Penerapan Algoritma K-Means untuk Pemetaan Kebutuhan Obat Sesuai Prioritas dalam Pengadaan

Laili Nur Azizah[1] , Lis Utari[2], Leny Tritanto Ningrum[3], dan Dwi Rahmiyati[4]*

[1,3,4]Program Studi Sistem Informasi, Universitas Binaniaga Indonesia

[2]Program Studi Teknik Informatika, Universitas Binaniaga Indonesia

**Penulis Korespondensi : lenytrinie@unbin.ac.i[*]*

ABSTRACT

Each Ministry and Institution (K/L) with a primary clinic has its own budget allocation to support operational health services. The budget is adjusted according to the size of the organization and the needs of its employees, including for purchasing medical supplies. Primary Clinics are expected to provide healthcare services according to standards and the needs of each scope within the budget. Drug procurement planning is essential. Segmentation/priority mapping of drug needs is essential as a reference for drug procurement, by mapping each type of drug based on the required drug needs and the available budget. Ideally, types of drugs with high usage should be prioritized to ensure their procurement. This study used input data on drug usage from January to June 2022 (213 types of drugs). The variables used were Total Usage and Number of Transactions, using the Clustering method with the K-Means algorithm. The final drug mapping results yielded three clusters: 197 low-priority drugs, 13 medium-priority drugs, and 3 high-priority drugs. The evaluation test of the clustering results using the Silhouette Coefficient obtained a value of 0.863, which is concluded as a "strong structure."

Article History

Received:
Revised:
Accepted: – font 9

Keywords

Clustering
Data Mining
Drug Grouping
Segmentation

ABSTRAK

Setiap Kementerian dan Lembaga (K/L) yang memiliki klinik pratama mempunyai alokasi anggaran tersendiri untuk menunjang kegiatan operasional pelayanan kesehatan. Besar anggaran disesuaikan dengan besar organisasi dan kebutuhan pegawainya, yang salah satunya untuk belanja persediaan obat-obatan. Klinik Pratama diharapkan dapat memberikan pelayanan kesehatan sesuai standar dan kebutuhan setiap ruang lingkup sesuai anggaran. Perencanaan pengadaan obat sangat diperlukan. Segmentasi/pemetaan prioritas kebutuhan obat sangat diperlukan sebagai acuan pengadaan obat, dengan cara memetakan setiap jenis obat berdasarkan kebutuhan obat yang diperlukan dan anggaran yang tersedia. Idealnya jenis obat dengan pemakaian tinggi merupakan kelompok obat yang perlu diprioritaskan agar terjamin pengadaannya. Penelitian ini menggunakan Input data pemakaian obat bulan Januari s.d Juni Tahun 2022 (213 jenis obat). Adapun variabel yang digunakan adalah Total Pemakaian dan Jumlah Transaksi, menggunakan metode Clustering dengan algoritma K-Means. Hasil akhir pemetaan obat diperoleh 3 cluster dengan jumlah 197 obat prioritas rendah, 13 obat prioritas sedang dan 3 obat prioritas tinggi. Uji evaluasi hasil clustering dengan Silhoutte Coefficient diperoleh nilai 0,863 yang disimpulkan dengan “strong structure”.

PENDAHULUAN

Kehadiran Unit Pelayanan Kesehatan (UPK) yang berbentuk klinik pratama pada instansi Kementerian atau Lembaga bertujuan untuk mendukung kebijakan pemerintah dalam menyediakan layanan kesehatan yang berkualitas dan merata bagi seluruh lapisan masyarakat tanpa diskriminasi. Masing-masing instansi yang menyelenggarakan klinik pratama mendapatkan alokasi anggaran tersendiri guna menunjang operasional pelayanan kesehatan di fasilitas tersebut. Besaran anggaran tersebut umumnya disesuaikan dengan ukuran organisasi serta kebutuhan pegawai yang dilayaninya. Salah satu penggunaan anggaran tersebut adalah untuk pengadaan obat-obatan. Klinik pratama di

