid Miner*

Abstract

This study aims to identify patterns of differences in the characteristics of inpatients at the North Sumatra University Hospital (USU Hospital) based on Social Security Administering Body (BPJS) class using the k-means algorithm approach. The data used in this study involves the number of patients with different BPJS classes during a certain period. The data analysis method uses the k-means algorithm to form groups of similar patients based on relevant variables, such as diagnosis, responsible doctor, and insurer. It is hoped that the results of the cluster analysis will provide a deeper understanding of the characteristic patterns of inpatients with different BPJS classes at USU Hospital. The findings from this research can provide valuable information for hospitals and the Social Security Administration in managing resources, planning health services, and improving service quality. In addition, this research can also provide a basis for the development of more effective personalized patient care strategies according to the needs of specific patient groups. By using this approach, it is hoped that this research can make a positive contribution to the efficient management of health resources, improve the quality of health services, and provide a basis for further research in the context of cluster analysis of inpatients with different health insurance groups.

Keywords: *Clustering, K-Means, Rapid Miner*

1. PENDAHULUAN

Data mining juga dikenal sebagai *Knowledge Discovery in Database* (KDD), suatu proses otomatis yang mencari data di dalam ruang memori yang besar untuk mengidentifikasi pola, menggunakan teknik seperti klasifikasi, asosiasi, atau Clustering. Data mining adalah proses mengidentifikasi pola dalam kumpulan data yang besar, dengan penerapan di berbagai bidang. Ini melibatkan penggunaan teknik seperti klasifikasi dan penambangan web untuk mengekstrak informasi berguna dari dokumen dan aktivitas web. Prosesnya meliputi pengumpulan, penyiapan, dan penyimpanan data, serta menganalisisnya untuk menemukan pola dan korelasi. Penambangan data telah menjadi hal yang penting dalam berbagai bidang seperti perbankan, ritel, dan layanan kesehatan, namun juga menghadirkan tantangan dan permasalahan.[1]. Istilah penambangan data dan penemuan pengetahuan berbasis data (KDD) sering digunakan secara bergantian untuk merujuk pada pencarian informasi dalam dataset besar, meskipun keduanya memiliki konsep yang berbeda namun saling terkait. Proses KDD, yang terdiri dari lima tahap berurutan, dimulai dengan seleksi data dari kumpulan operasional sebelum menjalani tahap pra-pemrosesan, di mana data disucikan dan diverifikasi. Tahap transformasi melibatkan encoding untuk menyesuaikan data dengan kebutuhan penambangan data, yang merupakan inti dari proses tersebut. Metode penambangan data yang beragam digunakan untuk menemukan pola menarik dalam data yang dipilih, dengan pemilihan metode yang tepat tergantung pada tujuan keseluruhan dan proses KDD. Terakhir, interpretasi pola informasi diperlukan agar dapat dipahami dengan mudah oleh pihak yang berkepentingan, dengan tahap ini juga mencakup pemeriksaan kesesuaian hasil dengan fakta atau asumsi yang sudah ada. [2]

Suatu metode dalam bidang data mining dikenal sebagai teknik pengelompokan, yang mencakup clustering. *Clustering* melibatkan pengelompokan data atau objek ke dalam cluster-cluster, di mana setiap cluster berisi data yang memiliki kemiripan tinggi dan berbeda dengan objek di cluster lainnya. Dalam konteks ini, terdapat dua metode pengelompokan yang terkemuka, yakni partisi dan hierarkis.[3]. *Clustering* bertujuan untuk mengidentifikasi kelas-kelas taksonomi atau batryologi, atau melakukan analisis topologi terhadap data yang ada. Dalam konteks data mining, penting untuk dicatat bahwa *Clustering* bukanlah proses klarifikasi, karena pada proses klarifikasi, data dikelompokkan ke dalam kelas-kelas yang sudah diketahui sebelumnya.[4]. Konsep *centroid* dalam clustering sangat penting untuk menilai keanggotaan dan konektivitas lokal dalam *cluster*. Pemilihan awal *centroid* berdampak signifikan terhadap kinerja algoritma pengelompokan K-means. Algoritme X-means, yang memperkirakan jumlah *cluster* tanpa pengetahuan sebelumnya, dapat digunakan untuk menentukan *centroid*, dengan evaluasi *Indeks Davies-Bouldin* menyediakan metode untuk memilih jumlah *centroid* yang optimal. [5]

