

SKRIPSI

**PERBANDINGAN ANALISIS *CLUSTERING HIERARCHICAL*
DAN *CLUSTERING K-MEDOIDS* DALAM PREDIKSI
POTENSI DAUN KELOR SEBAGAI
OBAT ANTIDIABETES**



**NURDIANI RESKI
E0220301**

**PROGRAM STUDI STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS SULAWESI BARAT
TAHUN 2024**

SURAT PERNYATAAN

Yang bertanda tangan di bawah ini:

Nama : Nurdiani Reski
Tempat/Tgl. Lahir : Tomemba, 16 Juli 2002
NIM : E0220301
Program Studi : Statistika

Menyatakan bahwa tugas akhir dengan judul “Perbandingan *Clustering Hierarchical* dan *Clustering K-Medoids* dalam Prediksi Potensi Senyawa Daun Kelor sebagai Obat Antidiabetes” disusun berdasarkan prosedur ilmiah yang telah melalui pembimbingan dan bukan merupakan plagiat dari karya ilmiah/naskah yang lain. Apabila di kemudian hari terbukti bahwa pernyataan ini tidak benar, maka saya bersedia menerima sanksi sesuai peraturan yang berlaku.

Majene, 20 Mei 2025

Yang menyatakan,



Nurdiani Reski

HALAMAN PENGESAHAN

Skripsi ini diajukan oleh:

Nama : Nurdiani Reski

NIM : E0220301

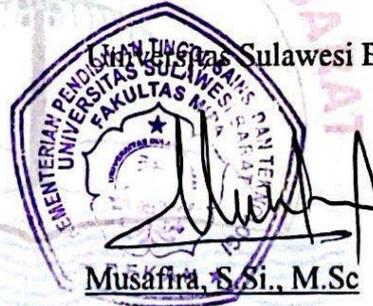
Judul : Perbandingan Analisis *Clustering Hierarchical* dan *K-Medoids* dalam Prediksi Potensi Daun Kelor sebagai Obat Antidiabetes.

Telah berhasil dipertanggungjawabkan di hadapan Tim Penguji (SK Nomor 79/UN55.7./HK04/2024 tanggal 25 Oktober 2024) dan diterima sebagai bagian persyaratan memperoleh gelar Sarjana S1 Statistika pada Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Sulawesi Barat.

Disahkan oleh:

Dekan FMIPA

Universitas Sulawesi Barat

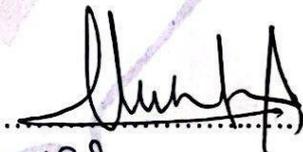


Musafira, S.Si., M.Sc

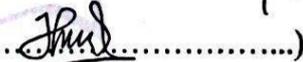
NIP.197709112006042002

Tim Penguji:

Ketua Penguji : Musafira, S.Si., M.Sc

(.....)

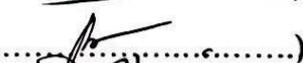
Sekretaris : Muh. Hijrah, S.Pd., M.Si

(.....)

Pembimbing 1 : Nur Hilal A. Syahrir, S.Si., M.Si

(.....)

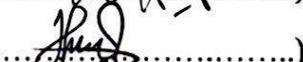
Pembimbing 2 : Retno Mayapada, S.Si., M.S

(.....)

Penguji 1 : Fardinah, S.Si., M.Sc

(.....)

Penguji 2 : Muh. Hijrah, S.Pd., M.Si

(.....)

Penguji 3 : Reski Wahyu Yanti, S.Si., M.Si

(.....)

ABSTRAK

Analisis kluster merupakan analisis teknik multivariat yang bertujuan untuk melakukan pengelompokan berdasarkan kesamaan objek pada berbagai karakteristik. Daun kelor diketahui mengandung beragam senyawa aktif yang baik untuk tubuh. Hal inilah yang menjadi alasan mengapa daun kelor sering digunakan sebagai obat herbal untuk mengatasi gangguan kesehatan tertentu, termasuk diabetes. Penelitian ini bertujuan untuk mengeksplorasi potensi daun kelor sebagai obat antidiabetes dengan membandingkan kemiripan senyawa obat daun kelor dengan obat antidiabetes menggunakan dua metode *clustering* yaitu *clustering hierarchical* dan *clustering K-Medoids*. Analisis dilakukan terhadap data ADMET senyawa daun kelor dan obat antidiabetes. Berdasarkan hasil analisis kluster, metode *clustering hierarchical* dengan pendekatan *ward* menghasilkan *silhouette coefficient* sebesar 0,1340, *dunn Index* sebesar 0,3854 dan *connectivity* sebesar 20,500. Sementara itu, metode *clustering K-Medoids* menghasilkan *silhouette coefficient* sebesar 0,1000, *dunn index* sebesar 0,2831 dan *connectivity* sebesar 38,500. Hal ini menunjukkan bahwa nilai *silhouette coefficient* dan *dunn index* pada *clustering hierarchical* lebih tinggi dibandingkan dengan *clustering K-Medoids* dan nilai *connectivity* lebih rendah dari *clustering K-Medoids*, sehingga kualitas *clustering* menggunakan *clustering hierarchical* lebih baik. Hasil kluster dari *clustering hierarchical* menghasilkan 10 kluster dimana kluster 2 berisikan anggota senyawa obat yaitu *glimpiride* dan senyawa daun kelor yaitu 358357, 191036, 1301760, 1930553, yang menunjukkan terdapat karakteristik yang serupa antara senyawa daun kelor dan senyawa obat antidiabetes. Artinya, senyawa daun kelor berpotensi sebagai obat antidiabetes pada data ADMET dalam studi ini.

