Jurnal Informatika Teknologi dan Sains (JINTEKS)
E-ISSN : 2686-3359
Submit Oktober 20 2024 | Revised Oktober 30, 2024 |
Accepted November 05, 2024
Hal. 794-801
Vol. 6; No. 4
November 2024

# PERBANDINGAN METODE OPTIMASI PENENTUAN SENTROID AWAL PADA ALGORITMA K-MEANS MENGGUNAKAN ELBOW PSO DAN SSE

**Muhamad Rodi[1*], Hendrik[2], Amir Bagja[3], M Nurul Wathani[4], Zaenul Amri[5]**
[1]Program Studi Sistem Informasi, STMIK Lombok
[2]Program Studi Megister PJJ Informatik, Universitas Amikom Yogyakarta
[3]Program Studi Sistem Informasi, Universitas Hamzanwadi
[4,5]Program Studi Informatika, Universitas Hamzanwadi
*email*: Muhamadrodi97@gmail.com[1*]

**Abstrak:** Peningkatan volume dan kompleksitas data menimbulkan tantangan dalam pemrosesan big data, khususnya dalam menemukan pola dan hubungan data secara manual. Dalam data mining, metode clustering seperti algoritma K-Means banyak digunakan untuk mengelompokkan data berdasarkan kemiripan karakteristiknya. Namun, ketergantungan algoritma K-Means pada pemilihan titik awal sentroid secara acak dapat menghasilkan hasil klasterisasi yang kurang optimal. Penelitian ini bertujuan untuk membandingkan hasil evaluasi dan waktu iterasi dari tiga metode optimasi, yaitu Elbow, Particle Swarm Optimization (PSO), dan Sum of Square Error (SSE) pada algoritma K-Means. Data yang digunakan adalah dataset Online Retail II dari UCI Machine Learning Repository. Metode Davies-Bouldin Index (DBI) digunakan sebagai alat evaluasi untuk menilai kevalidan klaster yang terbentuk. Berdasarkan hasil analisis, metode optimasi Elbow dan SSE menghasilkan nilai DBI sebesar 0,8500 dengan waktu iterasi yang lebih cepat dibandingkan PSO. Sementara itu, metode PSO memberikan nilai DBI terbaik, yaitu 0,7376, meskipun membutuhkan waktu iterasi yang jauh lebih lama. Hasil penelitian ini diharapkan dapat menjadi acuan untuk memilih metode optimasi yang tepat pada algoritma K-Means, berdasarkan kebutuhan waktu dan hasil evaluasi klaster.

**Kata Kunci :** K-Means, Elbow, PSO, SSE, DBI

***Abstract:*** *The increasing volume and complexity of data present challenges in big data processing, particularly in manually identifying data patterns and relationships. In data mining, clustering methods such as the K-Means algorithm are widely used to group data based on similar characteristics. However, K-Means' reliance on random initial centroid selection can yield suboptimal clustering results. This study aims to compare the evaluation results and iteration time of three optimization methods—Elbow, Particle Swarm Optimization (PSO), and Sum of Square Error (SSE)—on the K-Means algorithm. The dataset used is the Online Retail II dataset from the UCI Machine Learning Repository. The Davies-Bouldin Index (DBI) method is used as an evaluation tool to assess the validity of the formed clusters. Based on the analysis results, the Elbow and SSE optimization methods achieved a DBI score of 0.8500 with faster iteration times compared to PSO. Meanwhile, the PSO method provided the best DBI score of 0.7376, although it required significantly longer iteration time. The results of this study are expected to serve as a reference for selecting an appropriate optimization method for the K-Means algorithm based on time requirements and clustering evaluation outcomes.*

***Keywords*** **:** *K-Means, Elbow, PSO, DBI.*

## PENDAHULUAN

Saat ini, data semakin beragam dan rumit dengan jumlahnya yang terus bertambah. Ini membuat pengolahan big data akan menghadapi masalah dalam membaca data dan menemukan pola dan relasi data jika dilakukan secara manual. Dalam data mining, pengelompokan data dapat dilakukan dengan metode clustering, salah satunya menggunakan algoritma K-Means. Klasterisasi dengan algoritma K-Means telah sering digunakan untuk mengelompokkan siswa-siswa kelas unggulan [1], mendiagnosa penyakit daun anggur[2], perencanaan kebutuhan obat di klinik [3], populasi ayam petelur [4], pengelompokan film[5], pengelompokan kelas unggulan untuk peningkatan prestasi belajar siswa [6].