Rumah sakit adalah fasilitas kesehatan yang memberikan layanan kepada masyarakat. Perkembangan ilmu pengetahuan dalam bidang kesehatan, kemajuan teknologi, dan aspek sosial ekonomi masyarakat membentuk karakteristik institusi ini. Penting bagi rumah sakit untuk terus meningkatkan kualitas pelayanan kesehatan agar lebih mudah diakses oleh masyarakat. Tujuannya adalah mencapai derajat kesehatan masyarakat yang optimal melalui pemberian pelayanan kesehatan yang menyeluruh, meliputi promosi kesehatan, pencegahan penyakit, pengobatan dan rehabilitasi. Ini termasuk layanan rumah sakit, rawat jalan dan darurat.[6]. Rumah Sakit Universitas Sumatera Utara (RS USU) berada di Medan, Sumatera Utara, Indonesia. Ini adalah bagian dari Universitas Sumatera Utara dan menyediakan layanan medis serta praktik dan pengajaran mahasiswa kedokteran. Rumah sakit ini berafiliasi dengan Universitas Sumatera Utara (USU) dan berfungsi sebagai pusat kesehatan yang menyediakan berbagai layanan medis. RS USU berperan penting sebagai pemberi pelayanan kesehatan kepada masyarakat sekitar serta sebagai tempat pelatihan dan pembelajaran bagi mahasiswa kedokteran dan tenaga kesehatan lainnya di Universitas Sumatera Utara. Sebagai rumah sakit pendidikan, RS USU biasanya dilengkapi dengan peralatan dan teknologi kedokteran terkini untuk mendukung pengajaran dan penelitian di bidang pelayanan kesehatan. Selain itu, rumah sakit ini juga fokus dalam

memberikan pelayanan kesehatan yang komprehensif, baik rawat inap, rawat jalan, dan berbagai spesialisasi kedokteran.[7]

BPJS, atau Badan Penyelenggara Jaminan Sosial, adalah program asuransi kesehatan nasional pemerintah Indonesia. Undang-Undang Nomor 24 Tahun 2011 tentang Badan Penyelenggara Jaminan Sosial (BPJS) mewajibkan pembentukan badan hukum BPJS sebagai langkah untuk melaksanakan berbagai program jaminan sosial di tingkat nasional. Dengan kata lain, undang-undang ini menegaskan bahwa pembentukan badan hukum BPJS menjadi prasyarat penting dalam pelaksanaan program jaminan sosial yang bersifat nasional.[8]. Tujuannya untuk menjamin kesehatan seluruh warga negara dan menyediakan berbagai fasilitas kesehatan, antara lain Fasilitas Kesehatan Tingkat Pertama (FKTP) dan Fasilitas Kesehatan Rujukan Tingkat Lanjut (FKRTL).[9]. Namun, terdapat tantangan dalam penerapannya sehingga menimbulkan masalah kesehatan masyarakat. Meskipun terdapat tantangan-tantangan ini, program ini telah berhasil meningkatkan partisipasi, khususnya di kalangan masyarakat berpenghasilan rendah. [10]

Tabel 1. Data Pasien BPJS Rawat Inap 2022

No.	No.Registrasi	Penanggung	Nilai Penanggung
1	RI220101-0001	BPJS(KelasI)	3
2	RI220101-0003	BPJS(KelasIII)	1
3	RI220101-0004	BPJS(KelasIII)	1
4	RI220101-0005	BPJS(KelasIII)	1
5	RI220101-0006	BPJS(KelasIII)	1
6	RI220102-0001	BPJS(KelasII)	2
7	RI220102-0002	BPJS(KelasII)	2
8	RI220102-0008	BPJS(KelasII)	2
9	RI220102-0003	BPJS(KelasIII)	1
10	RI220102-0004	BPJS(KelasII)	2
11	RI220102-0005	BPJS(KelasII)	2
12	RI220102-0010	BPJS(KelasIII)	1
13	RI220102-0006	BPJS(KelasIII)	1
14	RI220102-0007	BPJS(KelasIII)	1
...	...	...	...

