Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 6, No. 1, November 2024, Page 14-24
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v6i1.5873

Perbandingan Algoritma K-Means dan K-Medoids untuk Clustering Pada Transaksi Penjualan Minimarket

Ajeng Shalwa Alganiu[*], Ayu Ratna Juwita, Rahmat, Sutan Faisal

Fakultas Ilmu Komputer, Program Studi Teknik Informatika, Universitas Buana Perjuangan Karawang, Karawang, Indonesia
Email: [1,*]if20.ajengalganiu@mhs.ubpkarawang.ac.id, [2]ayurj@ubpkarawang.ac.id, [3]rahmat@ubpkarawang.ac.id, [4]sutan.faisal@ubpkarawang.ac.id
Email Penulis Korespondensi: if20.ajengalganiu@mhs.ubpkarawang.ac.id
Submitted: 30/08/2024; Accepted: 15/11/2024; Published: 15/11/2024

Abstrak−Dalam berbelanja pembeli sering mengalami kesulitan dalam mencari kebutuhan sehari-hari. Salah satu penyebabnya adalah proses penataan produk di minimarket masih mengatur produk secara sembarangan dan tidak sesuai dengan pola belanja konsumen. Sebaliknya, pembeli umumnya cenderung membeli produk dalam bentuk paket kebutuhan sehari-hari, namun paket ini umunya belum ada di minimarket. Mengidentifikasi pola hubungan dalam data transaksi minimarket dapat membantu mengatasi masalah penataan produk dan penyusunan paket produk. Dengan menggunakan metode pengelompokan, objek dikelompokkan dalam kategori yang memiliki banyak kesamaan antara satu sama lain. Metode ini memungkinkan proses pengelompokan dilakukan. Beberapa metode yang ada dalam clustering diantaranya metode K Means dan K-medoids. Penelitian ini bertujuan untuk mengklasifikasikan data barang di minimarket tersebut yang bisa dijadikan acuan untuk perencanaan produk yang lebih terorganisir dengan baik. Pengelompokan data terbagi kedalam tiga kategori yaitu lambat, sedang dan cepat. Hasil yang diperoleh menunjukkan bahwa kedua algoritma tersebut menghasilkan indeks Davies-Bouldin Index yang berbeda, dengan algoritma K-Medoids memperoleh nilai lebih rendah yaitu sebesar 0,50387 sedangkan K-Means mendapatkan nilai sebesar 0,50391 yang dimana hasil clustering K-Medoids memiliki kualitas yang lebih unggul dibandingkan dengan K-Means. Dengan hasil pengelompokan data barang tersebut, minimarket dapat melakukan penyeimbangan stok barang untuk mencegah kelebihan atau kekurangan persediaan barang tersebut.

Kata Kunci: K-Means; K-Medoids; Minimarket; Davies-Bouldin Index

Abstract−When shopping, buyers often have difficulty finding daily necessities. One of the causes of this is because the product arrangement process in minimarkets is still carried out randomly and does not match consumer shopping patterns. On the contrary, buyers usually want to buy products through daily necessities packages, but these packages are usually not yet available in minimarkets. Identifying relationship patterns in minimarket transaction data can help overcome product arrangement and product packaging problems. By using the clustering method, objects are grouped into groups that have many similarities with each other. This method allows the grouping process to be carried out. Some of the methods in clustering include the K-Means and K-medoids methods. The purpose of this study is to group the data on goods in the minimarket which can be a guide for more neatly arranged product planning. Data grouping is divided into 3 categories namely slow, medium and fast. The results obtained show that the two algorithms produce different Davies-Bouldin Index values, with the K Medoids algorithm obtaining a lower value of 0.50387 while K-Means obtains a value of 0.50391 where the K-Medoids clustering results have better quality compared to K-Means. With the results of the grouping of these goods data, minimarkets can balance the stock of goods to prevent excess or shortage of inventory of these goods.

Keywords: K-Means; K-Medoids; Minimarket; Davies-Bouldin Index

1. PENDAHULUAN

Minimarket seringkali disebut juga sebagai swalayan atau toko kecil, tempat yang menyediakan berbagai macam produk dan layanan yang dibutuhkan dalam kehidupan sehari-hari oleh masyarakat. Konsep ini mirip dengan toko kelontong, tetapi minimarket dianggap sebagai bentuk yang lebih modern dari toko kelontong [1]. Pada umumnya minimarket yang ada di Indonesia belum menerapkan metode standard yang konsisten. Salah satu masalah umum dalam pengelolaan stok barang secara manual adalah ketidakpastian mengenai jumlah dan kondisi barang yang masih tersisa di gudang. Hal ini bisa menyebabkan barang terlalu lama tersimpan di gudang, terjadi kelebihan, kekurangan, atau bahkan kehabisan stok. Kerusakan, kesalahan dalam pencatatan pemasukan, kelalaian mencatat permintaan, pengeluaran barang yang tidak sesuai dengan pesanan, serta berbagaai kemungkinan lainnya yang bisa menyebabkan perbedaan antara catatan persediaan dan stok yang sebenarnya ada di gudang [2]Untuk mengatasi masalah tersebut, suatu teknik data mining diterapkan dengan menganalisis data penjualan untuk memperoleh informasi yang berguna pada proses perencanaan dan pengendalian persediaan barang. Proses pengolahan data dilakukan melalui metode *clustering* data menggunakan k-means & k-medoid.

Data mining merupakan suatu progres yang memanfaatkan statistik deskriptif dan matematika, kecerdasan buatan, serta perangkat teknologi untuk mengumpulkan pengetahuan, menentukan informasi yang bermanfaat, dan memperoleh informasi relevan dari basis data besar [3]. Data mining yang merupakan salah satu aspek dari teknologi informasi dapat digunakan untuk memprediksi, mengklasifikasi, maupun mengelompokkan data yang memliki dimensi tinggi. Penerapan berbagai teknik ini bertujuan untuk menghasilkan model yang efisien dalam mengelola atau menemukan pola dalam himpunan data yang besar. Beberapa teknik data mining yang popule meliputi klasifikasi, klastering, asosiasi, peramalan, dan estimasi [4]. Istilah "mining" dalam konteks ini mengacu pada upaya untuk mengekstraksi nilai yang berharga dalam jumlah besar. Oleh karena itu, Data Mining sebenarnya berasal dari bidang ilmu seperti kecerdasan buatan (artificial intelligence),

This Journal is licensed under a Creative Commons Attribution 4.0 International License

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 6, No. 1, November 2024, Page 14-24
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v6i1.5873

pembelajaran mesin, statistika, dan basis data. Dengan kata lain, data mining adalah proses ekstraksi pola-pola yang terdapat dalam data. Data mining menjadi semakin signifikan sebagai alat untuk mengubah data tersebut menjadi informasi yang bermanfaat[5].