Kata Kunci: *Clustering Hierarchical, Clustering K-Medoids, Daun Kelor, Diabetes.*

ABSTRACT

Cluster analysis is a multivariate technique analysis that aims to group objects based on similarities in various characteristics. Moringa leaves are known to contain various active compounds that are good for the body. This is the reason why Moringa leaves are often used as herbal medicine to treat certain health disorders, including diabetes. This study aims to explore the potential of Moringa leaves as an antidiabetic drug by comparing the similarity of Moringa leaf medicinal compounds with antidiabetic drugs using two clustering methods, namely hierarchical clustering and K-Medoids clustering. The analysis was carried out on ADMET data of Moringa leaf compounds and antidiabetic drugs. Based on the results of the cluster analysis, the hierarchical clustering method with the Ward approach produced a silhouette coefficient of 0.1340, a Dunn Index of 0.3854 and a connectivity of 20.500. Meanwhile, the K-Medoids clustering method produced a silhouette coefficient of 0.1000, a Dunn Index of 0.2831 and a connectivity of 38.500. This shows that the silhouette coefficient and dunn index values in hierarchical clustering are higher than K-Medoids clustering and the connectivity value is lower than K-Medoids clustering, so the quality of clustering using hierarchical clustering is better. The cluster results from hierarchical clustering produce 10 clusters where cluster 2 contains members of the drug compound, namely glimepiride and moringa leaf compounds, namely 358357, 191036, 1301760, 1930553, which shows that there are similar characteristics between moringa leaf compounds and antidiabetic drug compounds. This means that moringa leaf compounds have the potential as antidiabetic drugs in the ADMET data in this study.

Keywords: *Clustering Hierarchical, Clustering K-Medoids, Moringa leaves, Diabetes*

BAB I

PENDAHULUAN

1.1 Latar Belakang

Negara Indonesia merupakan salah satu negara berkembang dimana masyarakat banyak mengalami perubahan baik dari gaya hidup maupun pola makan. Gaya hidup yang tidak sehat seperti kurangnya aktivitas fisik, diet yang tidak sehat dan pola makan yang tidak seimbang menyebabkan obesitas. Hal ini mengakibatkan banyak muncul penyakit di dalam tubuh, salah satunya adalah penyakit Diabetes (Ardiani, 2021). Menurut *World Health Organization* (WHO), diabetes merupakan penyakit kronis yang terjadi ketika pankreas tidak menghasilkan cukup insulin atau ketika tubuh tidak dapat secara efektif menggunakan hasil insulin (hormon yang mengatur gula darah) (WHO, 2021).

WHO menyatakan bahwa jumlah kasus diabetes di negara berkembang kemungkinan akan meningkat lebih dari dua kali lipat dalam 30 tahun ke depan, dari 15 juta pada tahun 2000 menjadi 284 juta pada tahun 2030. Pada tahun 2019 diabetes menyebabkan kematian sebanyak 1,5 juta atau 48% terjadi sebelum usia 70 tahun dan sekitar 90% dari perkiraan 171 juta orang di dunia memiliki diabetes (WHO, 2021).

Penggunaan tanaman obat mengalami perkembangan yang semakin meningkat, terlebih dengan munculnya isu kembali ke alam (*back to nature*) sementara itu banyak masyarakat beranggapan bahwa penggunaan tanaman obat atau obat tradisional relatif lebih aman dibandingkan obat sintetis (Herman, dkk., 2019). Daun kelor (*Moringa oleifera*) merupakan salah satu tanaman dalam pengobatan tradisional untuk mengobati berbagai kondisi kesehatan yang banyak tumbuh di wilayah beriklim tropis, termasuk Indonesia. Hampir semua bagian dari tanaman ini bisa diolah menjadi obat herbal.

Inventarisasi tanaman obat untuk penderita diabetes melitus dan hipertensi sudah pernah dilakukan oleh Herman, dkk. (2019) di Desa Minanga, Kecamatan Bambang, Kabupaten Mamasa. Penelitian tersebut menunjukkan 11 jenis tanaman obat

untuk diabetes melitus dan 14 jenis tanaman obat untuk penyakit hipertensi yang digunakan oleh masyarakat di desa termasuk daun kelor.

Daun kelor diketahui mengandung beragam senyawa aktif yang baik untuk tubuh. Selain itu, daun kelor juga kaya akan vitamin dan mineral. Hal inilah yang menjadi alasan mengapa daun kelor sering digunakan sebagai obat herbal untuk mengatasi gangguan kesehatan tertentu, termasuk diabetes. Namun, senyawa daun kelor perlu diketahui karakteristiknya (Marhaeni, 2021).

Analisis kluster merupakan analisis teknik multivariat yang bertujuan untuk melakukan pengelompokan penempatan sekumpulan objek dalam dua atau lebih kelompok berdasarkan kesamaan objek pada berbagai karakteristik (Faradilla, 2022). Terdapat dua jenis metode analisis kluster, yaitu metode *hierarchical* dan metode non *hierarchical*.

Analisis kluster dengan metode *hierarchical* adalah analisis yang pengklasteran datanya dilakukan dengan cara mengukur jarak kedekatan pada setiap objek yang kemudian membentuk sebuah dendrogram. Terdapat beberapa metode dalam *clustering hierarchical* diantaranya yaitu, metode *single linkage*, *complete linkage*, *average linkage*, *centroid method*, dan *ward's method*. Salah satu kelebihan utama analisis *clustering hierarchical* adalah tidak adanya keharusan untuk menetapkan jumlah kluster secara eksplisit diawal analisis (Dewi dan Ahadiyah, 2022).