Berdasarkan data yang telah disajikan pada tabel 1, terlihat belum adanya pengelompokan terhadap penanggung biaya. Oleh karena itu, RS USU melakukan pengelompokan penanggung biaya berdasarkan tingkat Cluster yang telah ditentukan.

2. METODE PENELITIAN

1. Pengambilan Data

Data yang digunakan dalam penelitian ini adalah data yang diambil dari RS USU berupa data pasien rawat inap selama periode 2022 yang kemudian akan diolah menggunakan RapidMiner serta menggunakan pendekatan metode K-Means Clustering.

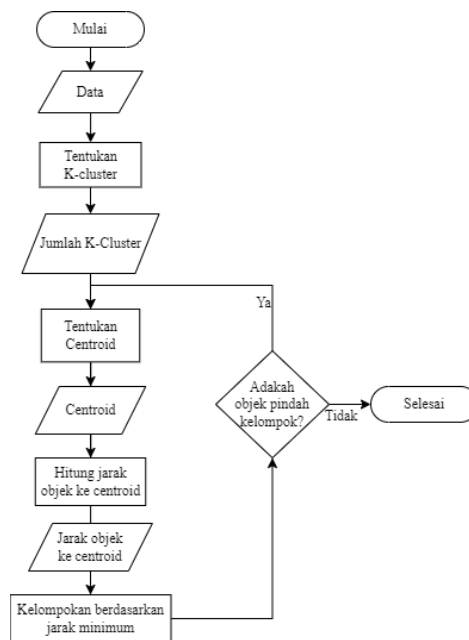
2. Metode Penelitian

Penelitian ini menggunakan Algoritma K-Means Clustering. K-Means Clustering adalah metode untuk mengelompokkan data non-hierarki (partisi), yang memiliki kemampuan untuk memisahkan data menjadi dua kelompok atau lebih. Metode ini membagi data ke dalam kelompok-kelompok. Kelompok yang sama terdiri dari data dengan atribut yang serupa, dan

kelompok lain terdiri dari data dengan atribut yang berbeda. [11]

3. Flowchart

Diagram alur atau *flowchart* adalah representasi visual dari langkah-langkah dan hubungan antar proses dalam suatu program, termasuk instruksi. Ilustrasi ini menggunakan simbol-simbol untuk menggambarkan secara spesifik setiap langkah atau proses [12]. *Flowchart* dari algoritma K-Means dapat dilihat pada gambar dibawah ini:



Gambar 1. Flowchart K-Means [13]

Adapun cara kerja algoritma K-Means dalam melakukan *Clustering* adalah sebagai berikut:

- Terangkan jumlah K cluster.
- Mulailah dengan menetapkan pusat cluster (*centroid*) K, yang dapat dipilih dari cluster yang sudah ada atau diperoleh secara acak.
- Untuk setiap data *record*, hitung dan tentukan jarak terdekat ke *centroid* awal menggunakan rumus berikut:

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

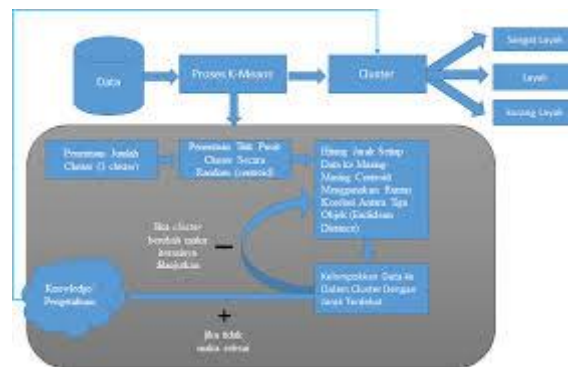
$D(i,j)$ = jarak dari data I kepusat cluster j

