# PERBANDINGAN ALGORITMA *K-MEANS* DAN *K-MEDOIDS* DALAM PENGELOMPOKAN KABUPATEN DAN KOTA BERDASARKAN TANAMAN BIOFARMAKA

**Arif Miftahul Anwar, Ade Rizki Rinaldi, Mulyawan**
Teknik Informatika, STMIK IKMI CIREBON
Jalan Perjuangan No. 10 B Majasem Kec. Kesambi Kota Cirebon
*Arifmiftah520@gmail.com*

**ABSTRAK**

Tumbuhan biofarmaka dikenal juga dengan tumbuhan obat yang mempunyai keunggulan penting dalam pengembangan industri farmasi, dan produk kesehatan yang diperoleh dari berbagai bagian tumbuhan seperti daun, batang, buah, dan akar. Salah satu strategi untuk meningkatkan dan mempertahankan produktivitas tanaman biofarmaka dengan membagi sektor tanaman menjadi beberapa kelompok berdasarkan kesamaan karakteristik. Salah satu pendekatan untuk mensegmentasi area ini dengan menggunakan analisis *cluster*. Tujuan dari penelitian ini adalah untuk mengetahui hasil *cluster* yang terbentuk dan penggunaan algoritma *K-Means* dan *K-Medoids* untuk mengelompokkan kabupaten/kota berdasarkan variabel biomedis tanaman. Dari hasil analisis *cluster* menggunakan *K-Means* menghasilkan 24 kabupaten/kota pada *cluster* 1.1 kabupaten pada *cluster* 2.1 kabupaten pada *cluster* 3, dan 1 kabupaten pada *cluster* 4. Sedangkan dengan menggunakan *K-Medoids*, *cluster* 1 menghasilkan 2 kabupaten, *cluster* 2 menghasilkan 1 kabupaten, *cluster* 3 menghasilkan 1 kabupaten dan *cluster* 4 memiliki 23 kabupaten/kota. Dari hasil pengelompokan kedua metode tersebut, ditentukan metode terbaik melalui perbandingan menggunakan *cluster variance* dalam pengelompokan produksi tanaman biofarmaka. Mencapai nilai DBI. 0.087 untuk algoritma *K-Means* dan 0.428 untuk algoritma *K-Medoids*. Hal ini menunjukkan bahwa pada pengujian DBI algoritma *K-Means* lebih baik dibandingkan algoritma *K-medoids* karena semakin rendah nilai yang diperoleh maka kualitas *cluster* yang dihasilkan semakin baik.

***Kata kunci**: Tanaman Biofarmaka, Analisis Cluster, Cluster K-Means, Cluster K-Medoids*

## 1. PENDAHULUAN

Di Indonesia tanaman obat sering digolongkan sebagai tanaman biofarmaka. Kategori tanaman biofarmaka ini mencakup 15 tanaman antara lain jahe, laos/lengkuas, kencur, kunyit, lampuyang, jahe, temlen, temu kunci, dlingo/dringo, kapulaga, mengkudu/pace, makota dewa, kejibeling, sambiloto, dan lidah buaya. [1]. Tanaman biofarmaka merujuk pada tumbuhan yang memiliki fungsi untuk mencegah dan mengobati penyakit. Tumbuhan ini umumnya dapat dimanfaatkan sebagai obat tradisional atau sebagai bahan baku untuk pembuatan obat kimiawi. Salah satu bagian tanaman yang sering digunakan untuk keperluan ini adalah daun. Daun-dedaunan seringkali memiliki kesamaan dalam warna dan bentuk.

Tanaman biofarmaka mempunyai manfaat yang besar dalam memenuhi kebutuhan manusia. Dalam industri farmasi, tanaman obat merupakan sumber utama bahan baku obat tradisional dan modern. Saat ini terdapat kecenderungan masyarakat untuk mengadopsi pengobatan tradisional. Hal ini disebabkan adanya peralihan ke gaya hidup yang lebih alami *(back to nature)* dan semakin mahalnya biaya pengobatan modern. Akibatnya, permintaan terhadap tanaman obat semakin meningkat tidak hanya di Indonesia tetapi juga di seluruh dunia. Tanaman obat mungkin tidak sepopuler jenis tanaman lainnya, seperti tanaman penghasil pangan seperti buah-buahan dan umbi-umbian, namun sangat populer terutama dalam konteks perubahan gaya hidup global *(back to nature)*. Tidak hanya berpengaruh pada pola konsumsi masyarakat, tetapi juga telah merambah ke berbagai sektor lainnya, termasuk pengobatan. Secara keseluruhan, terjadi perubahan signifikan dalam pola pengobatan masyarakat secara global, dengan meningkatnya minat terhadap obat-obatan tradisional yang terbuat dari bahan alam [1]. Dalam konteks ini, analisis data mengenai produksi tanaman biofarmaka dapat memberikan wawasan penting bagi industri farmasi dan peneliti dalam mengidentifikasi pola-pola produksi yang efisien dan berpotensi tinggi. Salah satu pendekatan analisis yang digunakan adalah klastering, yang memungkinkan pengelompokan tanaman berdasarkan karakteristik tertentu seperti pertumbuhan, kandungan senyawa, dan faktor-faktor lain yang memengaruhi produksi [1].