Algoritma K-Means mengelompokkan data berdasarkan kedekatan, di mana data-data yang memiliki karakteristik yang sama akan dimasukkan ke dalam cluster yang sama. Kedekatan data tersebut bisa dihitung dengan melihat seberapa dekat antara data satu dengan yang lain dan dengan sentroid data. Cara kerja algoritma K-Means adalah dengan mulai menentukan jumlah k cluster terlebih dahulu, kemudian menentukan titik sentroid data secara acak. Kemudian, data akan dialokasikan ke cluster terdekat dan proses tersebut akan diulang sampai menemukan sentroid yang stabil. Penentuan jumlah cluster dan pilihan sentroid awal yang ditentukan secara acak sangat mempengaruhi hasil dari algoritma K-Means. Masalah yang timbul pada algoritma K-Means adalah bahwa sentroid akhir tidak selalu menjadi pusat cluster yang sebenarnya. Dalam prakteknya, algoritma ini perlu dijalankan beberapa kali dengan sentroid awal yang berbeda-beda agar dapat memperoleh sentroid akhir yang dianggap paling optimal [7].

Jurnal Informatika Teknologi dan Sains (JINTEKS)
E-ISSN : 2686-3359
Submit Oktober 20 2024 | Revised Oktober 30, 2024 |
Accepted November 05, 2024
Hal. 794-801
Vol. 6; No. 4
November 2024

Indeks Davies-Bouldin (DBI) adalah salah satu teknik yang dipakai untuk menilai kevalidan klaster dalam sebuah metode pengelompokan. Penggunaan Davies-Bouldin Index bertujuan untuk memperbesar jarak antara cluster dan sekaligus meminimalkan jarak antar titik dalam sebuah cluster. Apabila jarak antar kluster mencapai maksimal, ini berarti bahwa karakteristik antara setiap kluster sedikit, sehingga perbedaan antarkluster akan terlihat lebih jelas. Jika jarak minimal antara cluster berarti bahwa setiap objek dalam cluster memiliki tingkat kesamaan karakteristik yang tinggi [5].

Oleh karena itu, diperlukan teknik optimisasi untuk meningkatkan hasil evaluasi yang dapat mengatasi masalah hasil evaluasi yang kurang optimal karena penentuan sentroid secara acak. Terdapat beberapa teknik optimasi yang bisa digunakan, salah satunya adalah teknik elbow [8], metode Artificial Bee Colony [9], metode Particle Swarm Optimization [10], metode Genetic Algorithm [11], metode Support Vector Machine [12], metode Sum of Square Error [13], metode Pillar [14], metode Fuzzy Metrics [15], metode Bee Colony Optimization [16]. Dari beberapa metode tersebut, peneliti akan memilih 3 metode yang sering digunakan oleh peneliti lain untuk dioptimalkan dengan menggunakan data dari dataset yang sama. Penelitian ini bertujuan untuk membandingkan hasil evaluasi dan waktu iterasi dari 3 metode optimasi tersebut, serta menarik kesimpulan mengenai metode optimasi mana yang lebih baik untuk diterapkan.

## TINJAUAN PUSTAKA

Beberapa peneliti sebelumnya telah meneliti mengenai optimasi penentuan titik pusat cluster dengan menggunakan beberapa metode[17], Menerapkan algoritma siku (elbow) untuk mengoptimalkan penentuan sentroid awal agar didapatkan jumlah cluster optimal dengan k=2. Pembagian hasil menghasilkan cluster 1 dengan sentroid sebesar 2.686 dan 15 anggota serta cluster 2 dengan sentroid sebesar 3.245 dan 40 anggota. Penelitian dengan metode optimasi yang serupa telah dilakukan oleh [18] dengan hasil optimal 4 kluster dengan proporsi masing-masing 30%, 40%, 23,3%, dan 6,67% untuk kluster 1, 2, 3, dan 4. Penelitian serupa juga dilakukan oleh [8] dengan menggunakan metode elbow, kita dapat mencapai hasil penentuan cluster optimal. Menurut metode elbow, Within Cluster Sum of Squares (WCSS) menurun secara tajam ketika nilai K naik dari 2 menjadi 3, dan persentase perubahan maksimum SSE adalah 71,29%. Karenanya, cluster yang paling optimal adalah 3. Menurut K = 3, sebagian besar siswa (62 atau 41,05 persen) termasuk dalam cluster Intelligent System, sementara jumlah siswa terbanyak kedua (59 atau 39,07 persen) masuk ke dalam cluster Multimedia. Penelitian yang sama menggunakan metode yang serupa dilakukan oleh [19] dengan hasil sebagai berikut: 1. Algoritma k-means clustering dengan metode elbow dapat menghasilkan 2 cluster optimal dalam pengelompokkan kasus Demam Berdarah Dengue (DBD) di Jawa Barat selama periode 2016-2021. 2. Pengelompokan hasil dibagi menjadi 2 kelompok yaitu kelompok 0 dengan kategori rendah yang terdiri dari 22 daerah kabupaten/kota, dan kelompok 1 dengan kategori tinggi yang terdiri dari 5 daerah kabupaten/kota. 3. Evaluasi klaster dengan menggunakan koefisien siluet menghasilkan skor sebesar 0,689 yang menunjukkan bahwa klaster termasuk dalam kategori standar dan data telah masuk ke klaster yang tepat. Penelitian dengan dataset yang berbeda juga telah dilakukan oleh [20] Dengan K-Means, berhasil membentuk kelompok dengan anggota cluster yang sesuai dan dijalankan menggunakan bahasa pemrograman Python pada data pengguna narkoba. Dengan menjalankan Python, informasi yang dihasilkan akan sesuai dengan keadaan yang ada karena berasal dari cluster optimal yang dihasilkan dari perhitungan sum square of error (SSE), di mana nilai selisih tertinggi terjadi pada k=3.