X_{ki} = data ke-I pada atribut data ke-k

X_{kj} = titik pusat ke-j pada atribut ke-k

- Hitung dan perbarui *centroid* untuk setiap grup cluster.
- Ulangi langkah a ke b, jika anggota cluster tidak mengalami perubahan (*konvergen*), maka proses dihentikan.[14]

Adapun langkah-langkah dalam pengolahan data dengan menggunakan metode K-Means *Clustering*:



Gambar 2. Alur kerja K-Means Clustering

Gambar adalah langkah-langkah dari algoritma K-Means Clustering. Pada tahap awal, data dikumpulkan dan melalui proses pembersihan serta transformasi data. Selanjutnya, terjadi pengolahan data menggunakan algoritma K-Means Clustering dengan menentukan jumlah Cluster secara acak, menentukan titik pusat Cluster, menghitung jarak setiap data, dan mengelompokkan data ke dalam Cluster. Proses ini menghasilkan pengetahuan berupa Clustering.[15]

3. HASIL DAN PEMBAHASAN

1. Centroid

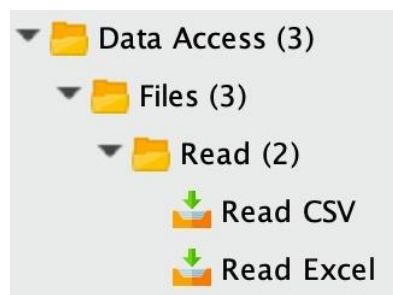
Pusat *centroid* dalam pengelompokan K-means adalah elemen penting yang berdampak signifikan terhadap kinerja algoritma [16]. Nilai mean atau sentral dari data muncul ketika penerapan algoritma K-Means. Penentuan nilai sentralitas ini diperoleh dengan mengikuti nilai minimum untuk *cluster* sedikit (C0), nilai mean untuk *cluster* sedang (C1), dan nilai maksimum untuk *cluster* banyak (C2).

Tabel 2. Data Centroid Awal

Data Cluster	Nilai Penanggung
(C0) Cluster sedikit	2
(C1) Cluster sedang	3.052
(C2) Cluster banyak	1

2. Implementasi RapidMiner

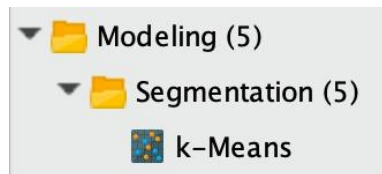
Mengacu pada informasi pada tabel 2 data *centroid* awal, *clustering* data pasien BPJS rawat inap 2022 diterapkan menggunakan RapidMiner. Penerapan metode K-Means pada RapidMiner akan dijelaskan pada gambar berikut.



Gambar 3. Operator Pembaca Data CSV

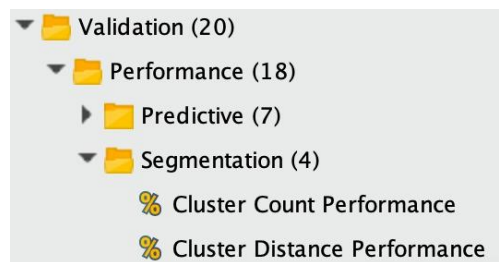
Operator RapidMiner pada gambar 3 digunakan untuk memasukkan data penelitian yang

bersumber dari file CSV.



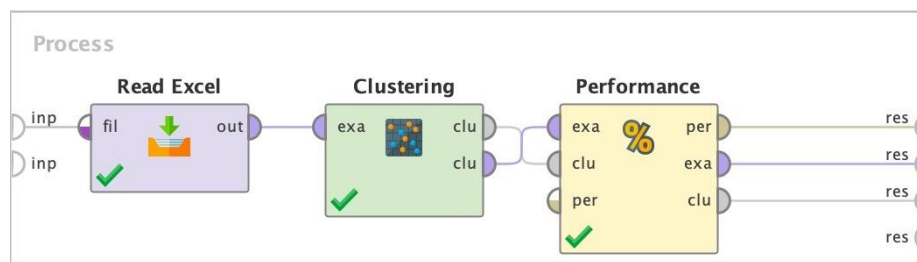
Gambar 4. Operator Clustering

Operator RapidMiner pada Gambar 5 digunakan untuk mengakses fungsi *Clustering* dari algoritma K-Means..