Clustering hierarchical sebelumnya sudah pernah digunakan oleh Dewi dan Ahadiyah, (2022) pada penentuan kelompok kabupaten/kota di Jawa Timur berdasarkan indikator pendidikan. Metode terbaik pada penelitian ini adalah metode *average linkage* dengan nilai korelasi *cophenetic* sebesar 0,807. Penelitian yang serupa juga dilakukan oleh Matdoan dan Van Delsen (2020) yaitu melakukan analisis kluster dengan metode *clustering hierarchical* untuk klasifikasi kabupaten/kota di Provinsi Maluku berdasarkan Indeks Pembangunan Manusia. Berdasarkan Penelitian ini, diperoleh hasil bahwa metode terbaik dalam analisis yang dilakukan adalah *ward's method*.

Berbeda dengan metode *hierarchical*, metode non *hierarchical* dimulai dengan menentukan terlebih dahulu jumlah kluster yang diinginkan. Setelah jumlah kluster ditentukan, proses kluster baru dilakukan. Pada metode non *hierarchical* terdapat 2 jenis analisis yaitu *clustering K-Means* dan *clustering K-Medoids*. *Clustering K-Medoids* merupakan metode yang dapat mengatasi kelemahan *clustering K-Means* yaitu sensitif terhadap *outlier* karena suatu objek dengan suatu nilai yang besar mungkin secara substansial menyimpang dari distribusi data (Luthfi dan Wijayanto, 2021).

Meiriza dkk. (2023) membandingkan algoritma *clustering K-Means* dan *clustering K-Medoids* untuk pengelompokan Program BPJS Ketenagakerjaan. Hasil dari penelitian ini menunjukkan bahwa *clustering K-Medoids* menghasilkan kelompok yang lebih stabil dan *robust* dibandingkan dengan *clustering K-Means*. Sebelumnya penelitian serupa juga sudah pernah dilakukan Triansyah dkk. (2022) yaitu perbandingan *clustering K-Means* dan *clustering K-Medoids*, namun pada data tingkat stres pada manusia. Hasil dari penelitian ini menunjukkan *K-Medoids* adalah pilihan yang lebih baik daripada *K-Means* untuk pengelompokan tingkat stres pada manusia, karena menghasilkan determinan kovarians yang lebih kecil, menunjukkan bahwa pengelompokan bekerja lebih baik.

Sementara itu, Rahma dan Muhajir (2024) dalam penelitiannya membandingkan metode *clustering hierarchical* dan *clustering K-Medoids* pada data Satuan Kerja Perangkat Daerah (SKPD) berdasarkan jumlah aduan yang sudah ditanggapi dan yang belum ditanggapi pada tahun 2021. Hasil penelitian menunjukkan kedua metode merupakan metode yang sama baik dalam melakukan *clustering* karena diperoleh nilai *silhouette coefficient* pada kedua metode sebesar 0.9146.

Berdasarkan uraian di atas, maka peneliti tertarik untuk melakukan perbandingan antara dua metode pada *clustering* yaitu *clustering K-Medoids* dan *clustering hierarchical* dalam mengelompokkan senyawa dari daun kelor dan obat antidiabetes untuk mengetahui apakah ada senyawa dalam daun kelor yang mirip

dengan obat antidiabetes yang sudah dikenal, yang mungkin akan mendorong penggunaan daun kelor sebagai alternatif atau obat tambahan untuk diabetes.

1.2 Rumusan Masalah

Berdasarkan latar belakang di atas, maka masalah dari penelitian ini adalah sebagai berikut:

1. Bagaimana perbandingan metode *clustering hierarchical* dan *clustering K-Medoids* dalam identifikasi senyawa aktif dari daun kelor yang berpotensi sebagai obat anti diabetes?
2. Bagaimana menentukan senyawa antidiabetes yang potensial dari daun kelor berdasarkan hasil klasterisasi?

1.3 Tujuan Penelitian

1. Untuk mengetahui perbandingan metode *clustering hierarchical* dan *clustering K-Medoids* dalam identifikasi senyawa aktif dari daun kelor yang berpotensi sebagai obat antidiabetes.
2. Untuk mengetahui senyawa antidiabetes yang potensial dari daun kelor berdasarkan hasil klasterisasi.

1.4 Manfaat Penelitian

Temuan dari penelitian ini dapat membantu dalam pengembangan lebih lanjut terhadap senyawa-senyawa dari daun kelor sebagai obat antidiabetes. Dengan mengetahui kelompok senyawa yang paling menjanjikan, penelitian lebih lanjut dapat difokuskan untuk menguji efikasi dan keamanan potensial dari senyawa-senyawa tersebut. Penelitian ini tidak hanya berpotensi untuk memberikan wawasan baru tentang senyawa-senyawa dari daun kelor sebagai obat antidiabetes, tetapi juga dapat memberikan landasan untuk pengembangan lebih lanjut dalam bidang pengobatan herbal dan farmasi.

1.5 Batasan Masalah

Batasan masalah digunakan untuk menghindari terjadinya penyimpangan dalam melakukan penelitian. Adapun batasan masalah dalam penelitian ini adalah sebagai berikut:

1. Data yang digunakan dalam penelitian ini adalah data senyawa-senyawa aktif yang terdapat dalam daun kelor dengan mengelompokkan senyawa aktif obat antidiabetes.
2. Senyawa obat antidiabetes dari obat *Metformin*, *Glimepiride*, dan *Acarbose*.
3. Metode analisis yang digunakan adalah *clustering hierarchical* dan *clustering K-Medoids*.
4. *Software* yang digunakan dalam penelitian ini adalah R studio dan microsoft excel

BAB II

KAJIAN PUSTAKA

2.1 *Clustering*

Secara umum, *clustering* adalah teknik statistik yang digunakan untuk mengelompokkan objek atau data ke dalam kelompok-kelompok yang homogen berdasarkan kesamaan karakteristik atau atribut tertentu. Tujuan dari *clustering* adalah untuk mengidentifikasi pola atau struktur dalam data yang tidak terlihat secara kasat mata (Sindi dkk., 2020).