## 2. TINJAUAN PUSTAKA

Hasil literature review yang telah dilakukan pada jurnal-jurnal penelitian terkait topik Perbandingan Algoritma *K-means* dan *K-medoids* Terhadap Produksi Tanaman Biofarmaka Di Jawa Barat dapat dijabarkan sebagai berikut :

Paper [1] menjelaskan perbandingan algoritma *K-means* dan *K-medoids* dalam pengelompokan bahan tanaman biofarmasi di Jawa Timur. Tujuan dari penelitian ini adalah untuk memperjelas hasil analisis deskriptif data tanaman biofarmasi dan memahami secara mendalam situasi dan karakteristik tanaman

biofarmasi. Metode *K-means* dan *K-medoids* kemudian digunakan untuk mengelompokkan bahan tanaman biofarmasi untuk memberikan wawasan lebih lanjut mengenai pola distribusi dan kemungkinan pengelompokan di dalamnya. Perbandingan metode *K-means* dan *K-medoids* untuk *clustering* produk ini juga akan membantu menentukan metode yang paling tepat untuk keperluan analisis lebih lanjut. Hasil *clustering* yang dihasilkan metode *K-means* dan *K-medoids* serupa. Klaster 1 terdiri dari dua kabupaten, klaster 2 terdiri dari 32 kabupaten/kota, dan klaster 3 hanya berisi satu kabupaten.

Paper [2] membahas tentang perbandingan algoritma K-means dan K-medoids dalam mengklasifikasikan kasus Covid-19 di Indonesia. Jurnal ini menjelaskan langkah-langkah penelitian, termasuk akuisisi dan transformasi data, serta penggunaan algoritma yang digunakan (kesalahan jumlah kuadrat, koefisien siluet, indeks Davis-Bouldin, dll). Hasil penelitian menunjukkan bahwa K-means mempunyai kinerja yang baik dalam menciptakan cluster yang optimal. Jurnal ini bertujuan untuk memberikan informasi sebaran kasus Covid-19 di Indonesia dengan menggunakan teknik clustering.

Paper [3] menjelaskan perbandingan algoritma clustering K-means dan K-medoids dalam mengklasifikasikan siswa dengan gangguan jiwa. Tujuan penelitian ini adalah untuk menemukan metode yang optimal dalam mengelompokkan siswa tunagrahita ringan, sedang, dan berat. Penelitian menunjukkan bahwa K-means lebih efektif dalam mengklasifikasikan siswa dengan gangguan jiwa. Namun penelitian ini mempunyai beberapa permasalahan. Metode ini menggunakan kumpulan data yang relatif kecil dan tidak mempertimbangkan faktor lain yang mungkin memengaruhi klasifikasi. Penelitian ini merekomendasikan penggunaan dataset yang lebih besar untuk penelitian selanjutnya.

Paper [4] menjelaskan penerapan algoritma K-means dan K-medoids untuk mengelompokkan data negara berdasarkan faktor sosial ekonomi dan kesehatan. Tujuan dari jurnal penelitian ini adalah untuk mengelompokkan data nasional berdasarkan faktor sosial ekonomi dan kesehatan dengan menggunakan algoritma K-means dan K-medoids. Hasil penelitian menunjukkan bahwa algoritma K-means dengan nilai K=5 mengungguli K-medoids dengan nilai Davies-Bouldin Index (DBI) sebesar 0,095. Oleh karena itu, cluster optimal ditentukan dengan menggunakan algoritma K-means. Studi ini akan membantu Anda memanfaatkan data kelompok negara secara strategis dan efektif.

Paper [5] membahas peramalan impor bahan tanaman biofarmaka menggunakan tiga algoritma penambangan data: Naive Bayes, SVM, dan regresi linier. Penelitian ini menggunakan data Badan Pusat Statistik pada tahun 2014 hingga 2019. Hasil perbandingan menunjukkan bahwa algoritma SVM memiliki akurasi tertinggi sebesar 93,3% dan nilai AUC sebesar 1. Prediktor terbaik impor bahan tanaman biofarmaka adalah jahe, jahe, dan dolingo.

Paper [6] Membahas perbandingan algoritma clustering K-means dan K-medoids untuk pengelompokan program pekerja bukan penerima upah pada program BPJS Ketenagakerjaan. Penelitian ini bertujuan untuk menemukan algoritma clustering yang lebih efektif dalam program pengelompokan karyawan non-gaji. Hasil penelitian menunjukkan bahwa algoritma K-medoids menghasilkan kelompok yang lebih stabil dan kokoh dibandingkan dengan K-means.

