Penerapan Metode Silhouette Coefficient Untuk Evaluasi Clustering Obat

Solmin Paembonan¹, Hisma Abduh²

^{1,2} Universitas Andi Djemma, Palopo, Indonesia
Email: ¹solmin.p@unanda.ac.id, ²isma.syakirah21@gmail.com

Abstract

Dalam penelitian ini menggunakan metode k-means, metode ini dapat digunakan untuk menjadikan beberapa obat yang mirip menjadi suatu kelompok data tertentu. Salah satu cara untuk mengetahui tingkat kemiripan data adalah melalui perhitungan jarak antar data. Semakin kecil jarak antar data semakin tinggi tingkat kemiripan data tersebut dan sebaliknya semakin besar jarak antar data maka semakin rendah tingkat kemiripannya. Tujuan akhir clustering adalah untuk menentukan kelompok dalam sekumpulan data yang tidak berlabel, karena clustering merupakan suatu metode unsupervised dan tidak terdapat suatu kondisi awal untuk sejumlah cluster yang mungkin terbentuk dalam sekumpulan data, maka dibutuhkan suatu evaluasi hasil clustering. Berdasarkan evaluasi yang dilakukan terhadap hasil clustering dengan nilai dari silhouette coefficient = 0,4854.

Kata Kunci

Clustering; K-means;
Silhouette coefficient.

Abstract

In this study using the k-means method, this method can be used to make several similar drugs into a certain data group. One way to determine the level of similarity of the data is through the calculation of the distance between the data. The smaller the distance between the data, the higher the level of similarity between the data and vice versa, the greater the distance between the data, the lower the similarity level. For a number of clusters that may be formed in a data set, an evaluation of the results of clustering is needed. Based on the evaluation carried out on the results of clustering with the value of the silhouette coefficient = 0.4854..

Keywords

Clustering; K-means;
Silhouette coefficient.

1. PENDAHULUAN

Menurut SK Menteri Kesehatan. No.193/Kab/B.VII/ 71 tanggal 9 Juni 1971, yang disebut dengan obat ialah suatu bahan atau paduan bahan-bahan untuk digunakan dalam menetapkan diagnosis, mencegah, mengurangi, menghilangkan, menyembuhkan penyakit, luka atau kelainan badaniah dan rohaniah pada manusia atau hewan, memperelok badan atau bagian badan manusia.

Obat dapat diganti dengan persetujuan dokter penulis resep, apoteker dan pasien/konsumen dapat langsung berkonsultasi dengan dokter untuk meminta persetujuan penggantian obat. Dokter berhak mengganti obat bermerek atau obat paten dengan obat paten lainnya atau dengan obat generik, dan dokter boleh mengganti obat generik dengan obat paten. Setelah mendapat persetujuan dari dokter dan pasien maka apoteker berhak mengganti obat yang dibutuhkan konsumen.

Proses penemuan pola dapat dilakukan dengan mengelompokkan data kedalam cluster-cluster sehingga data-data yang memiliki kemiripan berada pada cluster yang sama dan data-data yang tidak memiliki kemiripan terletak pada cluster yang berbeda (Abul Hasan 2011).

K-means cluster sebagai solusi untuk pengklusteran karakteristik dari objek yang digunakan untuk meringkas pengolahan objek dalam jumlah yang besar dari jumlah obat sehingga lebih memudahkan untuk mendeskripsikan sifat-sifat atau karakteristik dari masing-masing kelompok obat. (Jiptake, 2012) menyatakan k-means clustering menghasilkan akurasi yang cukup tinggi dan membutuhkan waktu yang cukup singkat. Dalam penelitian ini menggunakan metode k-means untuk menjadikan beberapa obat yang mirip menjadi suatu kelompok data tertentu. Obat dengan

karakteristik yang sama akan dikelompokkan dalam satu cluster dan obat dengan karakteristik yang berbeda dikelompokkan kedalam cluster yang lain.

Penggunaan metode silhouette dalam penelitian ini, dimana pada metode ini digunakan untuk merepresentasikan seberapa tepat suatu objek data ditempatkan pada sebuah cluster.