Clustering adalah proses membagi sekumpulan objek data ke dalam kelompok-kelompok yang disebut *cluster*. Objek-objek dalam sebuah klaster memiliki karakteristik yang mirip satu sama lain, namun berbeda dengan objek-objek di cluster lain. Pembagian tidak dilakukan secara manual, melainkan menggunakan algoritma *clustering*. Oleh sebab itu, *clustering* sangat bermanfaat dan dapat mengidentifikasi grup atau kelompok yang tidak diketahui dalam data [6]. Algoritma K-Means *clustering* adalah metode yang lebih sederhana dalam pengelompokkan data atau mudah diimplementasikan. Proses algoritma ini melibatkan partisi atau pembagian sejumlah data ke dalam K *cluster* berdasarkan jarak terdekat ke titik rata-rata (*mean*) dari setiap klaster, yang kemudian dihitung ulang pada setiap iterasi berikutnya. Keunggulan dari metode K-Means *clustering* adalah waktu komputasinya yang cukup cepat [7]. Selain itu, algoritma K-Medoids juga tergolong dalam kelompok metode partisi *clustering* yang merupakan variasi dari pendekatan K-Means. Perbedaan antara dua algoritma ini ada pada cara merepresentasikan pusat klaster. K-Medoid menggunakan objek sebagai medoid yang berfungsi sebagai pusat klaster untuk setiap kelompok, sementara K-Means menggunakan nilai rata-rata sebagai pusat klaster [8]. Algoritma K-Medoids, yang juga dikenal sebagai algoritma PAM (*Partitioning Around Medoid)*, di kembangkan oleh Leonard Kaufman dan Peter J. Rousseeuw. Algoritma ini serupa dengan K-Means sebab keduanya merupakan algoritma partisional yang membagi dataset menjadi beberapa klaster. Ketidaksamaan antara algoritma K-Means dan algoritma K-Medoids berada pada penetapan pusat klaster. Algoritma K-Means menggunakan nilai rata-rata dari setiap klaster menjadi pusatnya, sedangkan algoritma K-Medoids menggunakan objek data yang berfungsi sebagai perantara (*medoid*) untuk pusat klaster [9]

Sebelumnya, telah dilakukan penelitian oleh Farahdinna (2024) tentang pengelompokan produk asuransi perusahaan nasional menggunakan metode *clustering* dengan judul "Perbandingan Algoritma K-Means dan K-Medoids Dalam Klasterisasi Produk Asuransi Perusahaan Nasional". Fokus penelitian ini adalah pada produk-produk terbaik dan menentukan produk yang cocok dengan keperluan nasabah. Hasil dari penelitian tersebut menghasilkan dua nilai DBI terkecil yang didapatkan dengan menggunakan metode K-Means yaitu 0,018 dengan jumlah klaster k=5, dan metode K-Medoids yaitu sebesar 0,027 dengan jumlah klaster k=2. Dengan demikian, metode K-Means merupakan pilihan terbaik untuk membentuk cluster yang optimal pada pengelompokan produk asuransi dari perusahaan nasional [10]. Selanjutnya, penelitian lain yang dilakukan oleh Fathia (2022) [11], berjudul "Perbandingan Algoritma K-Means Dan K-Medoids Dalam Pengelompokkan Tingkat Kebahagiaan Provinsi Di Indonesia" juga menerapkan algoritma K-Means & K-Medoids untuk mengelompok data. Temuan dari penelitian tersebut menghasilkan komparasi pendekatan K-Means & K-Medoids untuk mengidentifikasi jumlah klaster yang optimal membuktikan bahwa metode K-Medoids paling tepat pada penelitian ini. Dalam proses klaterisasi, diketahui jika nilai validitas DBI yang paling tinggi terjadi pada klasterisasi K-Means yaitu sebesar 0,752 dan K-Medoids sebesar 0,648 dengan menghasilkan 2 cluster yaitu cluster (1) terdiri dari 17 provinsi yang dianggap sebagai provinsi yang paling bahagia, sementara cluster (0) mencakup 17 provinsi yang dianggap kurang bahagia. Pada penelitian lain yang dilakukan Salsabilla (2024), algoritma K-Means & K-Medoids serta metode Elbow untuk membentuk klaster hasil produksi buah-buahan berdasarkan jenis produksi di kabupaten Kotawaringin Timur. Terdapat 3 klaster yang mengacu pada hasil produksi buah. Hasil uji kinerja klaster menunjukkan bahwa pengelompokkan dengan 3 klaster menghasilkan DBI dengan nilai yang lebih rendah dengan nilai 0,296 menggunakan algoritma K-Means sedangkan algoritma K-Medoids dengan nilai DBI yang mencapai 0,507. yang dimana algoritma K-Means terbukti menjadi pilihan terbaik untuk klasterisasi hasil produksi buah-buahan di kabupaten kota Waringin Timur berdasarkan pada penilaian nilai DBI yang telah dihasilkan [12]. Pada penelitian selanjutnya dilakukan oleh Ena Tasia (2023), tentang suatu wilayah yang terkena rawan banjir di Kabupaten Rokan Hilir pada tahun 2019 yang menggunakan metode clustering. Berdasarkan penggunaan tools RapidMiner pada klaster k=2 hingga k=6, metode K-Means terbukti lebih optimal dibandingkan dengan K-medoid pada data kejadian banjir di Rokan Hilir, dengan k optimal sebanyak 3 dan nilai Davies-Bouldin Index (DBI) sebesar 0,218. Sementara itu, K-Medoid menunjukan klaster optimal pada k=4 dengan validitas sebesar 0,525 [13]

Meninjau dari beberapa penelitian sebelumnya, metode *Davies Bouldin Index* (DBI) bermanfaat dalam mengidentifikasi jumlah klaster paling optimal untuk proses data *mining clustering*. Sebelum melakukan proses *clustering* pada data transaksi penjualan, disarankan untuk melakukan proses awal yaitu *Exploratory Data Analysis* (EDA). EDA merupakan tahap awal dalam analisis data yang mencakup pencarian pola, pengecekan keberadaan nilai yang hilang, pencilan, dan duplikat dalam data. Proses EDA menggunakan ringkasan statistik dan visualisasi data untuk mengumpulkan informasi sebelum dilakukan analisis lebih lanjut. Langkah-langkah dalam proses EDA termasuk mengkategorikan data menjadi variabel numerik dan kategorikal, mengevaluasi korelasi antar atribut, dan memvisualisasikan atribut-atribut data [14]. Dalam menentukan jumlah klaster terbaik digunakan metode Elbow. Metode Elbow adalah pendekatan yang digunakan dalam algoritma K-Means untuk mengidentifikasi jumlah klaster yang paling optimal. Pada grafik Elbow, terdapat titik di mana penurunan nilai menjadi lebih landai, membentuk lengkungan tajam. Titik tersebut menandakan jumlah klaster terbaik atau nilai K yang optimal [15]. Untuk mengevaluasi kualitas klaster yang dihasilkan dari K-Means & K-Medoids, digunakanlah teknik evaluasi *Davies Bouldien Index* (DBI). Metode ini mengevaluasi hasil dari klaster yang

This Journal is licensed under a Creative Commons Attribution 4.0 International License

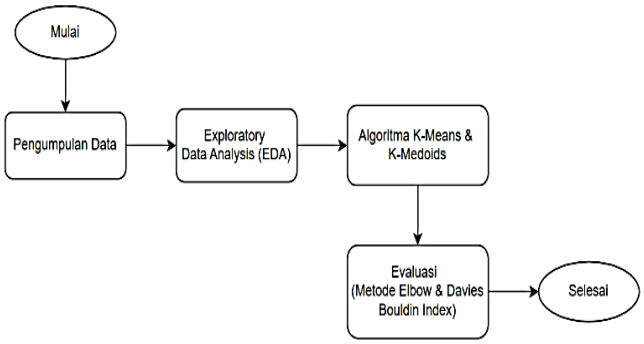
Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 6, No. 1, November 2024, Page 14-24
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v6i1.5873

telah dibentuk. DBI adalah metode untuk menilai validitas internal dengan cara mengukur seberapa baik klasterisasi yang dihasilkan, dengan menghitung jumlah banyaknya kriteria atau fitur dari dataset [16].