lingkungan Kementerian/Lembaga diharapkan mampu memberikan layanan kesehatan yang sesuai dengan standar serta kebutuhan spesifik instansinya, dengan mempertimbangkan ketersediaan dana yang ada. Oleh karena itu, agar penggunaan anggaran menjadi lebih efisien dan tepat sasaran, diperlukan perencanaan pengadaan obat yang terstruktur dan matang. Secara prinsip, perencanaan obat merupakan proses penetapan jenis dan jumlah obat yang akan diadakan untuk menjamin keberlangsungan pelayanan kesehatan, disesuaikan dengan keterbatasan anggaran yang tersedia. Dalam proses ini, penentuan skala prioritas sangat penting, yakni dengan mengidentifikasi kebutuhan utama secara sistematis, sebelum mempertimbangkan kebutuhan tambahan yang masih memungkinkan untuk ditunda pemenuhannya. [1]. Penentuan prioritas dalam proses identifikasi kebutuhan obat harus dilakukan secara cermat agar dana yang tersedia dapat dimanfaatkan secara optimal untuk pengadaan obat sesuai dengan kebutuhan riil. Berdasarkan studi-studi terdahulu, banyak peneliti yang menerapkan metode *K-Means Clustering* untuk mengatasi tantangan dalam manajemen persediaan obat. Salah satunya adalah penelitian yang dilakukan oleh Gustientiedina, M. Hasmil Adiya, dan Yenny Desnelita berjudul *"Penerapan Algoritma K-Means untuk Clustering Obat–obatan di RSUD Pekanbaru"*. Hasil penelitian tersebut menunjukkan bahwa algoritma *K-Means* efektif dalam mengelompokkan data obat, yang kemudian menghasilkan informasi mengenai klasifikasi obat berdasarkan tingkat konsumsinya, yaitu tinggi, sedang, dan rendah. Informasi ini menjadi dasar penting dalam proses pengambilan keputusan terkait perencanaan dan pengendalian stok medis di lingkungan rumah sakit.

Masalah yang dihadapi saat ini adalah belum dilakukannya pemetaan kebutuhan obat berdasarkan tingkat prioritas dalam proses pengadaan, sehingga jenis obat yang paling dibutuhkan belum dapat diidentifikasi secara akurat. Banyaknya variasi obat yang tersedia, ditambah dengan keterbatasan anggaran, menyebabkan pihak pengelola persediaan mengalami kesulitan dalam menentukan kelompok obat yang seharusnya diprioritaskan, sehingga pengadaan sering kali hanya didasarkan pada estimasi tanpa landasan analisis yang mendalam. Perkiraan semacam ini cenderung kurang tepat apabila tidak didukung oleh evaluasi atas data konsumsi obat pada periode sebelumnya. Untuk mengatasi hal tersebut, segmentasi kebutuhan obat berdasarkan data pemakaian dari bulan-bulan sebelumnya dapat dilakukan melalui metode *clustering*, khususnya dengan algoritma *K-Means*. Permasalahan ini menjadi dasar dilakukannya penelitian dengan judul *"Penerapan Algoritma K-Means untuk Pemetaan Kebutuhan Obat Sesuai Prioritas dalam Pengadaan"*.

Data mining merupakan suatu proses untuk mengekstraksi pola-pola yang bermakna serta pengetahuan tersembunyi dari kumpulan data dalam jumlah besar. Istilah ini digunakan untuk menggambarkan kegiatan penemuan informasi yang sebelumnya tidak diketahui dari basis data yang tersedia. [2]. Salah satu metode yang digunakan dalam *data mining* adalah *clustering*, atau dalam bahasa Indonesia dikenal sebagai teknik pengelompokan. *Clustering* bertujuan untuk mengelompokkan sekumpulan data atau objek ke dalam beberapa klaster berdasarkan kesamaan karakteristik tertentu dalam himpunan data yang besar. Tujuan utamanya adalah membentuk kelompok-kelompok data di mana setiap klaster terdiri dari objek-objek yang memiliki kemiripan satu sama lain [3].

Tujuan dari penelitian ini adalah untuk mendapatkan pemetaan kebutuhan obat sesuai prioritas dalam pengadaan dan mendapatkan perhitungan yang lebih akurat dalam pemetaan kebutuhan obat sesuai prioritas.

METODE

Metode Penelitian

Proses *data mining* terdiri dari beberapa tahapan yang harus dilalui untuk memperoleh pengetahuan baru yang bernilai. Pengetahuan yang dihasilkan dari proses ini diharapkan dapat memberikan kontribusi yang lebih optimal dalam pengambilan keputusan. Secara umum, tahapan dalam *data mining* meliputi pemilihan data (*data selection*), pembersihan atau praproses data (*pre-processing/cleaning*), transformasi data (*transformation*), proses penambangan data itu sendiri (*data mining*), dan tahap evaluasi hasil (*evaluation*). [4].