Paper [7] menjelaskan penggunaan algoritma clustering K-means dan K-medoids untuk mengelompokkan kejahatan berdasarkan provinsi di Indonesia. Tujuannya adalah untuk menentukan apakah area tersebut memerlukan pemantauan tambahan. Data yang digunakan dalam penelitian ini diperoleh dari Badan Pusat Statistik Indonesia. Algoritma K-means menghasilkan 6 anggota pada cluster dengan tingkat kriminalitas tinggi dan 28 anggota pada cluster dengan tingkat kriminalitas rendah, sedangkan algoritma K-medoids menghasilkan 7 anggota pada cluster dengan tingkat kriminalitas tinggi dan 27 anggota pada cluster dengan tingkat kriminalitas rendah. dihasilkan. Kinerja setiap algoritma bergantung pada data yang diproses.

Paper [8] Membahas tentang Perbandingan algoritma K-means dan K-medoids untuk mengelompokkan armada truk berdasarkan produktivitas. Proses evaluasi kinerja armada secara manual memakan waktu dan tidak akurat, sehingga memerlukan teknik pemrosesan data yang lebih cepat dan akurat. Teknik data mining dengan menggunakan teknik clustering diterapkan pada armada kelompok berdasarkan produktivitas. Indeks Davies-Bouldin digunakan untuk mengevaluasi validitas hasil clustering. Algoritma K-means dipilih untuk aplikasi web pengelompokan armada karena skor efektivitas DBI yang rendah. Kami menguji hasil pengelompokan dan menemukan tingkat kesesuaian sebesar 97%.

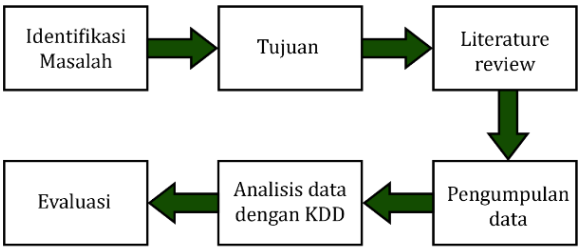
Paper [9] Membahas perbandingan algoritma K-means dan K-medoids untuk mengelompokkan data dari daerah miskin. Penelitian ini bertujuan untuk mengatasi masalah kemiskinan dengan menggunakan teknik data mining. Menerapkan data mining dapat membantu mengelompokkan data komunitas yang buruk. Metode yang cocok untuk ini adalah algoritma K-means dan K-medoids. Hasil penelitian menunjukkan bahwa kedua algoritma ini dapat mengelompokkan data komunitas nakal secara akurat dan efisien. Penelitian ini juga menyajikan langkah-langkah dan perhitungan yang terlibat dalam implementasi algoritma K-means.

Paper [10] menjelaskan perbandingan algoritma K-means dan K-medoids untuk mengelompokkan data obat menggunakan koefisien siluet. Tujuan dari jurnal penelitian ini adalah untuk mengelompokkan data obat untuk puskesmas dan mendukung perencanaan dan distribusi obat. Hasil penelitian

menunjukkan bahwa algoritma K-means memiliki koefisien siluet yang lebih tinggi dan pengelompokan yang lebih baik dibandingkan dengan K-medoids. Hasil klasterisasi dapat dijadikan acuan rekomendasi pengelolaan persediaan obat di puskesmas.

## 3. METODE PENELITIAN

Metode yang digunakan untuk menyelesaikan permasalahan yang ada adalah dengan mendemonstrasikan bagaimana model data mining dapat digunakan untuk menentukan hasil perbandingan produksi tanaman biofarmaka dengan menggunakan algoritma *K-means* dan *K-medoids*. Tahap metodologi penelitian diawali dengan identifikasi masalah, pengumpulan data, dan analisis data dengan menggunakan data mining dan evaluasi.




Gambar 1. Tahapan Metode Penelitian

Pada penelitian ini metode yang digunakan adalah metode *Knowledge Discovery Process* (KDD) sebagai proses dimana teknik *data mining* digunakan guna mengidentifikasi informasi serta pola yang relevan dalam sekumpulan data, proses *data mining* ini melibatkan penggunaan algoritma untuk menemukan pola yang dapat mengidentifikasi dalam suatu proses database.