Penelitian terkait dengan metode optimasi Particle Swarm Optimization (PSO) dilakukan oleh [21] Berdasarkan nilai rata-rata koefisien siluet, K-Means berbasis PSO lebih unggul daripada K-Means biasa, meskipun perbedaannya tidak signifikan, dengan rata-rata koefisien siluet sebesar 0,714114. Rata-rata waktu komputasi K-Means murni lebih cepat dengan rata-rata 0,748 detik dibandingkan K-Means berbasis PSO. Penelitian yang sama dengan data yang berbeda telah dilakukan oleh [22] dan hasilnya menunjukkan bahwa silhouette coefficient dan akurasi yang dihasilkan oleh PSO k-means lebih tinggi daripada k-means murni, yaitu sekitar 0,30731 dan 87%. Penelitian dengan cara yang sama juga dilakukan oleh [10] dengan hasil bahwa clustering menghasilkan 5 cluster. Berdasarkan evaluasinya 5 cluster yang sudah dioptimasi lebih baik di bandingkan dengan yang belum dioptimasi walaupun kinerja algoritmanya lebih lambat.

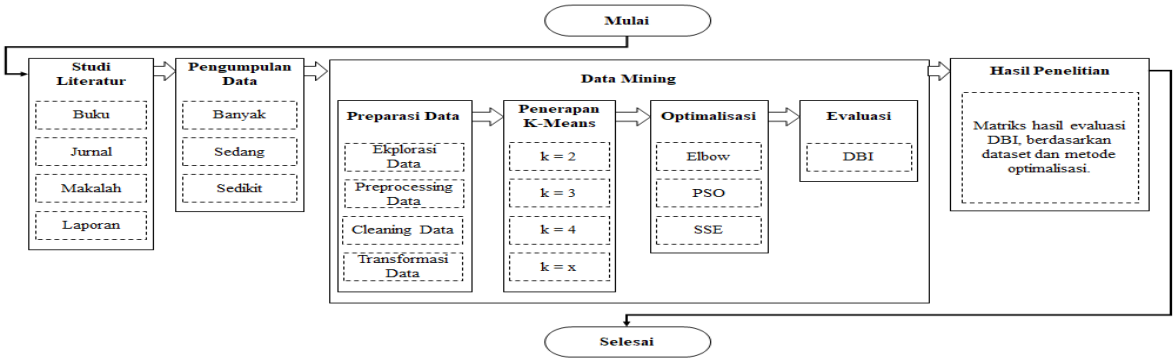
Penelitian terkait dengan metode Sum of Square Error (SSE) dilakukan oleh [13] Dengan hasil nilai optimal adalah ketika nilai K mengalami penurunan nilai SSE yang sangat signifikan sehingga membentuk grafik elbow, menunjukkan bahwa jumlah cluster terbaik adalah ketika K=3 dengan penurunan nilai SSE yang paling besar. Dan studi yang serupa dengan data lainnya dilakukan oleh [7] Berdasarkan hasil evaluasi untuk menentukan jumlah cluster yang optimal, didapatkan bahwa jumlah cluster terbaik menurut perhitungan metode elbow adalah tiga cluster, dengan nilai k=3 memiliki selisih paling besar di antara semua nilai uji k, yaitu sebesar 3116800259. Menurut perhitungan metode Davies Bouldin Index dan Silhouette Index, ditemukan bahwa jumlah cluster terbaik adalah dua cluster. Dapat diamati bahwa nilai Davies Bouldin Index terendah terjadi saat k = 2, dengan nilai

Jurnal Informatika Teknologi dan Sains (JINTEKS)
E-ISSN : 2686-3359
Submit Oktober 20 2024 | Revised Oktober 30, 2024 |
Accepted November 05, 2024
Hal. 794-801
Vol. 6; No. 4
November 2024

0.3228986726354396. Hasil dari perhitungan Silhouette Index yang mendekati 1 diperoleh pada k=2 dengan nilai 0,894.

## METODE

Dalam penelitian ini, beberapa langkah dilakukan untuk menjalankan proses pengolahan data. Langkah-langkah ini melibatkan penelitian literatur, pengumpulan data, dan hasil penelitian. Langkah-langkah yang dilakukan dalam penelitian ini dapat ditemukan di gambar 1 berikut.