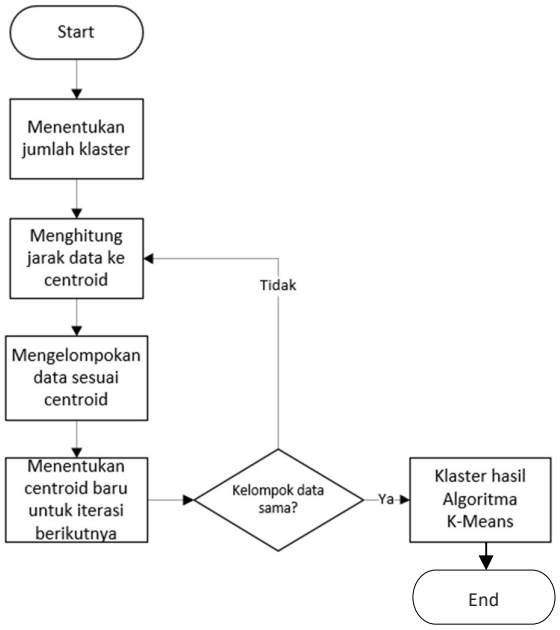
Penelitian ini menggunakan salah satu algoritma dalam pendekatan *unsupervised learning*, yaitu algoritma *K-Means*. *K-Means* merupakan metode *clustering* yang bersifat iteratif atau dilakukan secara berulang. Prosesnya diawali dengan pemilihan jumlah *cluster* secara acak yang

dilambangkan dengan *K*, yaitu jumlah kelompok yang ingin dibentuk [5]. Langkah-langkah pelaksanaan algoritma *K-Means* meliputi [6]: 1) Menetapkan jumlah klaster yang diinginkan dan diberi notasi *K* [7]; 2) Menghitung jarak antara setiap data dengan titik pusat (*centroid*) menggunakan rumus *Euclidean distance* [8]; 3) Mengelompokkan data berdasarkan kedekatannya dengan *centroid*, yakni berdasarkan jarak terpendek dari masing-masing data [9]; 4) Menentukan *centroid* baru untuk iterasi selanjutnya dengan menghitung nilai rata-rata dari seluruh data dalam masing-masing klaster; 5) Melakukan penghitungan ulang jarak data terhadap *centroid* baru menggunakan rumus *Euclidean distance*; dan 6) Mengelompokkan kembali data sesuai *centroid* baru berdasarkan jarak terkecil yang diperoleh dari perhitungan *Euclidean distance*, sehingga tiap data akan ditempatkan pada klaster dengan nilai jarak terdekat., rumus *Euclidean Distance* dapat dilihat pada persamaan (1).

$$\boldsymbol{deuclidean(x,y)} = \sqrt{\sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} \boldsymbol{(xi - yi)^2}} \ldots\ldots\ldots (1)$$

dimana x=x1,x2.....,xn dan y=y1,y2,.....yn
yang merupakan banyaknya n atribut (kolom) antara 2 record;

Alur algoritma K-Means dapat dilihat pada Gambar 1, yaitu alur cara kerja algoritma K-Means.



Gambar 1. Alur kerja algoritma K-Means

Berdasarkan studi literasi yang dilakukan algoritma K-Means mampu menghasilkan kelompok data yang dapat membantu dalam pengambilan keputusan sesuai permasalahan yang terjadi yaitu prioritas pengadaan obat.

Evaluasi

Klaster atau kelompok data hasil dari K-Means tidak dapat langsung digunakan namun perlu dilakukan evaluasi terlebih dahulu untuk mengetahui kualitas hasil klaster. Terdapat beberapa metode evaluasi yang dapat diterapkan pada hasil klaster algoritma K-Means diantaranya *Davies Bouldin Index* (DBI), *Dunn-Index*, dan *Silhouette Coefficient* [10][11][12]. Pada penelitian ini

metode pengujian yang digunakan untuk mengevaluasi klaster tersebut adalah *Silhouette Coefficient* yang akan mengukur seberapa optimal sebuah klaster terbentuk pada proses klasterisasi [13]. Adapun langkah pengujian hasil klaster dengan *Silhouette Coefficient* adalah: 1) hitung rata-rata jarak data, 2) hitung jarak rata-rata data ke-i dengan seluruh data di klaster lain, 3) setelah menghitung jarak data pada seluruh klaster yaitu tentukan nilai terkecil dengan mencari nilai paling minimal, 4) nilai *silhouette coefficient* dapat dihitung berdasarkan pembagian antara selisih rata-rata jarak data dan nilai paling minimal dengan nilai maksimal dari rata-rata jarak data dan nilai paling minimal.