Berdasarkan latar belakang yang telah dijelaskan, penelitian ini akan mengelompokkan data penjualan berdasarkan jenis tingkat penjualannya. Pengelompokan data terbagi menjadi 3 kategori yakni lambat, sedang dan cepat. Tujuan penelitian ini adalah untuk mengelempokkan data barang di minimarket tersebut yang dapat digunakan sebagai referensi untuk perencanaan produk yang lebih tertata secara rapi. Mengacu pada akurasi dan hasil yang baik dari metode yang diterapkan dalam penelitian sebelumnya, metode yang akan diterapkan adalah analisis klaster dengan algoritma K-Means & K-Medoids. Proses analisis dalam penelitian ini akan dilakukan memanfaatkan bahasa pemrograman *Python*.

2. METODOLOGI PENELITIAN

Metode penelitian ini membahas mengenai tahapan-tahapan penelitian yang diterapkan sebagai dasar untuk tahapan implementasi dan pengujian sistem. Selanjutnya, penelitian ini menjelaskan tentang tahapan-tahapan dalam bentuk diagram alur, yang dapat dilihat pada Gambar 1.



Gambar 1. Flowchart Penelitian

Gambar tersebut menunjukkan diagram alir dari proses analisis data untuk klustering menggunakan algoritma *K-Means* dan *K-Medoids*. Berikut adalah penjelasan langkah-langkah dalam diagram tersebut :

1. Mulai : Proses dimulai dari titik ini.
2. Pengumpulan Data : Langkah pertama adalah mengumpulkan data yang akan digunakan dalam analisis. Data ini bisa berupa data mentah yang akan diproses dan dianalisis lebih lanjut.
3. Exploratory Data Analysis (EDA): Setelah data terkumpul, dilakukan analisis eksploratif untuk memahami karakteristik data, mengidentifikasi pola, menangani nilai yang hilang, serta membersihkan data. Tahap ini penting untuk memastikan bahwa data siap digunakan untuk proses klustering.
4. Algoritma K-Means & K-Medoids: Setelah EDA selesai, data dianalisis menggunakan algoritma klustering *K-Means* dan *K-Medoids*. *K-Means* membagi data menjadi beberapa kluster berdasarkan jarak ke centroid, sementara *K-Medoids* memilih titik pusat yang benar-benar ada di dalam data sebagai representasi setiap kluster.
5. Evaluasi (Metode Elbow & Davies-Bouldin Index): Setelah proses klustering, evaluasi dilakukan menggunakan *Metode Elbow* dan *Davies-Bouldin Index* untuk menilai kualitas klustering. Metode Elbow membantu menentukan jumlah kluster optimal dengan melihat titik "siku" pada grafik yang menunjukkan penurunan jarak dalam kluster. Davies-Bouldin Index digunakan untuk mengukur seberapa baik kluster dibentuk dan seberapa jauh antar kluster.
6. Selesai: Proses analisis data selesai.
   Diagram ini menunjukkan alur sistematis untuk melakukan klustering, dari pengumpulan data hingga evaluasi hasil klustering untuk memastikan bahwa hasil akhir memiliki kualitas yang baik.

2.1 Pengumpulan Data

Pada tahap pengumpulan data, akan digunakan data yang relevan dengan penelitian. Dataset pada penelitian ini yaitu data transaksi penjualan di sebuah minimarket pada tahun 2021-2022 dan data ini diakses pada tanggal 22 januari 2024 pada pukul 11.00 WIB melalui platform web Kaggle di https://www.kaggle.com/datasets/ipunguhbpwt/clsuterpenjualan. Dataset ini terdiri dari 7403 baris dan memiliki 5 variabel yaitu kode_barang, nama_barang, jumlah_transaksi, total_penjualan dan rata_rata.

2.2 *Exploratory Data Analysis (EDA)*

Data yang telah diperoleh kemudian dianalisis melalui proses *Exploratory Data Analysis* (EDA). Proses ini mencakup pencarian pola, memeriksa apakah terdapat *missing value*, *outlier* dan duplikat dalam data. Langkah selanjutnya adalah melakukan visualisasi data untuk memperoleh informasi sebelum data tersebut diproses lebih lanjut.

Copyright © 2024 **Author**, Page 16
This Journal is licensed under a Creative Commons Attribution 4.0 International License

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 6, No. 1, November 2024, Page 14-24
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v6i1.5873

2.3 *Clustering* (Algoritma K-Means)

Pada penelitian ini, melibatkan penerapan model machine learning dengan memanfaatkan algoritma K-Means *Clustering*, yang termasuk salah satu teknik *unsupervised learning*. Pada tahapan ini, hasil klasterisasi di evaluasi dengan membagi pelanggan ke dalam segmen-segmen tertentu. Proses segmentasi pelanggan dimulai dari persiapan data, dilanjutkan dengan seleksi data yang akan di klaster, dan menetapkan jumlah klaster dengan menggunakan metode *Elbow.*
Berikut adalah tahapan-tahapan dalam penerapan algoritma K-Means:

a. Menetukan nilai k untuk total klaster yang akan dihasilkan
b. Menentukan pusat klaster awal yang akan dibentuk.
c. Mengukur jarak antara objek pada *centroid* menggunakan *Euclidean Distance.*

$$D_e = \sqrt{(x_i - s_i)^2 + (y_i - t_i)^2} \quad (1)$$

Dimana *De* merupakan *Euclidean Distance*, *i* merupakan banyaknya data, *(x, y)* merupakan titik data dan *(s, t)* merupakan titik pusat. Menghitung jarak *Euclidean* antara dua titik dalam ruang dua dimensi, yaitu titik pertama dengan koordinat (x_i, y_i) dan titik kedua dengan koordinat (s_i, t_i). Dalam rumus ini D_e merepresentasikan jarak *Euclidean* antara kedua titik tersebut. Komponen $(x_i - s_i)^2$ dan $(y_i - t_i)^2$ masing-masing mewakili perbedaan kuadrat antara koordinat x dan y dari kedua titik. Setelah menghitung perbedaan kuadrat antara koordinat x dan y, hasilnya dijumlahkan, kemudian diakarkan untuk mendapatkan jarak lurus antara kedua titik. Jarak *Euclidean* ini sering digunakan dalam analisis data dan pemrosesan citra untuk mengukur kedekatan atau perbedaan antara dua titik dalam ruang dua dimensi, yang sering kali berguna dalam klasifikasi, klustering, dan pengenalan pola.