**Gambar 1. Alur Penelitian**

### Studi Literatur

Dalam melakukan penelitian, perlu dilakukan studi literatur untuk mencari informasi dan pemahaman mengenai topik yang sedang dibahas serta eksperimen yang akan dilakukan. Informasi yang diperlukan untuk studi literatur bisa didapatkan dari berbagai sumber, seperti buku referensi, jurnal, makalah, laporan penelitian, dan sumber-sumber penelitian yang relevan sebelumnya. Hal ini bertujuan untuk mencapai tujuan dari suatu penelitian.

### Pengumpulan Data

Metode pengumpulan data yang digunakan pada Metode pengumpulan data kuantitatif yang digunakan dalam penelitian ini adalah melalui dataset Online Retail II yang terdiri dari 1.067.371 data penjualan online dengan 8 atribut. Data ini diambil dari UCI Machine Learning Repository seperti yang tercantum pada tabel 1 di bawah ini:

**Tabel 1. Dataset Online Retail II**

| Invoice | StockCode | Description | Quantity | InvoiceDate | Price | Customer ID | Country |
|---|---|---|---|---|---|---|---|
| 489434 | 85048 | 15CM CHRISTMAS GLASS BALL 20 LIGHTS | 12 | 01/12/2009 07:45 | 6,95 | 13085 | United Kingdom |
| 489434 | 79323P | PINK CHERRY LIGHTS | 12 | 01/12/2009 07:45 | 6,75 | 13085 | United Kingdom |
| 489434 | 79323W | WHITE CHERRY LIGHTS | 12 | 01/12/2009 07:45 | 6,75 | 13085 | United Kingdom |
| 489434 | 22041 | RECORD FRAME 7" SINGLE SIZE | 48 | 01/12/2009 07:45 | 2,1 | 13085 | United Kingdom |
| 489434 | 21232 | STRAWBERRY CERAMIC TRINKET BOX | 24 | 01/12/2009 07:45 | 1,25 | 13085 | United Kingdom |
| 489434 | 22064 | PINK DOUGHNUT TRINKET POT | 24 | 01/12/2009 07:45 | 1,65 | 13085 | United Kingdom |
| 489434 | 21871 | SAVE THE PLANET MUG | 24 | 01/12/2009 07:45 | 1,25 | 13085 | United Kingdom |
| 489434 | 21523 | FANCY FONT HOME SWEET HOME DOORMAT | 10 | 01/12/2009 07:45 | 5,95 | 13085 | United Kingdom |

Jurnal Informatika Teknologi dan Sains (JINTEKS)
E-ISSN : 2686-3359
Submit Oktober 20 2024 | Revised Oktober 30, 2024 |
Accepted November 05, 2024
Hal. 794-801
Vol. 6; No. 4
November 2024

Berdasarkan data pada tabel 1, rencananya akan dilakukan segmentasi pelanggan berdasarkan RFM (Recency, Frequency, dan Monetary) agar perusahaan dapat lebih efisien dalam menargetkan segmen pelanggan dengan menggunakan K-Means yang telah dioptimalkan.

**Data Mining**

Dalam penelitian ini, data mining dibagi menjadi tiga tahap utama, yaitu persiapan data, penggunaan algoritma K-Means dengan optimasi Elbow, penerapan PSO pada K-Means dengan optimasi SSE, dan terakhir, evaluasi dilakukan dengan DBI.

**Hasil Penelitian**

Hasil akhir dari penelitian ini adalah untuk memvisualisasikan dan menganalisis evaluasi DBI berdasarkan jumlah data yang berbeda pada dataset yang sama, serta metode optimasi yang diterapkan pada algoritma K-Means. Visualisasi terdiri dari tabel matrik dan diagram, diikuti dengan analisis untuk menentukan metode optimasi yang paling sesuai untuk jumlah data dan kondisi dataset yang berbeda.

## HASIL DAN PEMBAHASAN

Setelah mendapatkan dataset, langkah pertama yang dilakukan adalah melakukan eksplorasi dan persiapan terhadap data. Kedua langkah ini sangat krusial untuk memastikan bahwa data yang digunakan dalam analisis adalah tepat, relevan, dan siap digunakan. Dimana fokus utamanya adalah untuk memahami karakteristik dan sifat dari data yang ada. Ekplorasi data melibatkan pemahaman yang mendalam terhadap dataset yang digunakan, sementara preparasi data difokuskan pada membersihkan, mentransformasi, dan mengatur data agar dapat diproses lebih lanjut. Kedua tahap ini memiliki peran penting dalam menghasilkan hasil analisis yang akurat dan bermakna.