Gambar 5. Operator Centroid

Operator RapidMiner pada gambar 5 berfungsi untuk menciptakan pusat data pada dataset penelitian yang berasal dari file CSV.



Gambar 6. Urutan Operator

Pada gambar 6 merupakan urutan operator yang ada dalam proses yang akan dilakukan oleh kolom proses RapidMiner.

column index	attribute meta data information			
0	No.	☐ column ...	integer	attribute
1	No.Registrasi	☒ column ...	polynom...	attribute
2	Penanggung	☐ column ...	polynom...	attribute
3	Nilai Penangg	☒ column ...	integer	attribute
4	DPJP	☐ column ...	polynom...	attribute
5	ID DPJP	☐ column ...	integer	attribute

Gambar 7. Data Set Meta Data Information

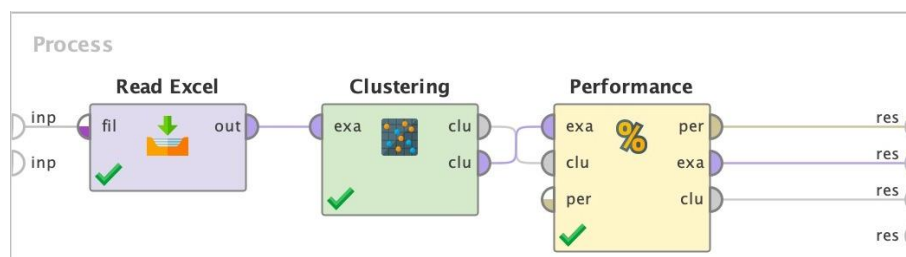
Pada gambar 7 merupakan atribut meta data *information* yang digunakan untuk pengaturan parameter List Data Set Meta Data *Information*.

Gambar 8. Parameter Clustering (K-Means)

Pada gambar 8, terdapat pengaturan parameter *Clustering* dimana nilai $k = 3$ dengan *Max Runs* = 10 dan *Measure Types* = *Mixed* dengan *Mixed Measure* = *MixedEuclidean*.

Gambar 9. Parameter Performance

Pada gambar 9, terdapat pengaturan parameter untuk pusat data di mana kolom *maximize* diaktifkan.



Gambar 10. Proses Pengelompokan Algoritma K-Means Dengan Nilai $K = 3$

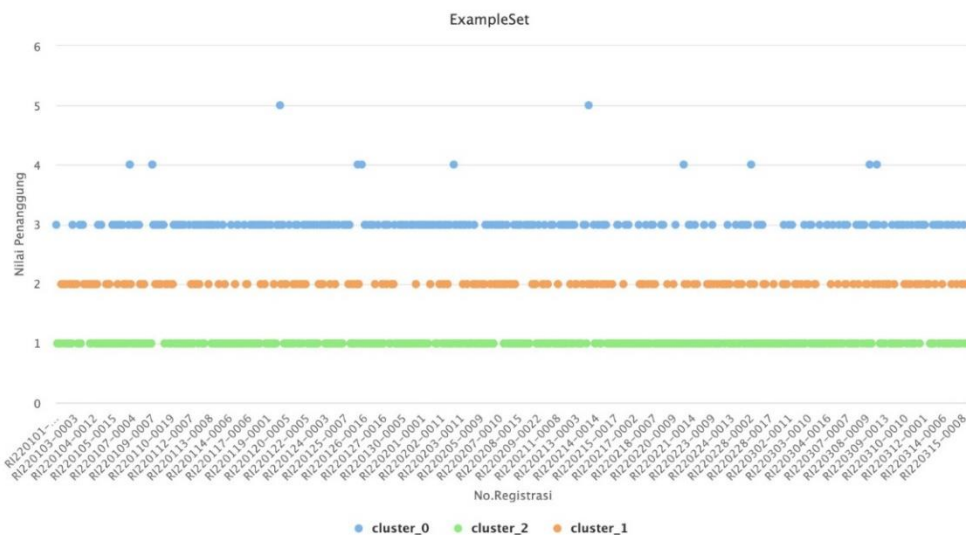
Pada gambar 10, terjadi penggabungan antara operasi *Read Excel* dengan K-Means dan *Output* yang akan dijalankan dengan nilai $K = 3$. Semua konfigurasi yang dibutuhkan disimpan dalam objek model. Dalam konteks ini, peneliti menggunakan data sampel mengenai Pasien BPJS Rawat Inap RS USU 2022 (sesuai dengan Tabel 1) yang memiliki atribut penanggung. Hasil akhir dari pengelompokan data dapat dilihat pada gambar di bawah ini.