Jarak antara data dapat dihitung menggunakan rumus *euclidean distance*. Untuk mengetahui kualitas hasil klaster lainnya dapat dihitung nilai *silhouette coefficient* pada masing-masing klaster [14]. Adapun kriteria kualitas klaster yang dihasil dari nilai *silhouette coefficient* [15] dapat dilihat pada Tabel 1, yaitu nilai *silhouette coefficient*.

Tabel 1. Nilai *Silhouette Coefficient*

No	Nilai Silhouette Coefficient	Kriteria Klaster
1	0,71 – 1,00	Strong Cluster
2	0,51 – 0,70	Medium Cluster
3	0,26 – 0,50	Weak Cluster
4	0 – 0,25	No Structure

HASIL DAN PEMBAHASAN

Metode yang digunakan adalah metode klasterisasi Algoritma *K- Means* yang akan menghasilkan sebuah klaster yang nantinya akan digunakan untuk mengelompokkan data kebutuhan obat sesuai prioritas. Dalam *data Mining* ada beberapa tahapan sebelum penerapan Algoritma *K-Means* yaitu:

a) Data Selection

Data yang berhasil dikumpulkan adalah data transaksi pemakaian obat dari bulan Januari 2022 hingga Juni tahun 2022. Data pemakaian obat tersebut sebanyak 1934 record dengan 213 jenis obat. Data tersebut diambil dari aplikasi persediaan dan kemudian melakukan eksport data dalam bentuk excel. Tabel 2 merupakan data Pemakaian obat dari bulan Januari s.d Juni Tahun 2022 . Dalam tahap data selection ini data yang dikumpulkan akan diseleksi atau dipilih menjadi data yang siap digunakan dalam proses klusterisasi. Pada tahap pemilihan data, data yang dipakai adalah transaksi pemakaian obat dari bulan Januari 2022 sampai dengan Juni 2022 dengan pemilihan field tgl.transaksi, kode obat, nama obat dan jumlah. Dilakukan pengecekan data apakah ada duplikat data, data yang tidak konsisten, data yang kosong dari data yang sudah dikumpulkan. Dalam tahap ini dilakukan proses penggabungan data pemakaian obat setiap bulan yaitu Januari s.d Juni 2022.

Tabel 2. Data Pemakaian Obat

No.	Tgl_Transaksi	Kode_Obat	Nama_Obat	Jumlah
1	03/01/2022	1010401001000094	Betadine 1L	1
2	03/01/2022	1010401001000033	Betadine Antiseptic 1000 ml	1
3	03/01/2022	1010401001000057	Betadine Kumur 100 ml	2
4	03/01/2022	1010401005000034	Bioplacenton 15 gr	1
...	...	...	...	...
1931	30/06/2022	1010401002000211	Sanmol	4
1932	30/06/2022	1010401002000077	Spasmal	8
1933	30/06/2022	1010401002000110	Zinc 20 mg	10
1934	30/06/2022	1010401002000210	Formuno	10

b) Pre-processing

Data yang dikumpulkan sebanyak 1934 record transaksi. Data tersebut adalah data pemakaian obat semester 1 (satu) di tahun 2022. Data tersebut kemudian diolah menjadi dataset yang akan

digunakan dalam proses data mining. Data tersebut di lakukan pengelompokkan berdasarkan kode_obat, kemudian dilakukan proses penjumlahan obat dan dilakukan perhitungan jumlah transaksi dari masing masing kode obat. Dalam tahap ini dilakukan proses perhitungan total pemakaian dari masing-masing obat yaitu 213 obat dan jumlah transaksi dari masing-masing obat. Sehingga didapat data seperti Tabel 3.

Tabel 3. Dataset hasil pre-processing

No.	Nama Obat	Total Pemakaian	Jumlah Transaksi
1	Acarbose 100 mg	6	2
2	Acetylcystein 200 mg	110	9
3	Acyclovir 5% Krim	2	1
4	Acyclovir salep	5	2
...	...	...	...
210	Vitamin B12	90	2
211	Vitamin B6	25	4
212	Xepazym	92	13
213	Zinc 20 mg	190	12

c) Transformation
Dalam penelitian ini tidak melakukan transformation data karena dataset sudah dalam bentuk angka (numerik).