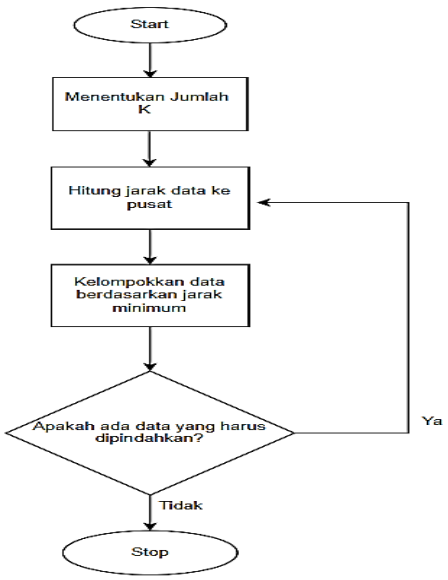
d. Kelompokkan data ke dalam klaster berdasarkan jarak terdekat.
e. Melakukan perulangan, lalu menentukan posisi titik pusat (*centroid*) baru dengan menggunakan rumus berikut :

$$\bar{v} = \frac{1}{N_i} \sum_{k=0}^{N_i} x_{kj} \quad (2)$$

Dimana $\bar{v}i$ j merupakan Rata-rata titik pusat pada klaster ke-i untuk variabel j, N_i merupakan Banyaknya anggota dalam klaster ke-i, *i*, *k* merupakan Indikator dari klaster, *j* merupakan Indikator dari variabel dan x_{kj} merupakan Nilai data ke-k pada klaster tersebut untuk variabel ke-j

Rumus diatas digunakan untuk menghitung rata-rata nilai dari sekumpulan data tertentu. Dalam rumus ini, $\bar{v}$ melambangkan nilai rata-rata dari semua data yang diamati, sedangkan N_i menunjukkan jumlah total elemen atau data dalam himpunan tersebut. Simbol $\sum_{k=0}^{N_i} x_{kj}$ menyatakan penjumlahan dari setiap nilai x_{kj} dalam himpunan, di mana k adalah indeks yang menunjuk pada setiap elemen mulai dari 0 hingga N_i. Setelah semua nilai dalam himpunan dijumlahkan, hasilnya dibagi dengan jumlah elemen N_i untuk memperoleh rata-rata $\bar{v}$. Rumus ini sering digunakan dalam analisis statistik atau data untuk menemukan nilai rata-rata dari sejumlah besar data, sehingga memudahkan dalam memahami karakteristik umum dari data tersebut.

f. Ulangi langkah 3 jika posisi *centroid* yang baru berbeda.

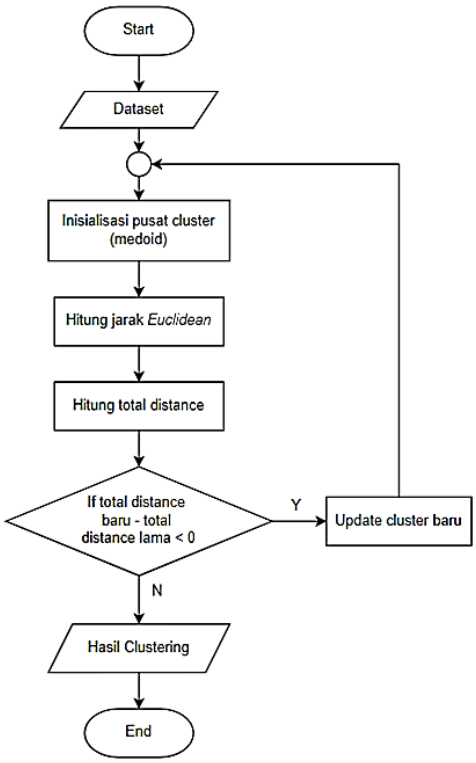


Gambar 1. Flowchart K-Means

Copyright © 2024 **Author**, Page 17
This Journal is licensed under a Creative Commons Attribution 4.0 International License

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 6, No. 1, November 2024, Page 14-24
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v6i1.5873

Gambar 2 menunjukkan alur kerja algoritma *K-Means Clustering*. Proses dimulai dengan menentukan jumlah kluster (*K*) yang diinginkan. Selanjutnya, jarak tiap data ke pusat kluster dihitung, dan data dikelompokkan berdasarkan jarak minimum ke pusat kluster tersebut. Algoritma kemudian mengecek apakah ada data yang perlu dipindahkan ke kluster lain. Jika ada, pusat kluster diperbarui, dan proses diulang. Jika tidak ada data yang perlu dipindahkan lagi, klustering selesai.



Gambar 2. Flowchart K-Medoid

Lalu, pada Gambar 3 menunjukkan alur proses algoritma *K-Medoids* untuk klustering data. Proses dimulai dengan dataset yang sudah disiapkan. Pertama, pusat kluster (*medoid*) diinisialisasi secara acak. Kemudian, jarak *Euclidean* dari setiap titik data ke medoid dihitung, diikuti dengan perhitungan total jarak (total *distance*). Jika total jarak baru lebih kecil dari jarak sebelumnya, maka kluster diperbarui dengan medoid baru dan perhitungan diulang. Proses ini terus berulang sampai tidak ada pengurangan dalam total jarak, menghasilkan klustering akhir yang optimal.

2.4 *Clustering* (Algoritma K-Medoids)

Algoritma K-Medoids adalah metode klasik dalam teknik partisi *clustering* yang membagi dataset yang terdiri dari n objek menjadi k *cluster* yang sudah ditentukan sebelumnya (Abhishek & Purnima, 2013). Algoritma ini bekerja berdasarkan prinsip untuk meminimalkan kesamaan antara setiap objek dengan titik acuan yang tepat.
Berikut adalah langkah-langkah untuk menyelesaikan algoritma K-Medoids :

a. Menginisialisasi pusat klaster dengan jumlah k (jumlah klaster).
b. Menentukan atau menghitung jarak (cost) memakai rumus *Euclidean Distance* sebagai berikut :

$$d_{\text{Euclidean}}(X, Y) = \sqrt{\sum_{i=1}^{n} (X_i - Y_i)^2} \quad (3)$$

c. Tentukan objek secara acak di setiap klaster sebagai pilihan medoid baru.
d. Menghitung jarak antara masing-masing objek di setiap klaster dengan kandidat medoid yang baru.
e. Tentukan jumlah simpangan (S) dengan cara menghitung selisih jumlah *distance* yang baru dan jumlah *distance* yang lama. Apabila S < 0, ubah objek dengan data klaster untuk membentuk k objek baru sebagai medoid.
f. Kembali ke langkah 3 sampai 5 hingga medoid tidak lagi berubah, sehingga dihasilkan klaster beserta anggotanya masing-masing.

2.5 Evaluasi

2.5.1 Metode Elbow

Metode Elbow diterapkan untuk menginterpretasikan dan menguji konsistensi performa jumlah klaster yang optimal dengan mempertimbangkan nilai SSE. Dengan memanfaatkan hasil perbandingan jumlah klaster yang

This Journal is licensed under a Creative Commons Attribution 4.0 International License

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 6, No. 1, November 2024, Page 14-24
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v6i1.5873

terbentuk di titik tertentu, metode elbow berpotensi untuk membantu menentukan jumlah klaster yang terbaik [17]. Pada titik tertentu, grafik akan menunjukkan penurunan yang signifikan dengan adanya suatu lekukan yang dikenal sebagai "siku". Penentuan nilai k yang optimal dilakukan dengan membandingkan nilai SSE (Sum of Square Error) yang ditampilkan pada grafik [18]. Nilai tersebut kemudian menjadi jumlah klaster optimal atau k yang terbaik. Dengan meningkatnya jumlah klaster K, nilai SSE cenderung menurun. Rumus SSE untuk algoritma K-Means dapat ditemukan dalam formula 2.