Dimulai dari tahap seleksi data yaitu mencakup pengumpulan dataset dan pemilihan atribut yang diperlukan, kemudian pembersihan data, penggabungan atribut menjadi satu dataset, standarisasi data, *preprocessing* data, proses pengolahan atau teknik *Clustering* menggunakan perbandingan algoritma *K-means dan K-medoids,* kemudian mengevaluasi DBI dan menentukan jumlah *cluster* yang dihasilkan dari kedua algoritma tersebut. Berikut adalah beberapa langkah dalam prosses *Knowledge Discovery in Database Process* (KDD):

a. *Data selection*
   Mengumpulkan data dengan mengunjungi web resmi BPS Jawa Barat dan mengidentifikasikan sumber-sumber data sekunder yang relevan dengan topik atau penelitian. Sumber data sekunder dapat berupa jurnal ilmiah, laporan pemerintah, basis data, buku, artikel berita, dan lain sebagainya. Setelah data diperoleh dari aktivitas pengumpulan data, langkah selanjutnya adalah melakukan pemilihan atribut yang sesuai dengan apa yang dibutuhkan seperti jumlah ternak, jenis ternak, wilayah geografis, tahun, dan atribut lainnya yang relevan.

*b. Data Preprocessing*
   Tahap preprocessing melakukan proses pembersihan data yang meliputi penghapusan atribut asing dan kolom yang berisi nilai null yang tidak diperlukan dalam perhitungan selanjutnya. Hal ini untuk memastikan bahwa data yang diolah benar-benar data yang relevan. Setelah selesai melakukan proses pembersihan pada setiap subdataset, langkah selanjutnya adalah melakukan integrasi data, yaitu menggabungkan beberapa kumpulan data parsial menjadi kumpulan data lengkap yang siap untuk proses penghitungan pengelompokan.

c. *Transformation*
   Setelah tahap *preprocessing*, langkah berikutnya adalah Proses transformasi data. Data transformasi dilakukan dengan memberikan inisialisasi terhadap data yang memiliki nilai nominal menjadi bernilai numerik [11]. Transformasi data merujuk pada proses mengubah beberapa subdataset menjadi satu format yang seragam, bertujuan untuk menyederhanakan proses *data mining*. Selain itu, terdapat juga penyesuaian pada nama atribut dengan menggunakan kode tertentu guna mempermudah dalam presentasi dan pengolahan data.

d. *Data mining*
   Proses selanjutnya Menerapkan algoritma *K-means* dan *K-medoids* dalam penelitian ini, dan melakukan pengujian dengan RapidMiner. Pengujian dengan RapidMiner merupakan langkah penting dalam memahami dan menguji algoritma pengelompokan *K-means* dan *K-medoids*. Langkah yang dilakuakan yaitu Konfigurasi algoritma dengan mengatur parameter-parameter seperti jumlah *cluster* (K), metode inisialisasi pusat *cluster*, dan pengaturan lainnya yang sesuai dengan penelitian.

e. Mengevaluasi DBI
   *Davies-Bouldin Index (DBI)* adalah metode penilaian yang digunakan untuk mengukur kualitas pengelompokan dalam model pengelompokan untuk data tertentu. Metode ini awalnya dikembangkan oleh David L. Davies dan Donald W. Bouldin pada tahun 1979 dan tetap menjadi salah satu metode evaluasi *clustering* yang paling umum digunakan.
   *Davies-Bouldin Index (DBI)* merupakan metrik untuk menentukan jumlah *cluster* yang optimal setelah proses pengelompokan selesai. Konsep DBI ini bertujuan untuk memaksimalkan jarak antara satu *cluster* dengan *cluster* lainnya sekaligus meminimalkan jarak antar objek dalam suatu *cluster*. Semakin kecil nilai DBI yang diperoleh (non-negatif $\geq 0$), maka semakin baik pula hasil *clustering* yang diperoleh dengan menggunakan algoritma *clustering K-means* dan *K-medoids*. DBI memungkinkan Anda mengevaluasi seberapa optimal skema pengelompokan untuk klaster Anda dengan memilih DBI yang lebih rendah.

## 4. HASIL DAN PEMBAHASAN

### 4.1. Hasil Penelitian

Pada bagian ini akan membahas tahapan-tahapan hasil yang diperoleh untuk pengklasteran komoditas Tanaman Biofarmaka di Provinsi Jawa Barat, dengan menggunakan metode *Cluster K-means* dan *Cluster K-medoids*. Selain itu, bagaimana melakukan pembandingan antara hasil pengelompokan yang dihasilkan oleh algoritma *K-means* dan *K-medoids* untuk menentukan mana yang memberikan pengelompokan yang lebih optimal dan sesuai dengan karakteristik produksi tanaman biofarmaka di Provinsi Jawa Barat.