Clustering merupakan suatu proses pengelompokkan sekumpulan data atau record, observasi, atau mengelompokkan kelas yang memiliki kesamaan objek. Perbedaan *clustering* dengan klasifikasi yaitu tidak adanya variabel target dalam melakukan suatu pengelompokan pada proses *clustering*. *Clustering* sering dilakukan sebagai untuk langkah awal dalam proses data mining saat melakukan suatu metode analisis. Terdapat banyak algoritma *clustering* yang telah digunakan oleh peneliti sebelumnya seperti, *K Means*, *Improved K-Means*, *Fuzzy C-Means*, *Density-Based Spatial Clustering Algorithm with Noise* (DBSCAN), *K-Medoids* (PAM), *Clustering Large Application based on Randomized Search* (CLARANS) dan *Fuzzy Subtractive*. Setiap algoritma memiliki kelebihan dan kekurangan masing-masing, namun prinsip algoritma sama, yaitu mengelompokkan data sesuai dengan karakteristik dan mengukur jarak kemiripan antar data dalam satu kelompok (Sindi dkk., 2020).

2.2 Analisis *Clustering Hierarchical*

Metode *clustering hierarchical* adalah teknik kluster yang dipakai untuk mengelompokkan objek atau elemen secara tersusun berlandaskan kemiripan sifat dan kluster yang diinginkan belum diketahui jumlahnya (Lutfhi dan Wijayanto, 2021). Kemiripan sifat ditentukan dengan mengukur jarak kedekatan antar objek, yaitu jarak *euclidean* yang dirumuskan sebagai berikut (Herman dkk., 2019).

$$d(x, y) = \sqrt{\sum_{i=1}^p (x_{ik} - y_{jk})^2} \quad (2.1)$$

- $d(x, y)$: kuadrat jarak *euclidean* antar dua objek x dan y ,
 x_{ik} : nilai objek x pada variabel ke- k
 y_{jk} : nilai objek y pada variabel ke- k
 p : jumlah total variabel (fitur)
 $\sum_{i=1}^p$: penjumlahan dari variabel ke 1 hingga ke- p
 $(x_{ik} - y_{jk})^2$: selisih nilai antar dua objek pada variabel ke- k , yang dikuadratkan

Terdapat dua metode dari *clustering hierarchical* yaitu metode *agglomerative* (pemusatan) dan metode *divisive* (penyebaran). Metode *agglomerative* (pemusatan) merupakan metode *clustering* data yang diawali dengan per satu pengamatan sebagai klasternya sendiri, kemudian mengelompokkan setiap pengamatan ke dalam kelompok yang semakin besar. Sedangkan, metode *divisive* adalah kebalikan dari metode *agglomerative* yang dimulai dengan semua pengamatan dalam satu klaster besar, kemudian secara bertahap memisahkan pengamatan tersebut menjadi klaster-klaster yang lebih kecil hingga setiap pengamatan menjadi klaster individu (Rachmatin dan Sawitri 2019). *Output* dari metode *agglomerative* dapat divisualisasikan dengan bentuk diagram yang disebut dendrogram yang terdapat tingkatan antara objeknya. Ada 5 metode *agglomerative* dalam pembentukan klaster yaitu sebagai berikut:

1. *Single Linkage* (jarak terdekat), metode klaster ini memberikan hasil jika klaster digabungkan berdasarkan jarak antar objek yang paling dekat antara dua klaster

$$d(ij)k = \min (d_{ik}d_{jk}) \quad (2.2)$$

dengan,

$d(ij)k$: jarak antara klaster (ij) dan klaster k

d_{ik} dan d_{jk} : jarak antara tetangga terdekat klaster i dan k , serta klaster j dan k

2. *Average Linkage* (jarak rata-rata), metode klaster ini menyatukan objek berdasarkan jarak rata-rata pasangan objek masing-masing pada kumpulan antara dua kelompok.

$$d(ij)k = \frac{\sum_n \sum_m d_{nm}}{N_{(ij)}N_k} \quad (2.3)$$

dengan,

d_{nm} : jarak antara objek ke- n dalam klaster (ij) dan objek ke- m dalam klaster k

$N_{(ij)}$ dan N_k : jumlah objek dalam klaster $(i j)$ dan k

3. *Complete Linkage* (jarak terjauh), metode klaster ini menampilkan hasil jika kelompok-kelompok digabungkan berdasarkan jarak antar objek yang paling jauh di antara dua klaster.

$$d(ij)k = \max (d_{ik}d_{jk}) \quad (2.4)$$

4. *Centroid Method*, metode klaster ini menggunakan rata-rata jarak pada setiap kelompok yang dibentuk. Setiap terjadi kelompok baru, maka *centroid* yang baru akan dihitung kembali hingga terbentuk kelompok yang stabil.

$$d(C_i, C_j) = \sqrt{\sum_{m=1}^p \left(\frac{1}{n_i} \sum_{k=1}^{n_i} x_{ikm} - \frac{1}{n_j} \sum_{k=1}^{n_j} x_{jkm} \right)^2} \quad (2.5)$$

dengan,

$d(C_i, C_j)$: jarak antara dua klaster

p : jumlah fitur (dimensi) dari data.

n_i dan n_j : jumlah titik data dalam klaster

x_{ikm} : nilai dari fitur ke- m untuk titik data ke- k dalam klaster C_i

x_{jkm} : nilai dari fitur ke- m untuk titik data ke- k dalam klaster C_j

5. *Ward's method*, metode klaster ini menerapkan perhitungan yang *complete* dan memaksimalkan kehomogenan dalam suatu kelompok. ukuran yang digunakan adalah *Sum of Square* (SSE).

$$SSE = \sum_{i=1}^n (X_i - \bar{x})' (X_i - \bar{x}) \quad (2.6)$$

dengan,

X_i : vektor data untuk titik ke $-i$

\bar{x} : vektor rata-rata dari dari semua titik data dalam klaster.