Penerapan K-Means

Dari data pemakaian obat yang didapat dari gabungan laporan pemakaian obat setiap bulan yang diambil dari sistem yang sudah ada yaitu Sistem Aplikasi persediaan seperti yang tertera pada tabel 2 yang kemudian diolah menjadi dataset seperti tabel 3 akan dibuat cluster untuk memetakan kebutuhan obat sesuai prioritas berdasarkan variabel yang ditentukan, yaitu Total Pemakaian dan Jumlah Transaksi. Penerapan Algoritma *K-Means* bertujuan untuk membagi data obat menjadi beberapa kelompok, dengan tahapan sebagai berikut :

1) Menyiapkan dataset
Dataset yang digunakan pada perhitungan ini menggunakan *variabel* data total pemakaian dan jumlah transaksi dari data pemakaian obat selama bulan Januari 2022 s.d Juni 2022 yang telah diproses. Dataset tersebut terdiri dari 213 baris obat dan 2 kolom *variabel* yaitu total pemakaian dan jumlah transaksi.
2) Menentukan jumlah klaster
Berdasarkan hasil analisa kebutuhan pengguna**,** dari tabel dataset Pemetaan Kebutuhan Obat Bdibuat 3 (tiga) kelompok yaitu kelompok prioritas tinggi, kelompok prioritas sedang dan kelompok prioritas rendah
3) Menentukan centroid awal
Untuk penentuan *centroid* awal dilakukan secara acak dengan memilih data yang ada dalam dataset untuk dijadikan *centroid* awal. Dalam perhitungan pemetaan kebutuhan obat sesuai prioritas dalam pengadaan ini. Data yang dipilih menjadi *centroid* awal dalam perhitungan ini adalah adalah data ke-7, data ke-10 dan data ke-27.
4) Menghitung jarak dengan pusat data (centroid)
Setiap data input dihitung jaraknya terhadap masing-masing pusat klaster menggunakan rumus *Euclidean distance*, sehingga dapat diketahui seberapa dekat suatu data dengan *centroid* yang ada. Proses ini dilakukan dengan menerapkan persamaan matematis *Euclidean distance* untuk menentukan jarak terpendek dari tiap data ke pusat klaster.

Iterasi 1

Data ke-1 pada iterasi 1

Menghitung jarak data ke-1 dengan *centroid 1*

$c_1 = \sqrt{(6-2)^2 + (2-2)^2}$
$= \sqrt{(16+0)}$
$= \sqrt{16} = 4$

Menghitung jarak data ke-1 dengan *centroid 2*

$c_2 = \sqrt{(6-410)^2 + (2-23)^2}$
$= \sqrt{(163216+441)}$
$= \sqrt{163657} = 404{,}55$

Menghitung jarak data ke-1 dengan *centroid 3*

$c_3 = \sqrt{(6-2500)^2 + (2-187)^2}$
$= \sqrt{(6220036+34225)}$
$= \sqrt{6254261} = 2500{,}85$

Hasil perhitungan *euclidean distance* untuk data ke-1 adalah:
C1 = 4
C2 = 404,55
C3 = 2500,85

Proses penghitungan jarak data dengan euclidean distance dilanjutkan seterusnya sampai dengan data ke 213 dan masing-masing data memiliki jarak terdekatnya.

5) Mengelompokkan data berdasarkan jarak minimum ke centroid
Setelah diselesaikan proses hitung jarak data selanjutnya adalah pengelompokan data berdasarkan jarak minimum ke centroid seperti dapat dilihat pada Tabel 4. Hasil Pengelompokan klister.

Tabel 4. Hasil Pengelompoka Klaster Iterasi ke-1

No.	Nama Obat	C1	C2	C3	Jarak Terdekat	Hasil
1	Acarbose 100 mg	4,00	404,55	2500,85	4,00	**C1**
2	Acetylcystein 200 mg	108,23	300,33	2396,62	108,23	**C1**
3	Acyclovir 5% Krim	1,00	408,59	2504,92	1,00	**C1**
4	Acyclovir salep	3,00	405,54	2501,85	3,00	**C1**
...	...	...	...	...	...	...
213	Zinc 20 mg	188,27	220,27	2316,62	188,27	**C1**

Tabel 4 menampilkan hasil perhitungan masing-masing jarak antara data ke pusat klaster C1, C2, dan C3 menggunakan rumus *euclidean distance*, sehingga didapat jarak terdekat dari masing-masing data obat dengan pusat klaster kemudian kelompokan data ke klaster terdekat.