Cluster Model

Cluster 0: 248 items
Cluster 1: 202 items
Cluster 2: 449 items
Total number of items: 899

Gambar 11. Pengelompokan Hasil menggunakan K-Means

Pada gambar 11, secara tegas menunjukkan keberadaan 3 *Cluster*, dimulai dari *Cluster* sedikit (C0), diikuti oleh *Cluster* sedang (C1), dan dilanjutkan oleh *Cluster* banyak (C2). *Cluster* sedikit (C0) memiliki 248 items, *Cluster* sedang (C1) memiliki 202 items, dan *Cluster* banyak (C2) memiliki 449 items dengan total keseluruhan 899 items. Dengan demikian, penulis berhasil meraih representasi visual dari hasil percobaan menggunakan RapidMiner seperti berikut:



Gambar 12. Grafik *Clustering* dalam Tampilan Plot

Pada gambar 12 dijelaskan bahwa diagram ringkasan di Tampilan Plot memainkan peran penting dalam memberikan representasi visual tentang bagaimana data disusun dan membantu dengan cepat memahami struktur data, mengidentifikasi pola pengelompokan, dan mengevaluasi performa model pengelompokan.

4. KESIMPULAN DAN SARAN

Clustering Pasien Rawat Inap Di RS USU dapat dilakukan dengan metode data *mining* menggunakan metode K-means menggunakan *tool* RapidMiner. Dengan menggunakan atribut evaluasi yaitu penanggung, pengelompokan dilakukan menjadi tiga *cluster*: (C0) *cluster* sedikit, (C1) *cluster* sedang, dan (C2) *cluster* banyak. Hasil agregat menunjukkan bahwa terdapat (C0) *cluster* sedikit memiliki 248 items, (C1) *cluster* sedang memiliki 202 items, dan (C2) *cluster* banyak 449 items. *Clustering* ini membantu dalam mengelompokkan Pasien Rawat Inap Di RS USU menjadi tiga kelompok, sehingga dapat lebih mudah dalam proses penempatan dan penanganan pasien.

Saran untuk penelitian selanjutnya ialah memperluas cakupan penelitian dengan melibatkan data pasien lebih lanjut atau menggunakan atribut tambahan dapat memberikan pemahaman yang lebih mendalam tentang karakteristik kelompok pasien. Penggunaan teknik data

mining lainnya juga dapat dieksplorasi untuk membandingkan hasil dan memperluas analisis serta meningkatkan validitas hasil clustering, penelitian dapat mempertimbangkan penggunaan metode evaluasi clustering tambahan, seperti indeks validitas internal atau eksternal.