$(X_i - \bar{x})$: vektor yang menunjukkan deviasi titik data ke- i dari rata rata klaster.

$(X_i - \bar{x})'$: *transpose* dari vektor $(X_i - \bar{x})$.

$\sum_{i=1}^n$: penjumlahan semua titik data i sampai n .

2.3 Analisis *Clustering K-Medoids*

Clustering K-medoids merupakan suatu algoritma yang digunakan untuk menemukan *medoids* didalam sebuah kelompok (klaster) yang merupakan titik pusat dari suatu kelompok (klaster). Algoritma *K-Medoids* lebih baik dibandingkan dengan *K-Means* karena pada *K-Medoids* kita menemukan k sebagai objek yang representatif untuk meminimalkan jumlah ketidaksamaan objek data, sedangkan pada *K-Means* menggunakan jumlah jarak *euclidean distances* untuk objek data.

Langkah-langkah algoritma *K-Medoids* sebagai berikut (Sindi dkk., 2020).

1. Inisialisasi pusat klaster sebanyak k (jumlah klaster).
2. Mengalokasikan setiap data (objek) ke klaster terdekat menggunakan ukuran jarak *euclidean*
3. Menentukan objek secara acak dalam setiap klaster sebagai kandidat medoid baru.
4. Mengevaluasi jarak setiap objek dalam klaster terhadap medoid yang baru terpilih.
5. Menghitung total simpangan (S) dengan menghitung nilai total distance baru – total distance lama, Jika $S < 0$, maka ganti objek dengan data klaster untuk memperoleh sekelompok k objek yang baru sebagai medoid.
6. Melakukan iterasi tahap ke 3 sampai 5 hingga tidak terjadi perubahan medoid, sehingga didapatkan klaster beserta anggota klaster masing-masing. Kemudian untuk mendapatkan nilai k di sebuah data yang ada di *clustering K-Medoids* dilakukan di dalam proses *clustering*.

2.4 Validitas *Clustering*

Pengujian validitas untuk memilih di antara lima metode yang ada pada *clustering hierarchical* salah satunya adalah dengan menggunakan koefisien korelasi

cophenetic. Koefisien korelasi *cophenetic* adalah koefisien korelasi antar elemen asli matriks jarak *euclidean* dengan elemen yang disajikan oleh dendrogram (*matriks cophenetic*). Nilai koefisien korelasi *cophenetic* berada di antara angka -1 dan 1 , jika nilai koefisien korelasi *cophenetic* mendekati 1 bermakna bahwa hasil dari proses *clustering* cukup baik (Herman dkk., 2019).

Pengujian validitas antara metode *clustering hierarchical* dan *clustering K-Medoids* salah satunya menggunakan uji validasi internal, yang merupakan pengujian data kluster tanpa membutuhkan informasi eksternal. Contoh dari validasi ini adalah *silhouette index*, *connectivity*, dan *dunn index*. Metode validasi *silhouette index* adalah ukuran validasi yang dapat mengevaluasi peletakan tiap objek pada setiap kluster dengan mengkomparasikan antara jarak rata-rata objek dalam satu kluster dan jarak antara objek dengan kluster yang berlainan (Herman dkk., 2019). Metode *connectivity* adalah melihat seberapa terhubung suatu titik data tetangga dengan kluster yang sama sedangkan metode *dunn index* adalah rasio jarak terkecil antara observasi pada kluster yang berbeda jarak terbesar pada masing masing kluster (Irwansyah dan Faisal, 2015).

Silhouette coefficient atau koefisien *silhouette* berada di antara angka -1 hingga 1 . Nilai tersebut dipakai untuk menentukan kedekatan kemiripan objek yang dikelompokkan dalam suatu kluster. Kluster semakin baik jika rata-rata nilai *silhouette coefficient* mendekati 1 . Sedangkan kluster dikatakan tidak baik jika nilai rata-rata *silhouette coefficient* mendekati -1 (Farissa dkk., 2021).

Metode yang menunjukkan seberapa dekat suatu objek dengan klasternya dibandingkan dengan kluster lainnya dikenal sebagai analisis *Silhouette*. rata-rata yang diperoleh dengan analisis *silhouette* menunjukkan seberapa optimal jumlah kluster yang dibuat. Rata-rata yang lebih tinggi menunjukkan bahwa jumlah k kluster optimal (Januzaj dkk., 2023).

Metode *elbow* atau *within cluster sums of square* merupakan salah satu metode yang dapat digunakan untuk menentukan jumlah kluster terbaik, yaitu dengan cara melihat persentase setiap kluster yang akan membentuk siku pada suatu titik tertentu.

Metode *elbow* biasa disajikan dalam bentuk grafik untuk mengetahui lebih jelas siku yang terbentuk. Tujuan dari metode *elbow* adalah untuk memilih nilai k yang kecil dan masih memiliki nilai *withinss* yang rendah (Qusyairi dkk., 2024)

Metode *Gap statistics* merupakan teknik yang dapat digunakan untuk menangani masalah dengan menentukan jumlah kluster optimal dalam pengklasteran. *Gap statistics* memperkenalkan nilai *Gap* (k) yang merupakan hasil pengurangan nilai rata-rata logaritma *within cluster sum of square* yang direplikasi pada dataset yang telah dibangkitkan dengan nilai logaritma *within cluster sum of square* dari data asli (Barokah, 2023).