$$SSE = \sum_{K=1}^{k} \sum_{X_i} |X_i - C_k|^2$$

Top of Form

Keterangan *K* merupakan jumlah klaster yang digunakan, X_i merupakan jumlah data ke-i, C_k merupakan banyaknya klaster i pada klaster ke-*k*,| | merupakan menghitung jarak *Euclidean*

2.5.2 *Davies-Bouldin Index (DBI)*

Evaluasi yang diterapkan untuk menilai seberapa baik klaster yang dihasilkan oleh algoritma *K-Means & K-Medoids* adalah *Davies-Bouldin Index*. Semakin kecil nilai dari DBI, maka semakin baik hasil *cluster* tersebut [19]. Evaluasi dengan menggunakan *Davies-Bouldin Index* melibatkan skema evaluasi klaster internal, dimana kualitas klaster dinilai berdasarkan jumlah serta kedekatan antara *cluster* hasil. *Davies-Bouldin Index* adalah salah satu teknik yang diterapkan untuk mengukur keakuratan hasil klaster dari sebuah teknik *clustering*. *Cohesion* didefinisikan sebagai jumlah data yang berhubungan dengan titik pusat klaster yang diikuti oleh klaster tersebut. Sementara itu, pemisahan diukur berdasarkan jarak antara titik pusat klaster satu dengan yang lainnya. Penilaian dengan *Davies-Bouldin Index* ini berusaha mengoptimalkan jarak antar klaster, yaitu antara klaster Ci dan Cj, secara bersamaan berusaha meminimalkan jarak antara titik dalam satu klaster. Apabila jarak antara klaster berada pada tingkat maksimal, ini menunjukan bahwa persamaan karakteristik di antara klaster adalah rendah, sehingga perbedaan di antara klaster menjadi lebih terlihat. Sebaliknya, apabila jarak dalam satu klaster minimal, hal ini menunjukan bahwa setiap objek dalam klaster tersebut memiliki tingkat kesamaan karakteristik yang tinggi [20]

3. HASIL DAN PEMBAHASAN

3.1 Hasil Pengumpulan Data

Tahap pengumpulan data menjadi tahap awal yang dilakukan pada bagian hasil serta pembahasan kali ini, adapun data yang digunakan didapatkan dari website Kaggle yang merupakan data hasil rekap penjualan minimarket. Dataset yang digunakan terdiri dari 7403 baris dengan total 5 kolom yang antara lain, kode_barang, nama_barang, jumlah_transaksi, total_penjualan, rata_rata. Data yang digunakan dibuat pada tahun 2021 lalu diakses oleh penelititan kali ini pada tanggal 22 Januari 2024 guna melakukan penelitian. Adapun pada Tabel 1 terdapat penjabaran variable meliputi nama variable serta deskripsi variable.

Tabel 1. Penjabaran Dataset

Dataset	
Nama Variabel dan Tipe Data	Deskripsi
kode_barang (Object)	Berisikan nilai unik dai setiap barang.
Nama_barang (Object)	Berisikan kumpulan nama-nama barang.
Jumlah_transaksi (Integer)	Berisikan jumlah transaksi setiap barang.
Total_penjualan (Integer)	Berisikan total penjualan dari setiap barang.
Rata-Rata (Float)	Berisikan nilai rata-rata dari setiap barang.

Pada Tabel 1, diketahui terdapat 2 kolom yang mendeskripsikan Nama dari variabel beserta dengan tipe datanya, serta terdapat deskripsi dari setiap variabel yang di jabarkan. Berdasarkan uraian pada Tabel 1, diketahui bahwa terdapat 5 kolom yang dapat diolah pada dataset antara lain, nama_barang, kode_barang, jumlah_transaksi, total_penjualan, rata-rata. Adapun variabel kode_barang berisikan nilai unik berupa numerik dari setiap bareng/produk, selanjutnya variabel nama_barang berisikan kumpulan nama-nama barang yang tersedia pada minimarket. Selanjutnya variabel jumlah_transaksi berisikan jumlah transaksi dari setiap produk yang dibeli oleh konsumen. Berikutnya variabel total_penjualan berisikan total penjualan dari setiap produk yang dibeli oleh konsumen. Terakhir variabel rata-rata berisikan jumlah rata-rata setiap produk yang dibeli oleh konsumen.

3.2 Hasil Exploratory Data Analysis (EDA)

Tahap ini diawali dengan pengecekan informasi dari 5 data pertama, hal ini dilakukan untuk sedikit melihat distribusi data yang akan diolah. Berikutnya dilakukan pengecekan data dengan menggunakan *function* "info" untuk mengetahui berapa jumlah data mulai dari variabel jumlah keseluruhan baris data, serta tipe data. Tahapan selanjutnya merupakan tahap pembersihan data, di mana pada langkah ini akan di pastikan bahwa data sudah

Copyright © 2024 **Author**, Page 19
This Journal is licensed under a Creative Commons Attribution 4.0 International License

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 6, No. 1, November 2024, Page 14-24
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v6i1.5873

tidak memiliki duplikat, *missing value* dan *outlier* agar pemodelan menjadi lebih akurat. Adapun Tabel 2 menjabarkan jumlah duplikat data yang ada pada dataset.

Tabel 2. Proses Penghapusan Duplikat Data

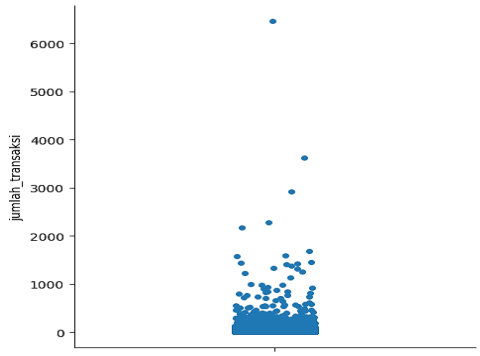
Sebelum Penghapusan Duplikat	Setelah Penghapusan Duplikat
7.403	7.391
Jumlah Duplikat Data = 12	

Dapat dilihat pada Tabel 2 terdapat kolom data sebelum dilakukan penghapusan duplikata data, serta kolom jumlah data yang sudah dilakukan penghapusan data. Pada kolom sebelum penghapusan duplikat terdapat 7.403 data setelah dilakukan penghapusan data jumlah baris menjadi 7.391 data seperti yang dijabarkan pada Tabel 2, di mana hal ini dikarenakan data yang diolah memiliki 12 baris data. Selanjutnya setelah data dibersihkan dari duplikat data, dilakukan pengecekan *missing value* dari data yang digunakan, di mana hasil pengecekan *missing value* dapat dilihat pada Tabel 3.

Tabel 3. Proses Penghapusan *Missing Value*

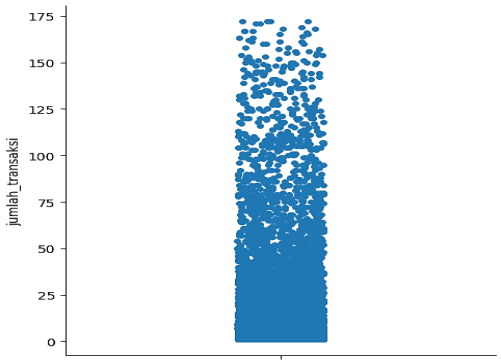
Sebelum Penghapusan *Missing Value*	Setelah Penghapusan *Missing Value*
7.391	7.372
Jumlah *Missing Value* = 19	

Hasil penjabaran proses penghapusan *missing value* pada Tabel 3 terdapat 2 kolom yang mendefinisikan jumlah data sebelum dan sesudah penghapusan *missing value*. Adapun data sebelum missing value dihapus sebesar 7.391 sedangkan setelah dilakukan penghapusan *missing value* data berkuran menjadi 7.372 baris, hal ini dikarenakan terdapat total 19 *missing value* di dalam dataset yang digunakan.
Selanjutnya dilakukan pengecekan outlier dengan menggunakan visualisasi catplot, di mana hasil dari visualisasi boxplot dapat dilihat pada Gambar 4.



