

Klasterisasi Supplier Berdasarkan Kinerja Menggunakan Algoritma K-Means

Alfina Fadhilah Afra^a, April Lia Hananto^b, Agustia Hananto^c, Bayu Priyatna^d

^{a,b,c,d}Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Buana Perjuangan Karawang
Email: si20.afinaafra@mhs.ubpkarawang.ac.id^a, aprilia@ubpkarawang.ac.id^b, agustia.hananto@ubpkarawang.ac.id^c, bayu.priyatna@ubpkarawang.ac.id^d

Submitted: 15-04-2025, Reviewed: 24-04-2025, Accepted 28-04-2025
<https://doi.org/10.47233/jteksis.v7i2.1935>

Abstract

Supplier performance evaluation has an important role in maintaining operational efficiency and product quality. PT Percetakan until now still uses a subjective approach in assessing suppliers. This study was conducted to classify suppliers based on performance using the *k*-Means algorithm, which was chosen because of its ability to process big data efficiently and provide more objective results. The Data used came from the general procurement of goods in 2019-2021, with variable order quantities, unit prices, and delivery time ranges. Cluster quality evaluation was conducted using Davies-Bouldin Index (DBI), Calinski-Harabasz Index (CHI), and Sum of Squared Error (SSE). The results showed a DBI value of 0.7462, which indicates a fairly good cluster separation; CHI of 53.9515 indicates a significant separation between clusters; and SSE of 4.8337 indicates the compactness of the cluster structure. Three supplier performance clusters were successfully formed. Cluster 1 shows suppliers with high performance, characterized by high order quantities, low unit prices, and fast delivery. Cluster 2 shows moderate performance with high unit price and inconsistent performance. Cluster 3 showed low performance, with few orders and slow delivery. 3D visualization of Scatter plots and Pairplots is used to illustrate the distribution of data and support the interpretation of results. This research contributes in providing an objective approach to supplier evaluation and selection, as well as supporting more accurate decision making in the industrial environment.

Keyword: Supplier Performance, Clustering, K-Means, Procurement, Supply Chain

Abstrak

Evaluasi kinerja *supplier* memiliki peran penting dalam menjaga efisiensi operasional dan kualitas produk. PT Percetakan hingga saat ini masih menggunakan pendekatan subjektif dalam penilaian *supplier*. Penelitian ini dilakukan untuk mengelompokkan *supplier* berdasarkan kinerja menggunakan algoritma *K-Means*, yang dipilih karena kemampuannya mengolah data besar secara efisien dan memberikan hasil yang lebih objektif. Data yang digunakan berasal dari pengadaan barang umum tahun 2019–2021, dengan variabel jumlah pesanan, harga satuan, dan rentang waktu pengiriman. Evaluasi kualitas kluster dilakukan menggunakan *Davies-Bouldin Index (DBI)*, *Calinski-Harabasz Index (CHI)*, dan *Sum of Squared Error (SSE)*. Hasil menunjukkan nilai *DBI* sebesar 0.7462, yang mengindikasikan pemisahan kluster yang cukup baik; *CHI* sebesar 53.9515 menunjukkan separasi antar kluster yang signifikan; dan *SSE* sebesar 4.8337 menunjukkan kompaknya struktur kluster. Tiga kluster kinerja *supplier* berhasil dibentuk. Kluster 1 menunjukkan *supplier* dengan kinerja tinggi, ditandai oleh jumlah pesanan tinggi, harga satuan rendah, dan pengiriman cepat. Kluster 2 menunjukkan kinerja sedang dengan harga satuan tinggi dan performa yang tidak konsisten. Kluster 3 menunjukkan kinerja rendah, dengan pesanan sedikit dan pengiriman lambat. Visualisasi *3D Scatter Plot* dan *Pairplot* digunakan untuk menggambarkan distribusi data dan mendukung interpretasi hasil. Penelitian ini memberikan kontribusi dalam menyediakan pendekatan objektif terhadap evaluasi dan seleksi *supplier*, serta mendukung pengambilan keputusan yang lebih akurat di lingkungan industri.