#### 4.1.1. *Dataset*

Mengumpulkan data dengan mengunjungi web resmi BPS Jawa Barat dan mengidentifikasikan sumber-sumber data sekunder yang relevan dengan topik atau penelitian.

| No | Kabupaten/Kota | Jahe | Laos | Kencur | Kunyit | Lempuyang | Temulawak | ... | Tahun |
|---|---|---|---|---|---|---|---|---|---|
| 1 | Bogor | 1846492 | 9595520 | 991711 | 613219 | 10969 | 2230 | ... | 2022 |
| 2 | Sukabumi | 2460730 | 151250 | 314467 | 980771 | 10000 | 7440 | ... | 2022 |
| 3 | Cianjur | 8432336 | 1176925 | 858371 | 872659 | 1321 | 4758 | ... | 2022 |
| 4 | Bandung | 606097 | 82058 | 45304 | 219384 | 2805 | 3765 | ... | 2022 |
| 5 | Garut | 30009597 | 165117 | 147243 | 3479861 | 0 | 1388 | ... | 2022 |
| 6 | Tasikmalaya | 693207 | 66135 | 44984 | 68407 | 11650 | 3328 | ... | 2022 |
| 7 | Ciamis | 1709593 | 144225 | 272330 | 775577 | 0 | 0 | ... | 2022 |
| 8 | Kuningan | 365977 | 191157 | 7615 | 101903 | 0 | 0 | ... | 2022 |
| 9 | Cirebon | 7614 | 18300 | 1400 | 11063 | 600 | 584 | ... | 2022 |
| 10 | Majalengka | 1555953 | 146111 | 6514 | 2086081 | 0 | 0 | ... | 2022 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 26 | Kota Tasikmalaya | 103720 | 20720 | 2330 | 10772 | 0 | 1400 | ... | 2022 |
| 27 | Kota Banjar | 55800 | 25475 | 30249 | 61900 | 0 | 0 | ... | 2022 |

Gambar 2. Dataset

#### 4.1.2. *Data selection*

Setelah data diperoleh dari aktivitas pengumpulan data, langkah selanjutnya adalah melakukan pemilihan atribut yang sesuai dengan apa yang dibutuhkan seperti jenis tanaman, lokasi produksi, jumlah produksi, dan faktor-faktor lain yang relevan. Dataset yang akan diolah di RapidMiner berasal dari aplikasi Microsoft Excel, selanjutnya, pilih file yang akan digunakan dalam penelitian. Setelah file dimasukkan, tentukan atribut masing-masing yang terdapat di dalamnya.

#### 4.1.3. *Data preprocessing*

Dalam tahap ini dilakukan *Preprocessing* data yaitu proses cleaning data yang hilang atau kosong sebelum ke proses selanjutnya, proses cleaning data yang hilang atau kosong terdapat tiga atribut data yang dihilangkang yaitu tahun, kajibeling dan dlingo, dikarenakan data tersebut berisi nilai null.

#### 4.1.4. Data transformasi

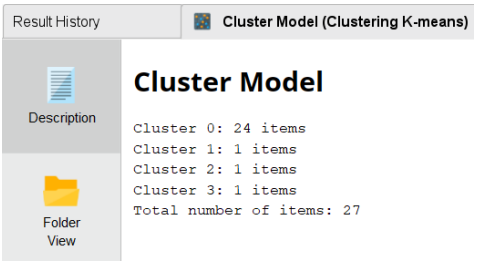
Pada tahap ini dilakukan proses transformasi data ke dalam bentuk format yang sesuai untuk diproses ke *data mining*. Bertujuan untuk memudahkan koordinasi data yang diproses oleh algoritma dan tools yang digunakan dalam penelitian yaitu rapidminer. Pada tahap transformasi atribut nomer digunakan untuk menggantikan fungsi dari atribut periode. Tetapi, ini tidak di perlukan dalam penelitian ini dikarenakan data yang akan digunakan untuk pengolahan data sudah berbentuk format yang sesuai untuk di proses ke *data mining*.

#### 4.1.5. Proses *Clustering*

Pengujian dengan RapidMiner merupakan langkah penting dalam memahami dan menguji algoritma *clustering K-means* atau *K-medoids*. Langkah-langkah yang dilakukan adalah melakukan konfigurasi algoritma dengan mengatur parameter seperti jumlah *cluster* (K), metode inisialisasi berpusat pada *cluster*, dan pengaturan lain yang sesuai dengan penelitian. Proses *clustering* menggunakan algoritma *K-means* dan *K-medoids* dilakukan pada data produksi 27 data produksi tanaman biofarmaka. *Clustering* menghasilkan pengelompokan data produksi yang dapat digunakan untuk menentukan wilayah mana yang memiliki tingkat produksi rendah, sedang, dan tinggi. Algoritma yang digunakan dalam proses *clustering* pada produksi tanaman biofarmaka di Jawa Barat adalah algoritma *K-means* dan algoritma *K-medoids* dengan menggunakan aplikasi RapidMiner.


Gambar 3. Menerapkan algoritma K-means

Setelah proses dijalankan, dapat dilihat hasil pada *cluster* model (*Clustering*), Untuk pilihan Description dapat dilihat pada Gambar 4.6 tampilan hasil *cluster* model *K-means*.