Gambar 4. Catplot Pengecekan Outlier

Dapat terlihat pada Gambar 4 visualisasi dari catplot pada fitur jumlah_transaksi, di mana berdasarkan visualisasi ini diketahui bahwa data yang diolah memiliki outlier, hal ini ditunjukan pada ukuran yang ada pada sumbu "y" mulai dari 0 sampai dengan 6000+ jumlah transaksi. Berdasarkan visualisasi yang terdapat paga Gambar 4 juga ditemukan sebuah nilai yang sangat extreme di atas 6000 sebanyak 1 data. Setelah ditemukannya outlier pada Gambar 4 dilakukan pembersihan outlier dengan menggunakan metode z-score. Hal ini dilakukan untuk memfilter posisi dari outlier, di mana berdasarkan visualisasi pada Gambar 4 outlier terletak pada bagian atas plot atau upper limit. Setelah dilakukan pembersihan outlier visualisasi berubah seperti yang ada pada Gambar 5.



Gambar 3. Catplot Setelah Pembersihan Outlier

Copyright © 2024 **Author**, Page 20
This Journal is licensed under a Creative Commons Attribution 4.0 International License

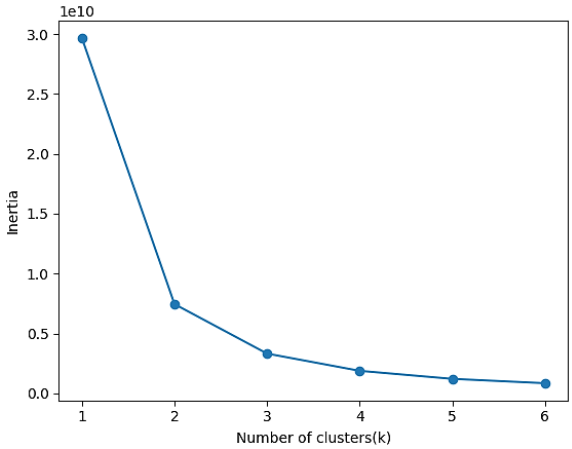
Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 6, No. 1, November 2024, Page 14-24
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v6i1.5873

Visualisasi pada Gambar 5 menunjukan penyebaran data pada fitur jumlah_transaksi setelah dilakukan penghapusan outlier dengan metode z-score. Hasil yang ditunjukan pada Gambar 5 menunjukan data yang lebih seragam tanpa adanya nilai yang terlalu extreme seperti pada Gambar 4, di mana sebelumnya nilai yang ada pada sumbu "y" lebih dari 6000, setelah outlier dihalangkan berkurang menjadi 175. Hal ini menandakan bahwa outlier/nilai extreme berhasil untuk di hapus dalam fitur jumlah_transaksi.

Selanjutnya setelah semua proses pembersihan data dilakukan, dilanjutkan kembali pada tahap encoding data, hal ini dilakukan karena data yang dapat di proses dengan python harus berbentuk numeric. Berdasarkan hal ini data pada fitur nama_barang akan di encode menjadi numeric berbentuk kategori dengan menggunakan library dari LabelEncoder. Hasil yang didapatkan pada proses encode data juga menunjukan bahwa data yang diolah memiliki 7079 kategori nama_barang.

3.3 Implementasi Algoritma K-Means

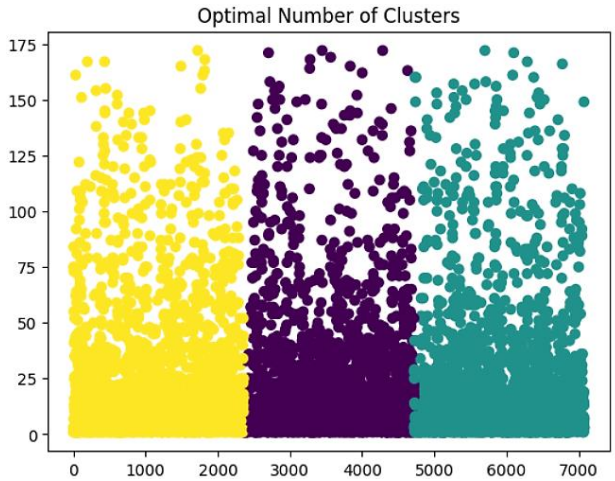
Tahap ini diawali dengan pembuatan kurva elbow untuk memberikan informasi rekomendasi nilai K terbaik berdasarkan inertia. Adapun hasil dari pembuatan kurva elbow dengan algoritma K-Means dapat terlihat pada Gambar 6.



Gambar 4. Elbow Method K-Means

Visualisasi diagram garis yang terdapat di Gambar 6 merupakan hasil penerapan dari metode elbow, di mana hal ini berguna untuk memperkirakan nilai cluster yang lebih baik. Dalam visualisasi pada Gambar 6, diketahui bahwa sumbu "x" menyimpan nilai dari setiap kluster, mulai dari 1 hingga cluster 6. Selanjutnya pada sumbu "y" menyimpan nilai inertia dari data yang digunakan. Berdasarkan hasil yang didapat pada visualisasi Gambar 6, diketahui bahwa nilai cluster yang direkomendasikan berdasarkan metode elbow yaitu 3.

Selanjutnya setelah rekomendasi kluster terbaik diketahui berikutnya merupakan tahap implementasi algoritma K-Means, di mana hasil dari visualisasi cluster dengan K-Means dapat terlihat pada Gambar 7.



Gambar.7 Hasil Clustering K-Means

Dapat terlihat pada Gambar 7, merupakan hasil clustering dari algoritma K-Means, diketahui juga bahwa hasil yang didapat berdasarkan visualisasi scatterplot menunjukan hasil yang kurang maksimal, di mana masih belum menunjukan keterpisahan kelas dengan baik. Adapun berikut pada Gambar 8 merupakan code implementasi dari algoritma K-Means.