2. METODE PENELITIAN

2.1 Deskripsi Sistem

Algoritma *k-means cluster* analisis termasuk dalam kelompok metode *cluster* analisis non hirarki yang membagi set data kedalam *cluster* yang tidak bertumpang-tindih antara satu *cluster* dengan *cluster* yang lain artinya setiap data hanya menjadi anggota satu *cluster* saja, dimana jumlah kelompok yang akan dibentuk sudah terlebih dahulu diketahui atau ditetapkan jumlahnya, salah satu cara untuk mengetahui tingkat kemiripan data untuk metode *k-means* adalah melalui perhitungan jarak antar data. Semakin kecil jarak antar data semakin tinggi tingkat kemiripan data tersebut dan sebaliknya semakin besar jarak antar data maka semakin rendah tingkat kemiripannya.

Jumlah *cluster* yang berbeda dari set data yang sama memberikan nilai evaluasi yang berbeda pula. Pemilihan metode *silhouette* digunakan untuk melihat kualitas *cluster* dimana seberapa baik suatu objek ditempatkan dalam suatu *cluster*.

2.2 Sistem yang diusulkan

Adapun diagram alir dokumen sistem yang diusulkan dapat dilihat pada Gambar 1.

Gambar 1. Sistem yang diusulkan

2.3 Clustering

Jain (2010) menyatakan *clustering* adalah proses pengelompokan suatu pola yang belum memiliki label dan dilakukan tanpa supervisi menjadi sebuah kelompok yang memiliki karakteristik tertentu. Hal ini didukung oleh Darmi dan Setiawan (2016). *Clustering* sangat penting pada beberapa permasalahan antar lain analisa pola, pembuatan keputusan, machine learning, data mining dan sebagainya.

Teknik *clustering* termasuk teknik yang sudah cukup dikenal dan banyak dipakai dalam *data mining*. *Data mining* melibatkan integrasi teknik dari berbagai disiplin ilmu seperti database dan data warehouse teknologi, statistik, pengenalan pola, jaringan saraf, visualisasi data, serta pemrosesan sinyal (Mestien & Sari 2015). Oleh karena itu, data mining dianggap salah satu bidang ilmu yang paling penting dalam database dan sistem informasi sebagai salah satu bidang yang paling menjanjikan dalam perkembangan interdisipliner dalam teknologi informasi. Kelebihan dari algoritma k-means terletak pada kecepatan untuk mencapai konvergen serta kemudahan dalam pengimplementasian (Kao dan Lee, 2009).

2.4 K-means

K-means merupakan algoritma *clustering* yang termasuk kedalam *partition clustering* merupakan metode *clustering* yang mengelompokkan berdasarkan tingkat kesamaan antar data (Alam, 2014) (Romsini dkk, 2018) (Handoko, 2016). Data yang berada pada *cluster* yang sama memiliki tingkat kesamaan yang tinggi sedangkan data yang berada pada *cluster* yang berbeda tingkat kemiripannya rendah. Penentuan kemiripan bisa dilakukan dengan banyak cara antara lain berdasarkan jarak, pola, dan kepadatan. *Euclidean distance* merupakan salah satu metode penghitungan tingkat kesamaan antar titik data yang berbasis pada jarak. Adapun langkah-langkah *clustering* dengan k-means dapat dijabarkan sebagai berikut (Kao, 2009):

1. Inisialisasi pusat *cluster k* secara random
2. Ulang
 - 2.1 Menghitung jarak masing-masing data dengan tiap-tiap *cluster*. Perhitungan jarak menggunakan *euclidian distance* dengan persamaan (1).

$$D(x_p, z_j) = \sqrt{\sum_{i=1}^d (x_{pi} - z_{ji})^2} \quad 1$$

Dimana:

x_p : titik data ke-p

z_j : titik pusat *cluster* ke-j

d : jumlah atribut dari tiap pusat *cluster*

- 2.2 Setiap data dimasukkan ke dalam *cluster* dengan jarak pusat *cluster* terkecil dengan titik data.