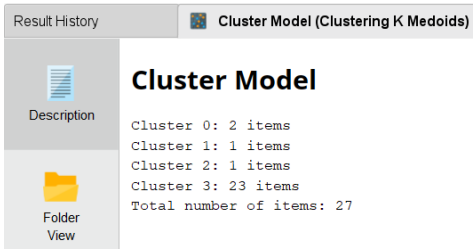

Gambar 4. Hasil Cluster Model K-means

Gambar 4 menunjukkan bahwa *cluster* yang digunakan dalam penelitian ini berjumlah empat kelas. Pengelompokan data berdasarkan algoritma *K-means*. Kelas pertama berjumlah 24 data, kelas kedua berjumlah 1 data, kelas ketiga berjumlah 1 data, dan kelas keempat berjumlah 1 data, sehingga total kumpulan data berjumlah 27 data.


Gambar 5. Menerapkan algoritma K-medoids

Setelah menjalankan prosesnya, dapat melihat hasilnya dalam model *cluster* (*clustering*). Untuk opsi

Jelaskan, Anda dapat melihat hasil model *cluster K-medoids* pada Gambar 6.




Gambar 6. Hasil Cluster Model K-medoids

Gambar 6 menunjukkan bahwa *cluster* yang digunakan dalam penelitian ini terdiri dari empat kelas. Pengelompokan data didasarkan pada algoritma *K-means*. Kelas pertama terdapat 2 buah data, kelas kedua terdapat 1 buah data, kelas ketiga terdapat 1 buah data, dan kelas keempat terdapat 23 buah data, sehingga diperoleh jumlah kumpulan data sebanyak 27 data.

Pada gambar 7 dan 8 terdapat hasil analisis *Clustering* produksi tanaman biofarmaka di Jawa Barat dengan menggunakan algoritma *K-means* dan *K-medoids*. Terlihat bahwa pada algoritma *K-means* terdapat 24 kabupaten/kota pada *cluster* 0.1 kabupaten pada *cluster* 1.1 kabupaten pada *cluster* 2 dan 1 kabupaten pada *cluster* 3. dengan jumlah total 27 kabupaten/kota. Sedangkan pada hasil analisis *Clustering* produksi tanaman biofarmaka menggunakan algoritma *K-medoids* terdapat 2 kabupaten/kota pada *cluster* 0.1 kabupaten pada *cluster* 1.1 kabupaten pada *cluster* 2 dan 23 kabupaten pada *cluster* 3. dengan jumlah total 27 kabupaten/kota.

| | | | | |
|---|---|---|---|---|
| *Cluster 0* | 1. | Bogor | 14. | Bekasi |
| | 2. | Sukabumi | 15. | Bandung Barat |
| | 3. | Kota Banjar | 16. | Pangandaran |
| | 4. | Bandung | 17. | Kota Bogor |
| | 5. | Majalengka | 18. | Kota Sukabumi |
| | 6. | Ciamis | 19. | Kota Bandung |
| | 7. | Kuningan | 20. | Kota Cirebon |
| | 8. | Cirebon | 21. | Kota Bekasi |
| | 9. | Sumedang | 22. | Kota Depok |
| | 10. | Indramayu | 23. | Kota Cimahi |
| | 11. | Subang | 24. | Kota Tasikmalaya |
| | 12. | Karawang | | |
| | 13. | Purwakarta | | |
| *Cluster 1* | 1. | Garut | | |
| *Cluster 2* | 1. | Cianjur | | |
| *Cluster 3* | 1. | Tasikmalaya | | |

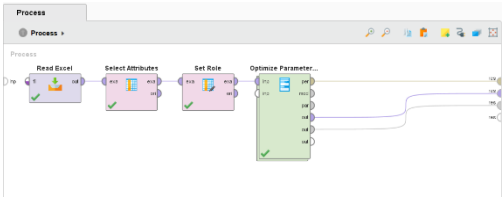
Gambar 7. Hasil Clustering K-means

| | | | | |
|---|---|---|---|---|
| *Cluster 0* | 1. | Cianjur | | |
| | 2. | Tasikmalaya | | |
| *Cluster 1* | 1. | Garut | | |
| *Cluster 2* | 1. | Sumedang | | |
| *Cluster 3* | 1. | Bogor | 13. | Bandung Barat |
| | 2. | Sukabumi | 14. | Pangandaran |
| | 3. | Bandung | 15. | Kota Bogor |
| | 4. | Ciamis | 16. | Kota Sukabumi |
| | 5. | Kuningan | 17. | Kota Bandung |
| | 6. | Cirebon | 18. | Kota Cirebon |
| | 7. | Majalengka | 19. | Kota Bekasi |
| | 8. | Indramayu | 20. | Kota Depok |
| | 9. | Subang | 21. | Kota Cimahi |
| | 10. | Purwakarta | 22. | Kota Tasikmalaya |
| | 11. | Karawang | 23. | Kota Banjar |
| | 12. | Bekasi | | |