2.5 Z-Score Standardization

Z-score adalah metode standardisasi yang digunakan untuk mengubah data menjadi skala baku dengan rata-rata nol dan simpangan baku satu. Proses ini sangat penting dalam prapemrosesan data, khususnya untuk metode klustering yang bergantung pada penghitungan jarak antar data. Dengan menggunakan *Z-score*, setiap variabel akan memiliki kontribusi yang setara terhadap hasil kluster, terlepas dari skala awalnya. Menurut Han dkk., 2012, *Z-score* standardization dilakukan dengan rumus:

$$Z = \frac{X - \mu}{\sigma} \quad (2.7)$$

Z : Nilai *Z-score*, yaitu standardisasi dari nilai data asli. Menunjukkan seberapa jauh nilai tersebut dari rata-rata dalam satuan standar deviasi

X : Nilai data asli (observasi individual).

μ : Rata-rata dari seluruh data (mean).

σ : Simpangan baku (standar deviasi) dari seluruh data.

2.6 Absorption, Distribution, Metabolism, Excretion, and Toxicity

Absorption, Distribution, Metabolism, Excretion, and Toxicity atau yang biasa disingkat ADMET merupakan kumpulan data yang berisi profil farmakokinetik suatu senyawa (molekul obat) dan memainkan peran penting dalam menentukan aktivitas farmakodinamiknya (Srivastava dkk., 2020). Dimana variabel yaitu *Water Solubility*, *Caco2 Permeability*, *Intestinal Absorption*, *Skin Permeability*, *VDss*, *Fraction*

Unbound, BBB Permeability, CNS Permeability (log Ps), Total Clearance, Max. Tolerated Dose, Oral Rat Acute Toxicity (LD50), Oral Rat Chronic Toxicity (LOAEL), TPyriformis Toxicity, dan Minnow Toxicity.

2.7 Diabetes

2.7.1 Pengertian Diabetes

Diabetes adalah penyakit kronis yang ditandai dengan kadar gula darah yang tinggi. Glukosa merupakan sumber energi utama bagi sel manusia. Glukosa yang menumpuk di dalam darah karena tidak diserap dengan baik oleh sel-sel tubuh dapat menyebabkan berbagai penyakit organ. Jika diabetes tidak dikontrol dengan baik, dapat timbul berbagai komplikasi yang membahayakan nyawa penderita. Kadar gula dalam darah dikendalikan oleh hormon insulin yang diproduksi oleh pankreas, yaitu organ yang terletak di belakang lambung. Pada penderita diabetes, pankreas tidak mampu memproduksi insulin sesuai kebutuhan tubuh. Tanpa insulin, sel-sel tubuh tidak dapat menyerap glukosa dan mengubahnya menjadi energi (Priyanto dkk., 2022).

2.7.2 Jenis Jenis Diabetes

Secara umum diabetes dibedakan menjadi dua jenis, yaitu diabetes tipe 1 dan tipe 2. Diabetes tipe 1 terjadi karena sistem kekebalan tubuh penderita menyerang dan menghancurkan sel-sel pankreas yang memproduksi insulin. Sedangkan diabetes tipe 2 adalah bentuk diabetes yang lebih umum. Diabetes tipe ini disebabkan oleh sel-sel dalam tubuh menjadi kurang sensitif terhadap insulin dan dengan demikian tidak dapat menggunakan insulin yang dihasilkan dengan baik (Priyanto dkk., 2022).

2.7.3 Pengobatan Diabetes

Pengobatan pada penderita diabetes harus mengatur pola makannya dengan memperbanyak konsumsi protein dari buah-buahan, sayur-sayuran, biji-bijian dan makanan rendah kalori, rendah lemak, juga dapat diganti dengan sorbitol, pemanis yang lebih aman bagi pasien. Penderita diabetes dan keluarganya dapat berkonsultasi dengan dokter atau ahli gizi untuk melakukan penyesuaian pola makan sehari-hari (Priyanto dkk., 2022).

Pengobatan diabetes juga dapat dilakukan dengan penggunaan obat medis dari dokter seperti obat *metformin*, *glimpirede*, *acarbose*, dan jenis lainnya (Meryta dkk., 2023). Pemanfaatan tanaman obat sebagai pengobatan alternatif masih banyak digunakan oleh masyarakat terutama dari kalangan menengah kebawah termasuk pengobatan pada diabetes. Tanaman obat yang digunakan juga bermacam-macam termasuk daun kelor (Herman dkk., 2019).

2.7.4 Pencegahan Diabetes

Diabetes tipe 1 tidak dapat dicegah karena pemicunya belum diketahui. Diabetes tipe 2 dan diabetes gestasional kini dapat dicegah dengan gaya hidup sehat. Beberapa hal yang dapat dilakukan untuk mencegah diabetes yaitu mengatur frekuensi dan menu makanan menjadi lebih sehat, menjaga berat badan ideal, rutin berolahraga dan rutin menjalani pengecekan gula darah, setidaknya sekali dalam setahun (Priyanto dkk., 2022).

2.8 Daun Kelor

2.8.1 Pengertian Kelor

Tanaman kelor adalah salah satu tanaman perdu dengan ketinggian 1-7 meter, tahan terhadap musim kering dengan toleransi terhadap kekeringan sampai 6 bulan serta mudah dibiakkan dan tidak memerlukan perawatan yang intensif. Di Indonesia, tanaman kelor memiliki beragam nama di beberapa wilayah di antaranya kelor (Jawa, Sunda, Bali, Lampung), maronggih (Madura), moltong (Flores), kelo (Bugis), ongge (Bima), murong atau barungai (Sumatera) dan hau fo (Timur) (Marhaeni, 2021).