- 2.3 Dilakukan perhitungan pusat *cluster* baru dengan persamaan (2)

$$z_j = \frac{1}{n_j} \sum_{x_p \in c_j} x_p \quad 1$$

Dimana:

n_j : jumlah titik data dalam *cluster j*

c_j : himpunan titik data dalam *cluster k*.

3. Proses terus berulang sampai kriteria berhenti dipenuhi

Ada beberapa kriteria yang dapat dijadikan acuan berhenti dalam proses *clustering* k-means yaitu ketika k-means telah memenuhi jumlah iterasi maksimum yang sudah ditetapkan, jika *centroid* tidak mengalami perubahan posisi terbilang kecil dalam jumlah iterasi tertentu, tidak ada lagi perpindahan *cluster* dari titik data.

2.5 Silhouette Coefficient

Metode ini merupakan metode evaluasi *cluster* yang menggabungkan metode *cohesion* dan *separation*. *Cohesion* diukur dengan menghitung seluruh objek yang terdapat dalam sebuah *cluster* dan *separation* diukur dengan menghitung jarak rata-rata setiap objek dalam sebuah *cluster* dengan *cluster* terdekatnya (Rendon, 2011) (Pramesiti, dkk 2017) (Wira dkk, 2019). Jarak antara data dihitung dengan menggunakan rumus *euclidean distance*. Untuk menyediakan informasi tentang kualitas hasil *clustering* pada proses *clustering*, dapat dihitung *silhouette* dari masing-masing *cluster* bahkan keseluruhan *cluster* dari hasil kerja suatu algoritma *clustering*. Nilai *silhouette* untuk keseluruhan data dengan jumlah *cluster k*, dapat didefinisikan sebagai *sil (k)* yang dihitung dengan persamaan (3) yakni rata-rata *silhouette value* untuk semua *cluster*.

$$sil(c) = sil(k) \frac{1}{|k|} \sum_{i=1}^k sil(c_i) \quad 2$$

Dimana:

$sil(k)$: nilai *silhouette* semua cluster

$|k|$: banyaknya *cluster k*

$sil(c_i)$: rata-rata nilai *silhouette*

2.6 Implementasi

2.5.1. Pembahasan hasil clustering

Setelah melakukan implementasi maka dapat diamati hasil dari *clustering* menggunakan *k-means* yang digunakan dalam pembentukan *cluster*. Dataset yang digunakan dalam penelitian ini diperoleh dari apotek Rs. Grestelina Makassar.

Gambar 1. Hasil *cluster*

Hasil proses *clustering* dapat dilihat dalam bentuk visualisasi grafik *scatter*, yang menampilkan visualisasi sebaran dataset seperti yang tersaji pada Gambar 1. Hasil *clustering* seperti yang tersaji pada Gambar 1 menampilkan informasi data dengan warna hitam merupakan keanggotaan dari *cluster* 1 dan keanggotaan pada *cluster* 2 menampilkan data yang berwarna hijau. Hasil *clustering* ini kemudian diuji menggunakan *silhouette coefisien* untuk semua *cluster* dengan nilai = 0,4854, dengan jumlah $k=2$.

Penggantian obat dapat dilakukan dengan persetujuan dokter penulis resep, apabila terjadi kekosongan, stok obat tidak tersedia, maka apoteker atau pasien dapat langsung berkonsultasi dengan dokter untuk meminta persetujuan penggantian obat. Apoteker dan pasien/konsumen dapat langsung berkonsultasi dengan dokter untuk meminta persetujuan penggantian obat. Menurut PP No.51 tahun 2009 tentang Pekerjaan Kefarmasian, pasal 24 poin b, mengganti obat merek dagang dengan obat generik yang sama komponen aktifnya atau obat merek dagang lain atas persetujuan dokter dan atau pasien. Begitupun dalam Permenkes No. HK.02.02/MENKES/068/I/ 2010/ pasal 7 bahwa apoteker dapat mengganti obat merek dagang/obat paten dengan obat generik yang sama komponen aktifnya atau merek dagang lain atas persetujuan dokter dan/atau pasien dan Permenkes No. HK.02.02/MENKES/068/I/2010/pasal 8 tentang dokter rumah sakit atau puskesmas dan unit pelaksana teknis lainnya dapat menyetujui pengantian resep obat generik dengan obat generik bermerek/bermerek dagang dalam hal obat generik tertentu belum tersedia.