6) Hitung kembali centroid
Setelah dilakukan pengelompokkan data kemudian dilakukan penentuan centroid baru. Penentuan nilai *centroid* baru diperoleh dengan menghitung rata-rata (*mean*) dari semua data pada setiap kelompok (*cluster*), sehingga didapat *centroid* baru untuk iterasi ke-2 dan iterasi berikutnya.

7) Ulangi langkah 4 sampai 6 sampai hasil pengelompokan tidak berubah
Tahapan selanjutnya yaitu mengulangi proses pada Langkah ke 4 sampai 6 hingga hasil pengelompokan data pada iterasi terakhir sama dengan pengelompokan data pada iterasi sebelumnya. Dalam penelitian ini perhitungan tidak dilanjutkan ke iterasi ke-5 dan berhenti di iterasi ke 4 karena hasil kelompok data sudah tidak berubah dan sama dengan iterasi sebelumnya.

Hasil pemetaan dibagi menjadi 3 klaster dengan inisial penyebutan nama klaster tersebut adalah C1, C2, dan C3. Pemetaan klaster tersebut dilihat berdasarkan jumlah transaksi yaitu C1 sebagai kelompok prioritas rendah dengan jumlah transaksi dibawah 20, C2 sebagai kelompok obat dengan prioritas sedang dengan jumlah transaksi diatas 20 dan dibawah 50, dan C3 sebagai kelompok obat dengan prioritas tinggi dengan jumlah transaksi diatas 50 dengan hasil kelompok dapat dilihat pada Tabel 5. Kelompok prioritas obat.

Tabel 5. Kelompok Klaster Prioritas

No.	Nama Obat	Total Pemakaian	Jumlah Transaksi	Hasil
1	Acarbose 100 mg	6	2	Prioritas Rendah
2	Acetylcystein 200 mg	110	9	Prioritas Rendah
3	Acyclovir 5% Krim	2	1	Prioritas Rendah
4	Allopurinol 100 mg	410	23	Prioritas Sedang
5	Amlodipine 10 mg	450	25	Prioritas Sedang
6	Amlodipine 5 mg	780	39	Prioritas Sedang
...	...	...	...	...
213	Sanmol	4326	200	Prioritas Tinggi

a) Evaluasi

Berdasarkan hasil pengembangan terkait metode yang diterapkan pada dilakukan uji hasil dengan menerapkan metode *Silhoutte Coefficient* dengan menggunakan aplikasi matlab dan hasil nilai *Silhouette Coefficient* dapat dilihat pada Tabel 6. Nilai *Silhouette Coefficent.*

Tabel 6. Nilai *Silhouette Coefficent*

Data ke-	Nilai Silhouette
1	0.93785
2	0.79195
3	0.93663
...	...
213	0.54612

Nilai rata-rata validasi dari setiap cluster berada pada point 0.8630 dan berdasarkan kategori nilai *silhouette coefficient* menurut Kaufman dan Roussewwuw (1990) nilai 0.7 <SC<= 1 masuk dalam kategori *Strong Structure* artinya Terdapat ikatan yang sangat baik (strong structure) antara objek dan kelompok yang terbentuk.

KESIMPULAN

Berdasarkan hasil analisis dan pembahasan yang telah dilakukan, dapat disimpulkan bahwa penggunaan variabel total pemakaian dan jumlah transaksi dalam metode *K-Means clustering* mampu mengelompokkan 213 jenis obat ke dalam tiga kategori prioritas. Hasil pengelompokan tersebut terdiri dari 197 obat dalam kelompok prioritas rendah, 13 obat dalam kelompok prioritas sedang, dan 3 obat dalam kelompok prioritas tinggi. Evaluasi menggunakan *Silhouette Coefficient* menghasilkan skor sebesar 0,863. Mengacu pada interpretasi yang dikemukakan oleh Kauffman dan Rousseeuw, nilai tersebut termasuk dalam kategori *strong structure*, yang menunjukkan bahwa klaster yang terbentuk memiliki struktur yang kuat dengan tingkat kemiripan data yang tinggi di dalam masing-masing kelompok.