Copyright © 2024 **Author**, Page 21
This Journal is licensed under a Creative Commons Attribution 4.0 International License

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 6, No. 1, November 2024, Page 14-24
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v6i1.5873

```
km = KMeans(n_clusters=3)
model_km= km.fit(train)
km_pred = model_km.predict(train)
plt.scatter(train.iloc[:, 0], train.iloc[:, 1], c=km_pred)
plt.title("Optimal Number of Clusters")
plt.show()
```

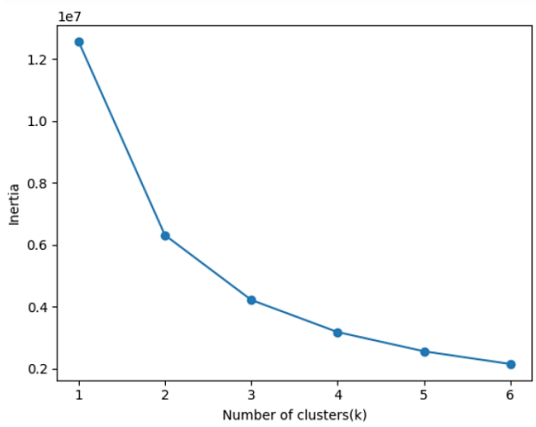
Gambar 5. Code Implementasi K-Means

Dapat diihat pada Gambar 8 merupakan code implementasi dari algoritma K-Means, di mana tahap ini diawali dengan pembuatan variabel “km” sebagai variabel yang akan menyimpan algoritma K-Means. Selanjutnya merupakan pembuatav variabel “model_km” yang akan digunakan untuk menyimpan model yang sudah dilatih. Selanjutnya merupakan pembuatan variabel “km_pred” guna menguji model yang sudah dilatih, berikutnya merupakan tahap visualisasi menggunakan scatter plot guna melihat distribusi hasil klustering.

3.4 Implementasi Algoritma K-Medoid

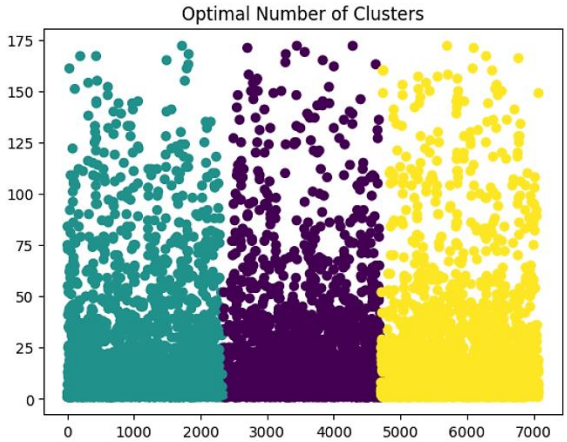
Selanjutnya merupakan tahapan dari implementasi algoritma K-Medoid, serupa dengan tahapan sebelumnya, tahap ini akan diawali dengan penerapan metode elbow guna mendapatkan rekomendasi kluster yang optimal berdasarkan nilai inersia. Visualisasi elbow dapat dilihat pada Gambar 8.

Hasil visualisasi pada Gambar 9 menunjukan diagram garis yang terdiri dari jumlah cluster pada sumbu “x” serta jumlah inertia pada sumbu “y”. Berdasarkan hasil yang terdapat pada Gambar 9, jumlah cluster yang di rekomendasikan dan akan digunakan pada tahap ini adalah 3.



Gambar 9. Visualisasi Elbow K-Medoid

Selanjutnya setelah ditemukan jumlah kluster rekomendasi berdasarkan metode elbow, dilanjutkan dengan penerapan algoritma K-Medoid, di mana hasil visualisasi cluster dengan menggunakan algoritma K-Medoid dapat dilihat pada Gambar 10.



Gambar 6. Hasil Clustering K-Medoid

Dapat dilihat hasil visualisasi pada Gambar 10, sama seperti hasil dengan menggunakan metode K-Means, K-Medoid juga memperoleh hasil pemisahan kluster yang kurang baik. Hal ini dibuktikan dengan tidak ada pemisahan kelas yang signifikan berbeda dari setiap clusternya seperti yang terlihat pada Gambar 10. Adapun hasil pada Gambar 11 merupakan code dari implementasi algoritma K-Medoids.

Copyright © 2024 **Author**, Page 22
This Journal is licensed under a Creative Commons Attribution 4.0 International License

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 6, No. 1, November 2024, Page 14-24
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v6i1.5873

```
kmed = KMedoids(n_clusters=3)
model_kmed = kmed.fit(train)
kmed_pred = model_kmed.predict(train)
plt.scatter(train.iloc[:, 0], train.iloc[:, 1], c=kmed_pred)
plt.title("Optimal Number of Clusters")
plt.show()
```

Gambar 7. Code Implementasi K-Medoids

Dapat dilihat pada Gambar 11, merupakan hasil code dari implementasi algoritma K-Medoids, di mana langkah pertama yang dilakukan pada proses ini adalah membuat variabel "kmed" guna menyimpan modul algoritma K-Medoids. Selanjutnya membuat variabel "model_kmed" guna menyimpan model yang dilatihh. Berikutnya dilakukan proses clustering yang disimpan pada variabel "kmed_pred". Selanjutnya merupakan tahap visualisasi scatter guna mengetahui hasil clustering.

3.5 Evaluasi

Di tahap ini, evaluasi dilakukan dengan menggunakan metode *davies-bouldin index* yang diterapkan untuk mengetahui performa dari model clustering yang dibuat dengan menggunakan metode K-Means & K-Medoid. Adapun hasil evaluasi menggunakan *davies-bouldin index* dapat dilihat pada Tabel 4.

Tabel 4. Hasil Davies-Bouldin Index

Algoritma	*Score*
K-Means	0.50391608
K-Medoid	0.50387082

Pada Tabel 4 terdapat uraian dari hasil pengujian performa algoritma K-Means dan K-Medoid, di mana pada Tabel 4 terdapat kolom nama algoritma serta kolom score ataupun nilai. Berdasarkan hasil yang didapat pada Tabel 4, diketahui bahwa algoritma K-Medoid menjadi algoritma yang mendapatkan performa lebih unggul dari pada K-Means. Ini dikarenakan nilai davis bouldin index yang didapatkan oleh algoritma K-Medoid lebih rendah sebesar 0.5038 dibandingkan dengan algoritma K-Means sebanyak 0.5039, yang menandakan keterpisahan yang didapat oleh algoritma K-Means sedikit lebih buruk.

4. KESIMPULAN

Hasil yang di dapat pada bagian ini menggunakan nilai k = 3, yang didapat berdasarkan rekomendasi metode elbow, hasil visualisasi yang didapat oleh setiap algoritma pada bagian hasil menunjukan bahwa model masih belum dapat membedakan kelas dengan baik. Hal ini didukung dengan hasil visualisasi pada algoritma K-Means dan K-Medoid yang masih tidak membentuk sebuah cluster dengan baik. Adapun visualisasi hasil kluster yang baik biasanya ditunjukan dengan adanya perbedaan behaviour dari setiap data. Sedangkan hasil yang didapat pada visualisasi dengan K-Means dan K-Medoid dalam penelitian ini, tidak menunjukan adanya perbedaan dari setiap cluster, yang menandakan hasil pengklusteran yang didapat masih belum baik. Adapun berdasarkan hasil yang didapat pada proses evaluasi menggunakan metode Davis Bouldin Index sebelumnya, diketahui bahwa algoritma K-Medoid menjadi algoritma yang lebih baik dibadingkan dengan algoritma K-Means. Adapun nilai Davis Bouldin Index yang didapat oleh algoritma K-Medoids sebesar 0.5038, sedangkan algoritma K-Means memperoleh nilai Davies-Bouldin Index sebesar 0.5039, di mana hal ini menunjukan hasil yang didapat oleh algoritma K-Means memiliki keterpisahan data yang lebih buruk.

REFERENCES

[1] H. Prastiwi, Jeny Pricilia, and Errissya Rasywir, "Implementasi Data Mining Untuk Menentuksn Persediaan Stok Barang Di Mini Market Menggunakan Metode K-Means Clustering," *Jurnal Informatika Dan Rekayasa Komputer(JAKAKOM)*, vol. 2, no. 1, pp. 141–148, Apr. 2022, doi: 10.33998/jakakom.2022.2.1.34.