DAFTAR PUSTAKA

- [1] S. Muawanah, U. Muzayanah, M. G. R. Pandin, M. D. S. Alam, and J. P. N. Trisnaningtyas, "Stress and Coping Strategies of Madrasah's Teachers on Applying Distance Learning During COVID-19 Pandemic in Indonesia," *Qubahan Acad. J.*, vol. 3, no. 4, pp. 206–218, 2023, doi: 10.48161/Issn.2709-8206.
- [2] F. Handayani, "Aplikasi Aplikasi Data Mining Menggunakan Algoritma K-Means Clustering untuk Mengelompokkan Mahasiswa Berdasarkan Gaya Belajar," *J. Teknol. dan Inf.*, vol. 12, no. 1, pp. 46–63, 2022, doi: 10.34010/jati.v12i1.6733.
- [3] A. A. Aldino, D. Darwis, A. T. Prastowo, and C. Sujana, "Implementation of K-Means Algorithm for Clustering Corn Planting Feasibility Area in South Lampung Regency," *J. Phys. Conf. Ser.*, vol. 1751, no. 1, 2021, doi: 10.1088/1742-6596/1751/1/012038.
- [4] 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.
- [5] M. Mughnyanti, S. Efendi, and M. Zarlis, "Analysis of determining centroid clustering x-means algorithm with davies-bouldin index evaluation," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 725, no. 1, 2020, doi: 10.1088/1757-899X/725/1/012128.
- [6] E. Ainun Novia, W. Isti Rahayu, and S. Fachri Pane, "Implementasi Algoritma K-Means Clustering Tingkat Kepentingan Tagihan Rumah Sakit Di Pt Pertamina (Persero)," *Jl. Sariasih*, vol. 54, p. 40151, 2020.
- [7] C. Pratiwi, S. Solin, and M. K. Zega, "Penilaian Pasien Rawat Inap Terhadap Pelayanan Makanan Instalasi Gizi RS. USU," *Pontianak Nutr. J.*, vol. 5, no. 1, pp. 171–176, 2022, [Online]. Available: <http://ejournal.poltekkes-pontianak.ac.id/index.php/PNJ/index>
- [8] R. Dewi and E. Israhadi, "Legal Aspects of BPJS as National Health Insurance," 2021, doi: 10.4108/eai.6-3-2021.2306412.
- [9] Pandi Rais, Shahrul bin Abd Shofi, Erina Sonia, and Sulis Setyoningsih, "the Islamic Law Review on Management of the Social Security Organizing Agency (Bpjs)," *Qawānīn J. Econ. Syaria Law*, vol. 4, no. 2, pp. 177–192, 2020, doi: 10.30762/q.v4i2.2469.
- [10] A. Purba, "Psychological Effect Of Bpjs Health Social Security participation In Perdagangan City Community," no. January 2014, pp. 3–7, 2020, doi: 10.4108/eai.14-3-2019.2291962.
- [11] I. Nuryani and D. Darwis, "Analisis clustering pada pengguna brand hp menggunakan metode k-means," vol. 1, no. 1, pp. 190–211, 2021.
- [12] Y. W. Syaifudin and R. A. Irawan, "Implementasi Analisis Clustering Dan Sentimen Data Twitter Pada Opini Wisata Pantai Menggunakan Metode K-Means," *J. Inform. Polinema*, vol. 4, no. 3, p. 189, 2018, doi: 10.33795/jip.v4i3.205.
- [13] L. G. Rady Putra and A. Anggrawan, "Pengelompokan Penerima Bantuan Sosial Masyarakat dengan Metode K-Means," *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 21, no. 1, pp. 205–214, 2021, doi: 10.30812/matrik.v21i1.1554.
- [14] N. F. Adani *et al.*, "Implementasi Data Mining Untuk Pengelompokan Data Penjualan Berdasarkan Pola Pembelian Menggunakan Algoritma K-Means Clustering Pada Toko Syihan," *J. Cyber Tech*, vol. x. No.x, no. x, pp. 1–11, 2019, [Online]. Available: <https://ojs.trigunadharma.ac.id/index.php/jct/article/view/4648%0Ahttps://ojs.trigunadharma.ac.id/index.php/jct/article/download/4648/791>
- [15] B. Harahap, "Penerapan Algoritma K-Means Untuk Menentukan Bahan Bangunan Laris (Studi Kasus Pada UD. Toko Bangunan YD Indarung)," *Reg. Dev. Ind. Heal. Sci. Technol.*

-
- Art Life*, pp. 394–403, 2019, [Online].
Available: <https://ptki.ac.id/jurnal/index.php/readystar/article/view/82>
- [16] C. Zhang, J. Wang, X. Li, F. Fu, and W. Wang, “Clustering Centroid Selection using a K-means and Rapid Density Peak Search Fusion Algorithm,” *Proc. IEEE Int. Conf. Softw. Eng. Serv. Sci. ICSESS*, vol. 2020-Octob, pp. 201–207, 2020, doi: 10.1109/ICSESS49938.2020.9237746.