2.5.2. Pengujian Hasil Clustering K-means

Pada proses *clustering* menggunakan *k-means* dengan parameter yang digunakan ialah penentuan jumlah *cluster (k)* dan maksimum iterasi. Parameter ini akan berpengaruh pada penentuan *centroid* yang akan digunakan pada proses perhitungan jarak data dengan *centroid* menggunakan *euclidean* hingga pada penentuan anggota *cluster*. Pengujian hasil *clustering* menggunakan metode *shilouette coefficient* dilakukan dengan jumlah *cluster* $k=2$ sampai $k=10$.

Dari hasil proses perhitungan *silhouette coefficient* terhadap data obat maka hasil *silhouette coefficient* yang maksimum adalah pada saat $k=2$ dengan nilai *silhouette* = 0,4854. Nilai rata-rata *silhouette coefisien* untuk semua *cluster* seperti yang terlihat pada Gambar 2 dan Tabel 1. Berdasarkan hasil pengujian yang di tunjukkan pada Gambar 2 dan Tabel 1 memberikan informasi mengenai jumlah *cluster* dan nilai *silhouette* yang maksimum dengan hasil *clustering* untuk $k=2$ dan nilai *silhouette* = 0,4854, dan ketika jumlah *cluster* (k) semakin besar maka nilai *silhouette* yang dihasilkan cenderung lebih kecil dibanding jumlah *cluster* sebelumnya.

Gambar 2. Grafik nilai *silhouette*

Nilai *silhouette* terlihat kembali meningkat ketika jumlah *cluster* = 6 dengan nilai *silhouette* = 0,4675, namun ketika nilai dari jumlah *cluster* bertambah, nilai *silhouette* kembali menurun. Nilai *silhouette* terkecil saat jumlah *cluster* = 8 dengan nilai *silhouette* = 0,4355.

Tabel 1 Nilai *silhouette*

No	Cluster	Nilai <i>silhouette</i>
1	K2	0,4854
2	K3	0,4746
3	K4	0,4407
4	K5	0,4571
5	K6	0,4675
6	K7	0,4452
7	K8	0,4355
8	K9	0,4399
9	K10	0,4423

2.5.3. Halaman Menu Utama

Halaman utama merupakan tampilan yang akan dijalankan pertama kali oleh *user*. Pada tampilan utama memiliki 2 menu yaitu menu *file*, *tools* dan 3 tombol yang terdiri dari tombol obat, *clustering* dan rekomendasi. Gambar 3 menampilkan halaman utama.

Gambar 3. Form menu utama

2.5.4. Halaman Pencarian Obat

Pencarian obat yang dibutuhkan pasien dilakukan dengan menginputkan nama komposisi obat, dosis obat, jenis obat dan harga obat pada aplikasi, selanjutnya muncul menu *dropdown* untuk memilih satuan obat dan jenis obat yang dicari. Halaman pencarian obat di tunjukkan pada Gambar 3.

Gambar 4. Form Pencarian Obat

2.5.5. Halaman Hasil Rekomendasi Obat

Pada halaman hasil pencarian obat menampilkan waktu pencarian obat dan urutan rekomendasi pengganti obat dengan nama obat, komposisi obat, jenis obat, satuan obat dan harga, seperti yang tersaji pada Gambar 4.