[2] M. M. Purba and Chaerul Rahmat, "Perancangan Sistem Informasi Stok Barang Berbasis Web Di Pt Mahesa Cipta," *Jurnal Sistem Informasi Universitas Suryadarma*, vol. 8, no. 2, Jun. 2021, doi: 10.35968/jsi.v8i2.721.

[3] Aqib Fharaj Zhaky, Sutan Faisal, and Yana Cahyana, "Segmentasi Jumlah Tenaga Kesehatan Berdasarkan Kecamatan di Kabupaten Karawang Menggunakan Metode K-Medoids," *Scientific Student Journal for Information, Technology and Science*, vol. V, no. 02, Jul. 2024.

[4] Sirojul Alam, Amril Mutoi Siregar, and Ayu Ratna Juwita, "Penerapan Algoritme C4.5 Untuk Klasifikasi Kasus Covid-19," *Scientific Student Journal for Information, Technology and Science*, vol. III No.1, Jul. 2022.

This Journal is licensed under a Creative Commons Attribution 4.0 International License

ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 6, No. 1, November 2024, Page 14-24
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v6i1.5873

[5] E. Widodo, "Pelita Teknologi Prediksi Penjurusan IPA, IPS dan BAHASA dengan Menggunakan Machine Learning Abstrak Informasi Artikel," *Jurnal Pelita Teknologi*, vol. 15, no. 1, pp. 37–48, Apr. 2020.

[6] K. Annisa, B. Serasi Ginting, and M. A. Syari, "Penerapan Data Mining Pengelompokan Data Pengguna Air Bersih Berdasarkan Keluhannya Menggunakan Metode Clustering Pada PDAM Langkat," *ALGORITMA: Jurnal Ilmu Komputer dan Informatika*, vol. 06, no. 01, Apr. 2022, doi: 10.30829/algoritma.v6i1.11624.

[7] S. Ramadhani, D. Azzahra, and T. Z, "Comparison of K-Means and K-Medoids Algorithms in Text Mining based on Davies Bouldin Index Testing for Classification of Student's Thesis," *Digital Zone: Jurnal Teknologi Informasi dan Komunikasi*, vol. 13, no. 1, pp. 24–33, May 2022, doi: 10.31849/digitalzone.v13i1.9292.

[8] F. Fathurrahman, S. Harini, and R. Kusumawati, "Evaluasi Clustering K-Means Dan K-Medoid Pada Persebaran Covid-19 Di Indonesia Dengan Metode Davies-Bouldin Index (DBI)," *Jurnal Mnemonic*, vol. 6, no. 2, pp. 117–128, Oct. 2023, doi: 10.36040/mnemonic.v6i2.6642.

[9] R. Siagian, P. Sirait, and A. Halim, "SISTEMASI: Jurnal Sistem Informasi Penerapan Algoritma K-Means dan K-Medoids untuk Segmentasi Pelanggan pada Data Transaksi E-Commerce The Implementation of K-Means and K-Medoids Algorithm for Customer Segmentation on E-commerce Data Transactions," May 2022. [Online]. Available: http://sistemasi.ftik.unisi.ac.id

[10] F. Farahdinna, I. Nurdiansyah, A. Suryani, and A. Wibowo, "PERBANDINGAN ALGORITMA K-MEANS DAN K-MEDOIDS DALAM KLASTERISASI PRODUK ASURANSI PERUSAHAAN NASIONAL," *Jurnal Ilmiah FIFO*, vol. 11, no. 2, p. 208, Nov. 2019, doi: 10.22441/fifo.2019.v11i2.010.

[11] C. Fathia Palembang, M. Yahya Matdoan, S. P. Palembang, and K. Kunci, "BULLET : Jurnal Multidisiplin Ilmu Perbandingan Algoritma K-Means Dan K-Medoids Dalam Pengelompokkan Tingkat Kebahagiaan Provinsi Di Indonesia," *BULLET : Jurnal Multidisiplin Ilmu*, vol. 01, no. 5, pp. 830–839, Nov. 2022.

[12] E. Prasetyaningrum and P. Susanti, "Jurnal Media Informatika Budidarma Perbandingan Algoritma K-Means Dan K-Medoids Untuk Pemetaan Hasil Produksi Buah-Buahan," vol. 7, pp. 1775–1783, 2023, doi: 10.30865/mib.v7i4.6477.

[13] E. T. Ena Tasia and M. Afdal, "Perbandingan Algoritma K-Means Dan K-Medoids Untuk Clustering Daerah Rawan Banjir Di Kabupaten Rokan Hilir," *Indonesian Journal of Informatic Research and Software Engineering (IJIRSE)*, vol. 3, no. 1, pp. 65–73, Mar. 2023, doi: 10.57152/ijirse.v3i1.523.

[14] N. Basuni and Amril Mutoi Siregar, "Comparison of the Accuracy of Drug User Classification Models Using Machine Learning Methods," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 7, no. 6, pp. 1348–1353, Dec. 2023, doi: 10.29207/resti.v7i6.5401.

[15] A. Satriawan, R. Andreswari, and O. N. Pratiwi, "Segmentasi Pelanggan Telkomsel Menggunakan Metode Clustering Dengan Rfm Model Dan Algoritma K-Means Telkomsel Customer Segmentation Using Clustering Method With Rfm Model And K-Means Algorithm," Apr. 2021.

[16] F. Tempola, M. Muhammad, and A. Mubarak, "Penggunaan Internet Dikalangan Siswa Sd Di Kota Ternate: Suatu Survey, Penerapan Algoritma Clustering Dan Validasi Dbi Use Of The Internet In The Elementary School Students In Ternate City: A Survey, Implemented Of Clustering Algorithm And Validation Dbi," vol. 7, no. 6, 2020, doi: 10.25126/jtiik.202072370.

[17] N. Sari, H. H. Handayani, and A. M. Siregar, "Implementasi Clustering Data Kasus Covid 19 Di Indonesia Menggunakan Algoritma K-Means," *Bianglala Informatika*, vol. 11, no. 1, pp. 7–12, Mar. 2023, doi: 10.31294/bi.v11i1.14762.

[18] Nopiti Yulistiani, Ayu Ratna Juwita, and Anis Fitri Nur Masruriyah, "Pengaruh Outlier pada Algoritma K-Medoid untuk Mengelompokan Rekanan Vendor dalam Pengadaan Barang," *Scientific Student Journal for Information, Technology and Science*, vol. V NO.2, 2024.

[19] N. A. Kamilah, T. Rohana, R. Rahmat, and A. Fauzi, "Implementasi Algoritma K-Means dan K-Medoids Dalam Klasterisasi Kasus Kekerasan Terhadap Perempuan," *Jurnal Media Informatika Budidarma*, vol. 8, no. 2, p. 810, Apr. 2024, doi: 10.30865/mib.v8i2.7558.

[20] M. Mughnyanti, S. Efendi, and M. Zarlis, "Analysis of determining centroid clustering x-means algorithm with davies-bouldin index evaluation," in *IOP Conference Series: Materials Science and Engineering*, Institute of Physics Publishing, Jan. 2020. doi: 10.1088/1757-899X/725/1/012128.

This Journal is licensed under a Creative Commons Attribution 4.0 International License