2.8.2 Zat Gizi Kelor

Salah satu bagian dari tanaman kelor yang telah banyak diteliti kandungan gizi dan kegunaannya baik untuk bidang pangan dan kesehatan adalah bagian daun. Di bagian tersebut terdapat ragam nutrisi, di antaranya kalsium, besi, protein, vitamin A, vitamin B dan vitamin C (Marhaeni, 2021).

Kandungan zat gizi daun kelor lebih tinggi jika dibandingkan dengan sayuran lainnya yaitu berada pada kisaran angka 17,2 mg/100 g Selain itu, di dalam daun kelor

juga terdapat kandungan berbagai macam asam amino, antara lain asam amino yang berbentuk asam aspartat, asam glutamat, alanin, valin, leusin, isoleusin, histidin, lisin, arginin, venilalanin, triptopan, sistein, dan methionin Kandungan fenol dalam daun kelor segar sebesar 3,4% sedangkan pada daun kelor yang telah diekstrak sebesar 1,6% . Penelitian lebih lanjut menyebutkan bahwa daun kelor mengandung antioksidan tinggi dan antimikrobia. Hal ini disebabkan oleh adanya kandungan asam askorbat, flavonoid, phenolic, dan karatenoid (Marhaeni, 2021).

BAB V

KESIMPULAN

5.1 Kesimpulan

1. Berdasarkan hasil kluster dengan dua metode menggunakan *clustering K-Medoids* dan *clustering hierarchical* diperoleh jumlah kluster optimal yang terbentuk pada *clustering K-Medoids* yaitu 8 dan *clustering hierarchical* yaitu 10. Berdasarkan hasil analisis kluster, metode *clustering hierarchical* dengan pendekatan *ward* menghasilkan *silhouette coefficient* sebesar 0,1340, *dunn index* sebesar 0,3854 dan *connectivity* sebesar 20,500. Sementara itu, metode *clustering K-Medoids* menghasilkan *silhouette coefficient* sebesar 0,1000 *dunn index* sebesar 0,2831 dan *connectivity* sebesar 38,500, dimana semakin tinggi nilai *silhouette coefficient* dan *dunn Index*, serta semakin rendah nilai *connectivity*, maka kualitas *clustering* dianggap semakin baik. Sehingga metode *clustering hierarchical* dengan *ward's method* merupakan metode terbaik untuk melakukan *clustering* pada data ADMET senyawa daun kelor dan obat antidiabetes.
2. Penerapan *clustering* untuk memprediksi potensi senyawa daun kelor sebagai obat antidiabetes dilakukan dengan melakukan pengelompokan kluster-kluster terhadap senyawa daun kelor dan senyawa obat antidiabetes berdasarkan variabel *Water Solubility, Caco2 Permeability, Intestinal Absorption, Skin Permeability, VDss, Fraction Unbound, BBB Permeability, CNS Permeability (log Ps), Total Clearance, Max. Tolerated Dose, Oral Rat Acute Toxicity (LD50), Oral Rat Chronic Toxicity (LOAEL), TPyriformis Toxicity, dan Minnow Toxicity*. Pada penelitian ini, dari beberapa kluster yg terbentuk baik menggunakan *clustering hierarchical* maupun *clustering K-Medoids*, terdapat 1 kluster yang berisikan senyawa daun kelor dan antidiabetes. Senyawa antidiabetes yang dimaksud adalah *glimepiride*. Hal ini menunjukkan, senyawa daun kelor dan antidiabetes dalam kluster yang sama memiliki karakteristik yang serupa sehingga dapat disimpulkan bahwa terdapat senyawa daun kelor yang berpotensi sebagai obat antidiabetes. Pada

metode *clustering hierarchical*, senyawa antidiabetes yang potensial dari daun kelor adalah 358357, 191036, 1301760, 1930553. Sementara pada metode *clustering K-Medoids*, diperoleh senyawa antidiabetes yang potensial dari daun kelor adalah 358357, 102564, 934955, 191036, 1301760, 1895718, 1930553, 1921979, 4030961.

5.2 Saran

Hasil penelitian yang telah dilakukan. Penelitian menyarankan untuk mengembangkan dan mempelajari algoritma pengelompokan tambahan seperti *agglomerative clustering* atau *Density-Based Spatial Clustering Algorithm with Noise* (DBSCAN) untuk menentukan apakah ada metode yang dapat memberikan hasil pengelompokan senyawa yang lebih baik terlepas dari bahwa *clustering hierarchical* adalah metode yang paling efektif dalam penelitian ini.