kode_obat	nama_obat	komposisi	satuan	jenis	harga
SANCS01	ANCEFA 250MG FORTE DRY SYR	Cefadroxil monohydrate	250 mg	tablet	6441.08
SQDRX02	QIDROX FORTE SYRUP	Cefadroxil monohydrate	250 mg	tablet	9300.00
SCEF01	CEFAT DRY SYRUP 60 ML	Cefadroxil monohydrate	250 mg	tablet	43670.00
SALXL01	ALXIL 60 ML SYRUP	Cefadroxil monohydrate	250 mg	tablet	44000.00

Gambar 5. Halaman Pencarian Obat

3. KESIMPULAN

Kesimpulan yang dapat diberikan setelah proses pengerjaan dari perancangan, implementasi dan uji coba yang telah dilakukan dengan metode *clustering k-means* yang digunakan untuk proses

rekomendasi pengganti obat bagi pasien. Hasil penelitian menunjukkan bahwa metode *k-means clustering* dengan jumlah $k = 2$ cluster, merupakan jumlah cluster yang ideal menggunakan *silhouette* dengan nilai 0,4854.

4. REFERENCES

- A. K. Jain, M. N. Murty, dan P. J. Flynn, "Data clustering: a review," *ACM Comput. Surv.*, vol. 31, no. 3, hlm. 264–323, Sep 1999.
- Darmi, Y. D., & Setiawan, A. (2016). Penerapan metode clustering k-means dalam pengelompokan penjualan produk. *Jurnal Media Infotama*, 12(2).
- E. Rendón, I. Abundez, A. Arizmendi, and E. M. Quiroz, "Internal versus External cluster validation indexes," vol. 5, no. 1, 2011.
- Handoko, K. (2016). Penerapan Data Mining dalam Meningkatkan Mutu Pembelajaran Menggunakan Metode K-MEANS Clustering. *Jurnal Nasional Teknologi dan Sistem Informasi*, 2(3), 31-40.
- Jiptake, B. R., & Gohokar, V. V. (2012). A Comparative Analysis of Fuzzy C-Means Clustering and K Means Clustering Algorithms. *International Journal of Computational Engineering Research*, 737-739.
- M. J. Abul Hasan dan S. Ramakrishnan, "A survey: hybrid evolutionary algorithms for cluster analysis," *Artif. Intell. Rev.*, vol. 36, no. 3, hlm. 179–204, Okt 2011.
- Metisen, B. M., & Sari, H. L. (2015). Analisis clustering menggunakan metode K-Means dalam pengelompokan penjualan produk pada Swalayan Fadhila. *Jurnal media infotama*, 11(2).
- Pramesti, D. F., Furqon, M. T., & Dewi, C. (2017). Implementasi Metode K-Medoids Clustering Untuk Pengelompokan Data Potensi Kebakaran Hutan/Lahan Berdasarkan Persebaran Titik Panas (Hotspot). *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer e-ISSN, 2548, 964X.Y.*
- Kao dan S.-Y. Lee, "Combining K-means and particle swarm optimization for dynamic data clustering problems," dalam *2009 IEEE International Conference on Intelligent Computing and Intelligent Systems*, Shanghai, China, 2009, hlm. 757–761.
- Rosmini, R., Fadlil, A., & Sunardi, S. (2018). Implementasi Metode K-Means Dalam Pemetaan Kelompok Mahasiswa Melalui Data Aktivitas Kuliah. *IT Journal Research and Development*, 3(1), 22-31.
- S. Alam, G. Dobbie, Y. S. Koh, P. Riddle, dan S. Ur Rehman, "Research on particle swarm optimization based clustering: A systematic review of literature and techniques," *Swarm Evol. Comput.*, vol. 17, hlm. 1–13, Agu 2014.
- Surat Keputusan Menteri Kesehatan Republik Indonesia Nomor 193/Kab/B.VII/71 tentang Pembungkusan Dan Penandaan Obat.
- Wira, B., Budianto, A. E., & Wiguna, A. S. (2019). Implementasi Metode K-Medoids Clustering Untuk Mengetahui Pola Pemilihan Program Studi Mahasiswa Baru Tahun 2018 Di Universitas Kanjuruhan Malang. *Rainstek: Jurnal Terapan Sains & Teknologi*, 1(3), 53-68.