**Data Cleansing**

Pembersihan data, yang juga dikenal sebagai data cleansing, merupakan langkah krusial dalam manajemen data yang bertujuan untuk menemukan, memperbaiki, dan menghapus kesalahan atau ketidaksesuaian dalam kumpulan data. Ini melibatkan sejumlah langkah untuk memastikan bahwa data tetap akurat dan bersih. Data cleansing dilakukan terhadap dataset yang memiliki kolom feature tanpa nilai. Melalui proses pembersihan data yang dilakukan, hasilnya diperoleh dengan menghapus data yang tidak memiliki nilai di kolom deskripsi dan kolom ID pelanggan, sehingga akhirnya diperoleh jumlah akhir sebanyak 824.364 baris data.

**Data Preparation**

Analisis yang akan dilakukan terhadap dataset dalam studi kasus ini adalah untuk mendapatkan segmentasi atribut pelanggan berdasarkan beberapa faktor, yaitu R (Recency): Jumlah hari sejak pembelian terakhir, F (Frequency): Jumlah transaksi, dan M (Monetary): Total transaksi (kontribusi pendapatan). Untuk memperoleh atribut-atribut tersebut, langkah-langkah yang dilakukan antara lain menciptakan atribut baru bernama "Monetary", kemudian diikuti dengan pembuatan atribut baru "Frequency".

Langkah berikutnya adalah menyatukan dua atribut baru antara dataframe monetary dan frequency. Langkah berikutnya adalah membuat atribut baru bernama "Recency" dengan mengonversi tipe data atribut "InvoiceDate" menjadi datetime terlebih dahulu. Kemudian, kita akan menampilkan tanggal maksimal untuk mengetahui kapan terakhir kali transaksi dilakukan oleh pelanggan. Setelah itu, kita akan menghitung waktu sejak transaksi terakhir pelanggan untuk mendefinisikan atribut Recency atau jumlah hari sejak pembelian terakhir. Karena atribut Recency hanya memerlukan informasi jumlah hari, maka setelah itu akan dilakukan ekstraksi, atau hanya akan diambil informasi tentang jumlah hari saja.

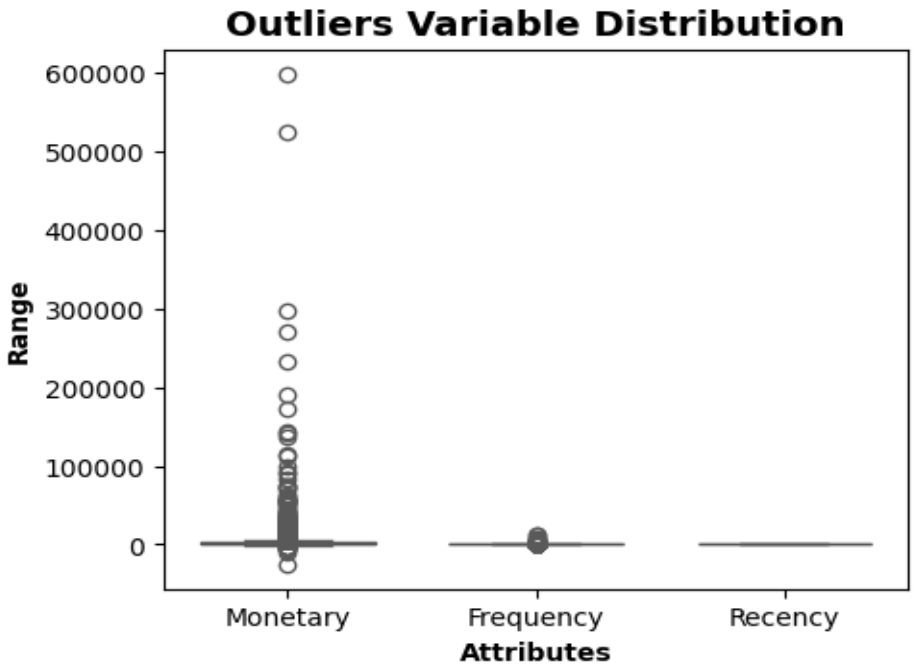
Langkah selanjutnya menggabungkan ketiga atribut baru tersebut sehingga didapatkan hasil seperti pada tabel 2 berikut ini:

**Tabel 2. Penggabungan Atribut MFR**

| CustomerID | Monetary | Frequency | Recency |
|---|---|---|---|
| 12346 | -64.68 | 48 | 325 |
| 12347 | 5633.32 | 253 | 1 |
| 12348 | 2019.4 | 51 | 74 |

| | | | |
|---|---|---|---|
| 12349 | 4404.54 | 180 | 18 |
| 12350 | 334.4 | 17 | 309 |

Setelah data seperti yang tercantum dalam tabel 1 telah diperoleh, langkah selanjutnya adalah menganalisis outlier dengan menggunakan boxplot. Diagram kotak atau boxplot adalah jenis visualisasi data statistik yang menunjukkan distribusi data serta beberapa statistik deskriptif penting. Diagram ini menunjukkan bagaimana data didistribusikan dari nilai terendah hingga nilai tertinggi dan juga menyoroti adanya nilai-nilai ekstrem. Setelah dilakukan analisis, hasilnya diperoleh seperti yang terlihat pada gambar 2 berikut ini:




**Gambar 2. Outlier Variabel**

Berdasarkan gambar 2, dapat disimpulkan bahwa terdapat data yang melenceng sehingga perlu dilakukan penghapusan data outlier tersebut. Langkah terakhir sebelum memodelkan adalah melakukan rescaling. Rescaling or penskalaan is the process of changing the range or scale of a variable in a set of data. Tujuan utama dari penskalaan adalah untuk memastikan variabel-variabel memiliki rentang nilai yang serupa atau cocok dengan metode analisis yang akan dilakukan. Dalam kasus ini, Standardisasi Scaling digunakan untuk setiap atribut dengan hasil yang ditunjukkan pada Tabel 3 di bawah ini:

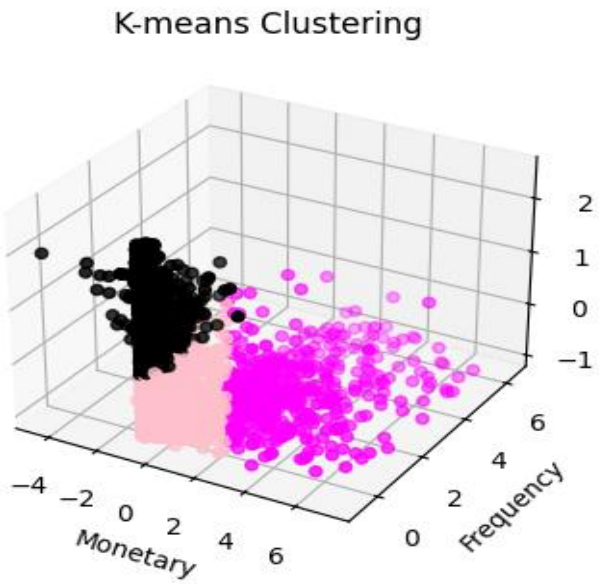
**Tabel 3. Data MFR Hasil Standarisasi**

| Monetary | Frequency | Recency |
|---|---|---|
| -0.66857 | -0.40729 | 0.56501 |
| 1.34691 | 0.88754 | -0.9622 |
| 0.0686 | -0.38834 | -0.61811 |
| 0.91227 | 0.42646 | -0.88207 |
| -0.52741 | -0.60309 | 0.48959 |
| -0.53925 | -0.57782 | 0.79597 |

| | | |
|---|---|---|
| 0.02255 | 0.00327 | -0.80194 |
| -0.50181 | -0.55888 | -0.01005 |

**Pemodelan K-Means Optimasi PSO**

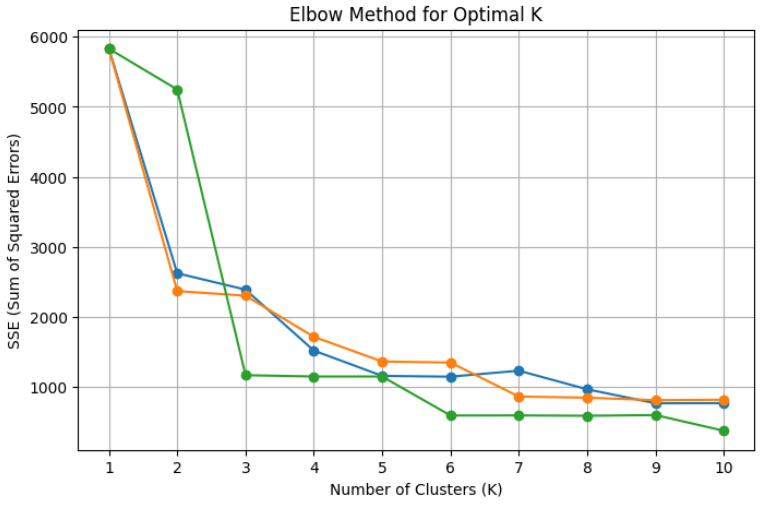
Dalam percobaan ini, pemodelan K-Means dengan optimasi PSO menggunakan beberapa parameter, antara lain nilai k yang bervariasi dari 2 hingga 10, serta nilai iterasi maksimum sebanyak 100. Hasil dari pemodelan K-Means yang telah dioptimalkan dengan PSO adalah k=3 yang optimal dan nilai DBI=0.7376 dengan waktu iterasi sepanjang 771.2143 detik. Untuk menggambarkan cluster menggunakan optimasi PSO seperti yang ditunjukkan pada gambar 3 di bawah ini:




**Gambar 3. Visual Cluster k=3**

**Pemodelan K-Means Optimasi Elbow**

Dalam uji coba ini, penggunaan pendekatan K-Means dengan optimasi Elbow dilakukan dengan parameter nilai k yang bervariasi dari 2 hingga 10 dan maksimum iterasi sebanyak 100. Hasil dari pemodelan dan parameter tersebut menunjukkan bahwa k=6 adalah nilai yang paling optimal dengan DBI=0.8500 dan iterasi membutuhkan waktu 0.0002 detik. Untuk memvisualisasikan cluster dengan menggunakan metode optimasi Elbow seperti yang terlihat pada gambar 4 di bawah ini:



**Gambar 4. Optimasi Elbow**

**Pemodelan K-Means Optimasi SSE**

Dalam percobaan ini, pemodelan K-Means dengan optimasi SSE menggunakan parameter nilai k mulai dari 2 hingga 10 dan dengan iterasi maksimum 100. Hasil dari pemodelan dan parameter tersebut menunjukkan bahwa nilai k yang paling optimal adalah 6 dengan DBI=0.8500 dan waktu iterasi 1.8401 detik untuk visualisasi cluster dengan optimasi SSE.

**Perbandingan Evaluasi DBI**

Dari 3 hasil percobaan di atas dapat kita analisa hasil dari masing-masing optimasi sebagaimana terlihat pada tabel 4 berikut ini:

**Tabel 4. Perbandingan Hasil Optimasi**

| Optimasi | k | DBI | Waktu (detik) |
|---|---|---|---|
| Elbow | 6 | 0.8500 | 0.0002 |
| PSO | 3 | 0.7376 | 771.2143 |
| SSE | 6 | 0.8500 | 1.8401 |

Dari tabel 3 di atas, kita dapat menyimpulkan bahwa optimasi K-Means menggunakan metode PSO menghasilkan nilai DBI terbaik sebesar 0.7376. Sementara itu, optimasi menggunakan metode Elbow dan SSE menghasilkan nilai DBI yang sama yaitu 0.8500. Namun, metode Elbow dan SSE lebih cepat dalam proses iterasinya dibandingkan dengan metode PSO, dengan perbedaan yang sangat signifikan.

## KESIMPULAN DAN SARAN

Berdasarkan hasil dan pembahasan, disimpulkan bahwa algoritma K-Means dengan optimasi Elbow lebih efisien dari segi waktu iterasi dibandingkan dengan optimasi menggunakan metode PSO. Waktu yang diperlukan untuk iterasi dalam optimasi Elbow adalah 0.0002 detik, sementara iterasi dalam optimasi PSO membutuhkan waktu sekitar 771.2143 detik, dengan SSE memakan waktu sekitar 1.8401 detik. Secara keseluruhan, evaluasi DBI pada algoritma K-Means baik dengan optimasi Elbow, SSE, maupun PSO menghasilkan hasil DBI yang paling optimal, yaitu 0.7376 dengan cluster optimal pada k=3.

Dalam penelitian berikutnya, disarankan untuk menggunakan dataset yang berbeda dan membandingkannya dengan menggunakan metode optimasi lainnya, sehingga dapat membandingkan hasil evaluasi DBI untuk menentukan clustering yang paling optimal dengan K-Means.

## DAFTAR PUSTAKA

[1] J. Hutagalung, "Pemetaan Siswa Kelas Unggulan Menggunakan Algoritma K-Means Clustering," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 9, no. 1, pp. 606–620, 2022, doi: 10.35957/jatisi.v9i1.1516.

[2] S. M. Javidan, A. Banakar, K. A. Vakilian, and Y. Ampatzidis, "Diagnosis of grape leaf diseases using automatic K-means clustering and machine learning," *Smart Agric. Technol.*, vol. 3, no. June 2022, p. 100081, 2023, doi: 10.1016/j.atech.2022.100081.

[3] L. ' Izzah and A. Jananto, "Penerapan Algoritma K-Means Clustering Untuk Perencanaan Kebutuhan Obat Di Klinik Citra Medika," *Progresif J. Ilm. Komput.*, vol. 18, no. 1, p. 69, 2022, doi: 10.35889/progresif.v18i1.769.

[4] E. Ramadanti and M. Muslih, "Penerapan Data Mining Algoritma K-Means Clustering Pada Populasi Ayam Petelur Di Indonesia," *Rabit J. Teknol. dan Sist. Inf. Univrab*, vol. 7, no. 1, pp. 1–7, 2022, doi: 10.36341/rabit.v7i1.2155.

[5] I. F. Ashari, R. Banjarnahor, D. R. Farida, S. P. Aisyah, A. P. Dewi, and N. Humaya, "Application of Data Mining with the K-Means Clustering Method and Davies Bouldin Index for Grouping IMDB Movies," *J. Appl. Informatics Comput.*, vol. 6, no. 1, pp. 07–15, 2022, doi: 10.30871/jaic.v6i1.3485.

[6] Y. H. Syahputra and J. Hutagalung, "Superior Class to Improve Student Achievement Using the K-Means Algorithm," *SinkrOn*, vol. 7, no. 3, pp. 891–899, 2022, doi: 10.33395/sinkron.v7i3.11458.