Gambar 8. Hasil Clustering K-medoids

### 4.1.6. Mengevaluasi DBI

*Indeks Davies-Bouldin* (DBI) adalah metode evaluasi klaster internal. Keandalan internal DBI mengukur efektivitas pengelompokan dengan mengukur kuantitas dan kualitas data yang dikumpulkan. Setelah data disiapkan, lanjutkan ke langkah berikutnya, pengolahan data. Data tersebut diolah menggunakan algoritma *K-means*. Pada penelitian ini jumlah *cluster* ditentukan sebanyak 2, 3, dan 4 dengan melakukan percobaan dengan jumlah *cluster* yang berbeda. Terakhir, berdasarkan hasil evaluasi, kami menentukan jumlah *cluster* yang akan digunakan dalam penelitian ini. Analisis data yang digunakan dalam penelitian ini menggunakan RapidMiner sebagai perangkat lunak untuk melakukan proses analisis. Metode *clustering* aplikasi Rapidminer serta algoritma *K-means* dan *K-medoids* ditunjukkan pada tampilan skor DBI pada Gambar 9.




Gambar 9. Tampilan mengevaluasi DBI

Tabel 1. Tabel Hasil Rekavitulasi DBI K-means

| Percobaan | Jumlah *Cluster* | Nilai DBI |
|---|---|---|
| 1 | 2 | 0.700 |
| 2 | 3 | 0.360 |
| 3 | 4 | **0.087** |
| 4 | 5 | 0.111 |
| 5 | 6 | 0.105 |
| 5 | 7 | 0.311 |
| 7 | 8 | 0.344 |
| 8 | 9 | 0.213 |
| 9 | 10 | 0.333 |

Tabel 2. Tabel Hasil Rekavitulasi DBI K-medoids

| Percobaan | Jumlah *Cluster* | Nilai DBI |
|---|---|---|
| 1 | 2 | 0.893 |
| 2 | 3 | 0.835 |
| 3 | 4 | **0.428** |
| 4 | 5 | 1.367 |
| 5 | 6 | 0.728 |
| 6 | 7 | 3.181 |
| 7 | 8 | 2.819 |
| 8 | 9 | 3.583 |
| 9 | 10 | 3.949 |

Hasil dari proses yang menggunakan algoritma *K-means* dan *K-medoids* untuk menentukan nilai DBI setiap *cluster* dan kelompok kabupaten dan kota berdasarkan hasil rangkuman *Davies Bouldin Index* (DBI) yang ditunjukkan dengan hasil. Jika k=2 yaitu 0.893, k =3 yaitu 0.835, dan k=4 yaitu 0.428, maka nilai k yang dipilih menjadi banyaknya *cluster* adalah nilai k= 4 pada percobaan ketiga. Nilai DBI

minimumnya adalah 0.428. Hal ini dikarenakan semakin kecil nilai DBI maka *cluster* yang dihasilkan akan semakin optimal. , Oleh karena itu, empat *cluster* digunakan dalam penelitian ini. Terlihat juga bahwa *K-means* dari vektor kinerja dan nilai *Davies Bouldin Index* (DBI) dari K-medoid merupakan metode validasi *cluster* dari hasil *clustering*. Nilai DBI untuk validasi menggunakan metode *K-means* sebesar 0.087. Nilai DBI yang diperoleh dengan metode *K-medoids* sebesar 0.428. Semakin rendah nilai DBI, semakin baik *cluster*nya.