Kata kunci: Kinerja Supplier, Klasterisasi, *K-Means*, Pengadaan Barang, Rantai Pasok

This work is licensed under Creative Commons Attribution License 4.0 CC-BY International license



PENDAHULUAN

Pengelolaan kinerja *supplier* dalam dunia industri sangat penting dilakukan untuk mendukung proses berjalannya pengadaan barang dan jasa. *Supplier* dengan kinerja baik dapat membantu perusahaan mencapai efisiensi operasional dan meningkatkan kualitas produk mereka. Sebaliknya, *supplier* dengan kinerja yang rendah dapat menyebabkan keterlambatan, peningkatan biaya bahkan dapat mengganggu rantai pasok [1].

PT Percetakan memiliki kebutuhan pengadaan barang seperti alat tulis kantor (ATK), perlengkapan produksi non-bahan utama, serta kebutuhan logistik lainnya. Namun saat ini PT Percetakan belum memiliki sistem penilaian kinerja secara sistematis. Karena selama ini perusahaan mengambil keputusan dalam pemilihan *supplier* hanya berdasarkan pada pengalaman, hubungan bisnis dan penilaian subjektif. Hal ini menimbulkan ketidakefisienan dalam mengambil keputusan dan

risiko yang tinggi dalam memilih *supplier* yang tidak konsisten dalam pengiriman barang ke perusahaan [2]. Menurut peneliti terkait seperti Efendi dan Wahyudin, pemilihan *supplier* dengan pendekatan subjektif dapat meningkatkan risiko ketidakkonsistenan dalam kualitas dan ketepatan waktu pengiriman barang, hal itu lah yang memperkuat penelitian ini [3].

Dari tahun 2019 hingga 2021, PT Percetakan telah bekerja sama dengan 67 *supplier* dengan 4.067 transaksi yang terlibat dalam pengadaan barang. Banyaknya data tersebut membuka peluang untuk dilakukan analisis lebih dalam. Data pengadaan barang tersebut mencakup variabel nama *supplier*, jumlah pesanan, harga satuan dan rentang pengiriman. Dengan pendekatan analitik berbasis data, perusahaan dapat mengoptimalkan proses penilaian kinerja *supplier* untuk memastikan kualitas, efisiensi dan keberlanjutan kerja sama dalam berbisnis [4].

Supplier sangat berperan penting dalam memenuhi kebutuhan Perusahaan [5]. Klasterisasi merupakan salah satu metode analisis data untuk membantu perusahaan dalam mengelompokkan *supplier* berdasarkan kinerjanya. Algoritma *K-means*, sebagai salah satu teknik dalam *data mining* menjadi solusi dalam proses klasterisasi kinerja *supplier* [6][7]. Metode ini dapat mengelompokkan *supplier* berdasarkan variabel tertentu seperti jumlah pesanan, harga satuan dan rentang pengiriman. Dengan algoritma ini, *supplier* dapat dikelompokkan ke dalam beberapa kategori berdasarkan kinerja yang serupa sehingga perusahaan dapat mengambil keputusan yang lebih tepat dalam mengelola rantai pasokan [8] [9].

Penelitian ini selaras dengan penelitian Bakri yang menggunakan algoritma *K-Means* untuk mengelompokkan kualitas batu bara di PLTU Sebalang. Hasil yang didapat dari penelitian tersebut menunjukkan bahwa algoritma *K-Means* efektif dalam mengelompokkan data menjadi tiga klaster yaitu kualitas baik, normal dan buruk [2]. Penelitian oleh Denis yang menerapkan metode *K-Means* untuk mengelompokkan data pembelian kendaraan ke dalam tiga klaster diantaranya barang yang paling diminati, sedikit diminati dan kurang diminati. Penelitian ini memperkuat memilih *K-Means* sebagai metode yang efektif untuk klasterisasi data [4]. Penelitian yang dilakukan oleh Dicky yaitu mengelompokkan obat dan menghasilkan 4 klaster [10].