DAFTAR PUSTAKA

- Barokah, M. S., 2023, Penentuan Jumlah Kluster Optimal Menggunakan Metode Gap Statistik, *Skripsi*, Program Pasca Sarjana Statistika, Univ. Gadjah Mada, Yogyakarta.
- Dewi, A. F., dan Ahadiyah, K., 2022, *Agglomerative Hierarchy Clustering* pada Penentuan Kelompok Kabupaten/Kota di Jawa Timur Berdasarkan Indikator Pendidikan, *Zeta Math Journal*, No.2, Vol.7, 57-63, : <https://doi.org/10.31102/zeta.2022.7.2.57-63>.
- Faradilla, S. B., 2022, Komparasi Analisis *K-Medoids Clustering* dan *Hierarchical Clustering*, (Studi Kasus : Data Kriminalitas di Indonesia Tahun 2020), *Skripsi*, Program Pasca Sarjana Statistika, Univ. Islam Indonesia, Yogyakarta.
- Farissa, R. A., Mayasari, R., dan Umaidah, Y., 2021, Perbandingan Algoritma *K-Means* dan *K-Medoids* untuk Pengelompokkan Data Obat dengan *Silhouette Coefficient* di Puskesmas Karangasambung, *Journal of Applied Informatics and Computing*, No.2, Vol.5, 109 -116, : <https://jurnal.polibatam.ac.id/index.php/JAIC/article/view/3237>.
- Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques* (3rd ed.). Elsevier.
- Herman, Murniati, dan Nur, A. S. S., 2019, Inventarisasi Tanaman Obat Tradisional untuk Penderita Diabetes Melitus dan Hipertensi Di Desa Minanga Kecamatan Bambang Kabupaten Mamasa, *Jurnal Farmasi Sandi Karsa*, No.1, Vol.5, 26-23, : <https://media.neliti.com/media/publications/286092-inventarisasi-tanaman-obat-tradisional-u-d96507a6.pdf>.
- Irwansyah, E., dan Faisal, M., 2015, *Advanced Clustering Teori dan Aplikasi* Edisi 1, Penerbit Deepublish, Yogyakarta
- Januzaj, Y., Beqiri, E., dan Luma, A., 2023, *Determining the Optimal Number of Clusters using Silhouette Score as a Data Mining Technique*, *International Journal of Online and Biomedical Engineering (iJOE)*, No.4, Vol.19, 174-182, : <https://www.ijoes.in/>.
- Luthfi, E., dan Wijayanto, A. W., 2021, Analisis Perbandingan Metode *Hierarchical*, *K-Means*, dan *K-Medoids Clustering* dalam Pengelompokkan Indeks Pembangunan Manusia Indonesia, *Inovasi: Jurnal Ekonomi, Keuangan, dan Manajemen*, No.4, Vol.17, 761-773, : <https://www.semanticscholar.org/paper/3d7185cac3f4bead333dbc2e95db0790b8684405>.

- Marhaeni, L. S., 2021, Daun Kelor (*Moringa oleifera*) sebagai Sumber Pangan Fungsional dan Antioksidan, *AGRISIA-Jurnal Ilmu-Ilmu Pertanian*, No.2, Vol.13, 0091-2302; <https://ejournal.borobudur.ac.id/index.php/3/article/view/882>.
- Matdoan, M. Y., dan Van Delsen, M. S. N., 2020, Penerapan Analisis Cluster dengan Metode Hierarki Untuk Klasifikasi Kabupaten/Kota Di Provinsi Maluku Berdasarkan Indikator Indeks Pembangunan Manusia, *Statmat: jurnal statistika dan matematika*, No. 2, Vol.2, 123-130, : <https://openjournal.unpam.ac.id/index.php/sm/article/view/4740/0>.
- Meiriza, A., Ali, E., Rahmiati, dan Agustin., 2023, Perbandingan Algoritma *K-Means* dan *K-Medoids* untuk Pengelompokan Program BPJS Ketenagakerjaan, *Indonesian Journal of Computer Science*, No.2, Vol.12, 714-728, : <https://www.researchgate.net/publication/358970135>.
- Rachmatin, D., dan Sawitri, K., 2019, Perbandingan antara Metode *Agglomeratif*, Metode *Divisif* dan Metode *K-Means* dalam Analisis Klaster, *Seminar Nasional Matematika UNPAR*, 9-1, : <http://eprints.itenas.ac.id/157/>.
- Rahma, M., dan Muhajir, M., 2024, Komparasi *Hierarchical Clustering* dan *K-Medoid Clustering* Terhadap Pengelompokan Satuan Kerja Perangkat Daerah (SKPD) Di Kabupaten Sleman Tahun 2021, *Emerging Statistics and Data Science Journal*, No.1, Vol.2, 41-52, : <https://typeset.io/documents/recent>.
- Priyanto, P., Yulianingsih, N., dan Asyari, H., 2022, Hubungan Pengetahuan Tentang Diabetes Mellitus dengan Kepatuhan Menjalani Pengobatan pada Pasien Diabetes Melitus di Kecamatan Kertasemaya Tahun 2021, *Jurnal Pengabdian Ilmu Kesehatan*, No.1, Vol.2, 17-24, : <https://ejournal.politeknikpratama.ac.id/index.php/JPIKes/article/view/337>.
- Sindi, S., Ningse, W. R. O., Sihombing, I. A., Zer, F. I. R., dan Hartama, D., 2020, Analisis Algoritma *K-Medoids Clustering* dalam Pengelompokan Penyebaran Covid-19 di Indonesia, *Jurnal Teknologi Informasi*, No.1, Vol.4, 166-173, : <http://jurnal.una.ac.id/index.php/jurti/article/view/1296>.
- Srivastava, V., Yadav, A., dan Sarkar, P., 2020, *Molecular docking and ADMET study of bioactive compounds of Glycyrrhiza glabra against main protease of SARS-CoV2*, *Materials Today: Proceedings*, No.8, Vol.49, 2999-3007, : <https://par.nsf.gov/servlets/purl/10297452>.

Triansyah, A., Herwindiati, D. E., dan Hendryli, J., 2022, Perbandingan *K-Means* dan *K-Medoids* untuk Klastering Tingkat Stess pada Manusia, *Jurnal Ilmu Komputer dan Informasi*, No. 2, Vol 10, : <http://dx.doi.org/10.24912/jiksi.v10i2.22534>.

Qusyairi, M., Hidayatullah, Z., dan Sandi, A., 2024, Penerapan *K-Means Clustering* dalam Pengelompokan Prestasi Siswa dengan Optimasi Metode *Elbow*, *Infotek: Jurnal Informatika dan Teknologi*, No.2, Vol.7, 500–510, : <https://doi.org/10.29408/jit.v7i2.26375>.

World Health Organization (WHO), 2021, Diabetes, <https://www.who.int/news-room/fact-sheets/detail/diabetes>, diakses tgl 24 Agustus 2024.