DAFTAR PUSTAKA

[1] D. Waluyo, *Ilmu Pengetahuan Sosial*. Jakarta: Pusat Perbukuan Departemen Pendidikan Nasional, 2008.

[2] J. Han, *Data Mining Concepts and Techniques*, Fourth. United States: Elsevier, 2011.

[3] B. Santoso, *Data Mining: Teknik Pemanfaatan Data untuk Keperluan Bisnis*, 1st ed. Yogyakarta: Graha Ilmu, 2007.

[4] E. T. Kusrini, Kusrini., Luthfi, *Algoritma Data Mining*, 1st ed. Yogyakarta: Andi Offset, 2009.

[5] L. T. Ningrum and D. Rahmiyati, "Analysis Sentiment of Twitter User on Indonesia's 2024 Presidential Election Using K-Means Algorithm," *J. Sisfotek Glob.*, vol. 13, no. 2, p. 102, 2023, doi: 10.38101/sisfotek.v13i2.9609.

[6] L. T. Ningrum, I. Irmayansyah, and L. Utari, "Penerapan Metode K-Means dan Euclidean Distance Untuk Seleksi Metode Judul Tugas Akhir," *Acad. J. Comput. Sci. Res.*, vol. 6, no. 1, p. 13, 2024, doi: 10.38101/ajcsr.v6i1.10766.

[7] M. S. Yang and K. P. Sinaga, "A feature-reduction multi-view k-means clustering algorithm," *IEEE Access*, vol. 7, pp. 114472–114486, 2019, doi: 10.1109/ACCESS.2019.2934179.

[8] L. U. Zogara, A. Sururi, and L. T. Ningrum, "Analysis of COVID-19 Information Based on Social Media Big Data Classification Using the K-Means Data Mining Method," *J. Sisfotek Glob.*, vol. 12, no. 1, p. 18, 2022, doi: 10.38101/sisfotek.v12i1.448.

[9] C. S. D. B. Sembiring, L. Hanum, and S. P. Tamba, "Penerapan Data Mining Menggunakan Algoritma K-Means Untuk Menentukan Judul Skripsi Dan Jurnal Penelitian (Studi Kasus Ftik Unpri)," *J. Sist. Inf. dan Ilmu Komput. Prima(JUSIKOM PRIMA)*, vol. 5, no. 2, pp. 80–85, 2022, doi: 10.34012/jurnalsisteminformasidanilmukomputer.v5i2.2393.

[10] D. A. I. C. Dewi and D. A. K. Pramita, "Analisis Perbandingan Metode Elbow dan Silhouette pada Algoritma Clustering K-Medoids dalam Pengelompokan Produksi Kerajinan Bali," *Matrix J. Manaj. Teknol. dan Inform.*, vol. 9, no. 3, pp. 102–109, 2019, doi: 10.31940/matrix.v9i3.1662.

[11] Y. Sopyan, A. D. Lesmana, and C. Juliane, "Analisis Algoritma K-Means dan Davies Bouldin Index dalam Mencari Cluster Terbaik Kasus Perceraian di Kabupaten Kuningan," *Build. Informatics, Technol. Sci.*, vol. 4, no. 3, pp. 1464–1470, 2022, doi: 10.47065/bits.v4i3.2697.

[12] S. Suraya, M. Sholeh, and D. Andayati, "Penerapan Metode Clustering Dengan Algoritma K-Means Pada Pengelompokan Indeks Prestasi Akademik Mahasiswa," *Skanika*, vol. 6, no. 1, pp. 51–60, 2023, doi: 10.36080/skanika.v6i1.2982.

[13] E. Irwansyah and M. Faisal, *Advanced Clustering: Teori dan Aplikasi*. Yogyakarta: Deepublish, 2015.

[14] S. Paembonan and H. Abduh, "Penerapan Metode Silhouette Coefficient untuk Evaluasi Clustering Obat," *PENA Tek. J. Ilm. Ilmu-Ilmu Tek.*, vol. 6, no. 2, p. 48, 2021, doi: 10.51557/pt_jiit.v6i2.659.

[15] H. Sa'diah, U. Enri, and T. Nur Padilah, "Penerapan Algoritme K-Means Dalam Segmentasi Daerah Rawan Kekerasan Anak Di Jawa Barat," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 2, pp. 1351–1357, 2023, doi: 10.36040/jati.v7i2.6838.