[7] M. Orisa, "Optimasi Cluster pada Algoritma K-Means," *Pros. SENIATI*, vol. 6, no. 2, pp. 430–437, 2022,

doi: 10.36040/seniati.v6i2.5034.

[8] M. Hamka and N. Ramdhoni, "K-Means Cluster Optimization for Potentiality Student Grouping Using Elbow Method," *AIP Conf. Proc.*, vol. 2578, no. November, 2022, doi: 10.1063/5.0108926.

[9] G. Yao, Y. Wu, X. Huang, Q. Ma, and J. Du, "Clustering of Typical Wind Power Scenarios Based on K-Means Clustering Algorithm and Improved Artificial Bee Colony Algorithm," *IEEE Access*, vol. 10, no. August, pp. 98752–98760, 2022, doi: 10.1109/ACCESS.2022.3203695.

[10] T. M. Dista and F. F. Abdulloh, "Clustering Pengunjung Mall Menggunakan Metode K-Means dan Particle Swarm Optimization," *J. Media Inform. Budidarma*, vol. 6, no. 3, p. 1339, 2022, doi: 10.30865/mib.v6i3.4172.

[11] Y. Shima, R. A. Kadir, and F. H. Ali, "A Novel Approach to the Optimization of a Public Bus Schedule Using K-Means and a Genetic Algorithm," *IEEE Access*, vol. 9, pp. 73365–73376, 2021, doi: 10.1109/ACCESS.2021.3080508.

[12] T. Y. Wen and S. A. Mohd Aris, "Hybrid Approach of EEG Stress Level Classification Using K-Means Clustering and Support Vector Machine," *IEEE Access*, vol. 10, pp. 18370–18379, 2022, doi: 10.1109/ACCESS.2022.3148380.

[13] L. P. Refialy, H. Maitimu, and M. S. Pesulima, "Perbaikan Kinerja Clustering K-Means pada Data Ekonomi Nelayan dengan Perhitungan Sum of Square Error (SSE) dan Optimasi nilai K cluster," *Techno.Com*, vol. 20, no. 2, pp. 321–329, 2021, doi: 10.33633/tc.v20i2.4572.

[14] D. Lestari, A. C. Fauzan, and Harliana, "Penerapan Algoritma Pillar Untuk Optimasi Penentuan Titik Awal Centroid Pada Algoritma K-Means Clustering," *J. Inf. Systerm Infromatics Eng.*, vol. 6, no. 1, pp. 15–24, 2022.

[15] X. Geng, Y. Mu, S. Mao, J. Ye, and L. Zhu, "An Improved K-Means Algorithm Based on Fuzzy Metrics," *IEEE Access*, vol. 8, pp. 217416–217424, 2020, doi: 10.1109/ACCESS.2020.3040745.

[16] I. Arfiani, H. Yuliansyah, and M. D. Suratin, "Implementasi Bee Colony Optimization Pada Pemilihan Centroid (Klaster Pusat) Dalam Algoritma K-Means," *Build. Informatics, Technol. Sci.*, vol. 3, no. 4, pp. 756–763, 2022, doi: 10.47065/bits.v3i4.1446.

[17] M. Al Ghifari, W. Trisari, and H. Putri, "Clustering Courses Based On Student Grades Using K-Means Algorithm With Elbow Method For Centroid Determination," vol. 8, no. 1, pp. 42–46, 2023.

[18] T. Santoso, A. Darmawan, N. Sari, M. A. F. Syadza, E. C. B. Himawan, and W. A. Rahman, "Clusterization of Agroforestry Farmers using K-Means Cluster Algorithm and Elbow Method," *J. Sylva Lestari*, vol. 11, no. January, pp. 107–122, 2023.

[19] D. Amelia, T. N. Padilah, and A. Jamaludin, "Optimasi Algoritma K-Means Menggunakan Metode Elbow dalam Pengelompokan Penyakit Demam Berdarah Dengue (DBD) di Jawa Barat," *J. Ilm. Wahana Pendidik.*, vol. 8, no. 11, pp. 207–215, 2022.

[20] A. Winarta and W. J. Kurniawan, "Optimasi cluster k-means menggunakan metode elbow pada data pengguna narkoba dengan pemrograman python," *J. Tek. Inform. Kaputama*, vol. 5, no. 1, pp. 113–119, 2021.

[21] T. P. Yoga, "Optimalisasi K-Means Berbasis Particle Swarm Optimization untuk Hasil Produksi Tanaman Sayuran di Indonesia," vol. 17, 2023.

[22] H. Harliana, R. M. Herdian Bhakti, O. Saeful Bachri, and F. Sofian Efendi, "Optimasi K-Means dengan Particle Swarm Optimization pada Pengelompokkan Daerah Stunting," *J. Ilm. Intech Inf. Technol. J. UMUS*, vol. 3, no. 02, pp. 95–101, 2021, doi: 10.46772/intech.v3i02.457.