## 5. KESIMPULAN DAN SARAN

Berdasarkan produksi tanaman biofarmaka di Provinsi Jawa Barat, algoritma *K-means* digunakan untuk membuat empat *cluster*, yaitu *cluster* 1 yang terdiri dari 24 kabupaten/kota, *cluster* 2 yang terdiri dari 1 kabupaten/kota, dan *cluster* yang terdiri dari 1 kabupaten/kota. mengelompokkan kabupaten dan kota, sehingga menghasilkan 4 1 kabupaten. Sedangkan hasil pengelompokan menggunakan algoritma *K-medoids* adalah *cluster* 1 sebanyak 2 kabupaten, *cluster* 2 sebanyak 1 kabupaten, *cluster* 3 sebanyak 1 kabupaten, dan *cluster* 4 sebanyak 23 kabupaten/kota dengan menggunakan empat *cluster*. Pencarian nilai DBI menghasilkan hasil pengelompokan kabupaten dan kota berdasarkan produksi tanaman biofarmaka menggunakan algoritma *K-means* dan *K-medoids* menggunakan empat *cluster* dari dataset tahun 2022. Anda bisa Nilai-nilai DBI adalah: Metodenya adalah 0.087. Nilai DBI yang diperoleh dengan metode *K-medoids* sebesar 0.428. Perbandingan algoritma *K-means* dan *K-medoids* untuk mencari nilai DBI pada kelompok kabupaten dan kota berdasarkan produksi tanaman biofarmaka menunjukkan bahwa *K-means* efektif dalam mengolah data terkait biofarmaka. Data tanaman lebih efektif di Jawa Barat tahun 2022. Dikarenakan nilai DBI pada algoritma *K-means* lebih rendah dibandingkan dengan algoritma *K-medoids*. Nilai DBI yang terkecil pada *K-means* membuktikan bahwa algoritma *K-means* lebih unggul daripada *K-medoids*. Ini disebabkan oleh fakta bahwa dalam pengujian DBI, semakin rendah nilai yang diperoleh, semakin baik kualitas *cluster* yang dihasilkan. Dari hasil pengujian yang telah dilakukan dan hasil kesimpulan, maka saran yang dapat disampaikan penulis yaitu: Saran peneliti pada penelitian ini adalah menggunakan aplikasi dan algoritma teknik data mining yang berbeda untuk mengelompokkan kabupaten dan kota berdasarkan produksi tanaman biofarmaka pada penelitian selanjutnya. Penelitian ini kemungkinan akan dilanjutkan pada tahun-tahun mendatang untuk menambah jumlah data yang digunakan guna memperoleh hasil yang lebih baik dan akurat..

## DAFTAR PUSTAKA

[1] A. Fauzy, Perbandingan Algoritma *K-Means* dan Algoritma *K-Medoids* dalam Pengelompokan Komoditas Tanaman Biofarmaka di Provinsi Jawa Tengah Tahun 2018. dspace.uii.ac.id, 2020.

[2] N. Puspitasari, G. Lempas, H. Hamdani, H. Haviuddin, and A. Septiarini, "Perbandingan Algoritma *K-Means* dan Algoritma *K-Medoids* Pada Kasus *Covid-19* di Indonesia," *Build. Informatics, Technol. Sci.*, vol. 4, no. 4, pp. 2015–2027, 2023, doi: 10.47065/bits.v4i4.2994.

[3] F. Harahap, "Perbandingan algoritma k means dan k medoids untuk clustering kelas siswa tunagrahita," *TIN Terap. Inform. Nusant.*, vol. 2, no. 4, pp. 191–197, 2021.

[4] D. Hastari, F. Nurunnisa, S. Winanda, and D. D. Aprillia, "Penerapan algoritma k-means dan k-medoids untuk mengelompokkan data negara berdasarkan faktor sosial-ekonomi dan kesehatan," *SENTIMAS Semin. Nas. Penelit. dan Pengabdi. Masy.*, pp. 274–281, 2023.

[5] R. Khalida and H. K. Bharata, "Analisa komparasi tiga metode data mining dalam prediksi impor komoditas tanaman biofarmaka: array," *J. Ilm. Komputasi*, vol. 19, no. 2, pp. 145–154, 2020.

[6] A. Meiriza and E. Ali, "Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Program BPJS Ketenagakerjaan," *Indones. J. Comput. Sci.*, 2023, [Online]. Available: http://3.8.6.95/ijcs/index.php/ijcs/article/view/3184

[7] H. D. Tampubolon, Suhada, M. Safii, Solikhun, and D. Suhendro, "Penerapan algoritma k-means dan k-medoids clustering untuk mengelompokkan tindak kriminalitas berdasarkan provinsi," *J. Ilmu Komput. dan Teknol.*, vol. 2, no. 2, pp. 6–12, 2021.

[8] A. Supriyadi, A. Triayudi, and I. D. Sholihati, "Perbandingan algoritma k-means dengan k-medoids pada pengelompokan armada kendaraan truk berdasarkan produktivitas," *JIPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.*, vol. 06, no. 02, pp. 229–240, 2021.

[9] J. Manurung, P. S. Ramadhan, and M. G. Suryanata, "Perbandingan algoritma k-means dan k-medoids untuk pengelompokkan data masyarakat miskin pada kantor camat hatonduhan," *J. CyberTech*, vol. 3, no. 9, pp. 1522–1531, 2020.

[10] R. A. Farissa, R. Mayasari, and Y. Umaidah, "Perbandingan algoritma k-means dan k-medoids untuk pengelompokkan data obat dengan silhouette coefficient di puskesmas karangsambung," *J. Appl. Informatics Comput.*, vol. 5, no. 2, pp. 109–116, 2021, doi: 10.30871/jaic.v5i1.3237.

[11] A. Nugraha, O. Nurdiawan, and G. Dwilestari, "Penerapan Data Mining Metode K-Means Clustering Untuk Analisa Penjualan Pada Toko Yana Sport," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 6, no. 2, pp. 1–7, 2022.