Sebagai perbandingan dengan metode yang berbeda, penelitian Firza menggunakan algoritma *K-Means* untuk mengelompokkan data siswa berdasarkan minat jurusan di sebuah sekolah kejuruan. Hasil yang didapat memberikan referensi untuk pembagian tempat magang. Penelitian oleh Firza menerapkan metode *K-Means* untuk

memetakan siswa berprestasi, menghasilkan pengelompokkan dengan akurasi sebesar 67% [11]. Peneliti oleh Kristina dan Irawan yang menghasilkan 3 klaster [12]. Penelitian yang dilakukan oleh Saputra dan Yusuf yaitu membandingkan metode *K-Means* dan *DBSCAN* yang menghasilkan 3 point [13]. Penelitian yang dilakukan oleh Fahmi yaitu pengambilan Keputusan dalam pemilihan *supplier* obat, dari 10 *supplier* yang dipilih menghasilkan 1 *supplier* dengan pemasok paling tertinggi [14].

Penelitian sebelumnya umumnya menggunakan pendekatan subjektif dalam evaluasi *supplier* dan jarang menerapkan algoritma *K-Means* pada data riil sektor pengadaan. Selain itu, evaluasi kualitas klaster belum banyak mengintegrasikan metrik *DBI*, *CHI*, dan *SSE* secara bersamaan. Penelitian ini menghadirkan kebaruan melalui pemanfaatan data aktual dari PT Percetakan, penerapan evaluasi klaster yang lebih komprehensif, serta visualisasi hasil klasterisasi untuk mendukung interpretasi dan pengambilan keputusan yang lebih objektif.

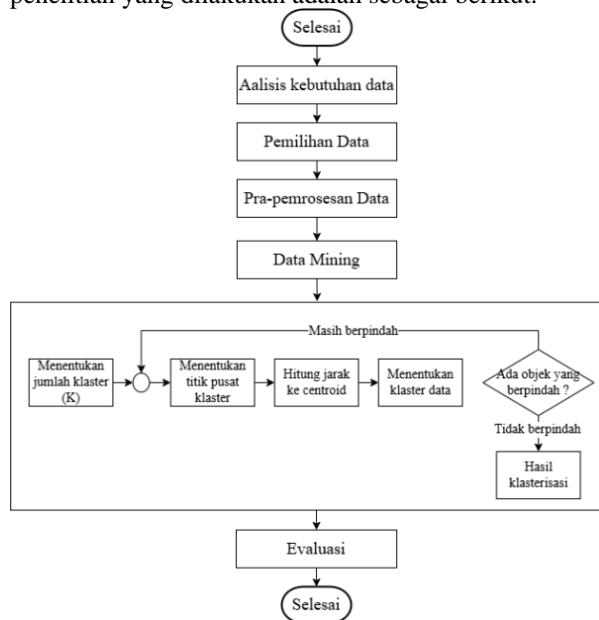
Knowledge Discovery in Database (KDD) merupakan proses mengidentifikasi data yang benar untuk mendukung pengambilan keputusan berdasarkan bukti yang asli atau sah [15]. Tahapan KDD yang pertama *Data Selection*, tahapan ini merupakan tahapan persiapan dalam pemilihan data. Data yang telah didapatkan akan dilakukan proses pemilihan variabel untuk digunakan pada tahap *pre-processing* [16]. Kedua *Pre-Processing*, pada tahap ini setelah data diseleksi, selanjutnya dilakukan proses *Preprocessing*, yaitu menghilangkan duplikat data dan menghapus data yang tidak valid untuk memastikan kualitas analisis [17]. Ketiga *Transformation*, meningkatkan kualitas data, menyesuaikan skala data dan mempermudah analisis data [18]. Keempat *Data mining*, Pada tahap ini, algoritma *K-Means* diterapkan untuk menemukan pola dan mengelompokkan *supplier* berdasarkan kinerja [19]. Kelima *Interpretation/Evaluation*, Evaluasi klaster yang dihasilkan oleh algoritma *K-Means* dilakukan dengan menggunakan *Davies-Bouldin Index (DBI)*, *DBI* digunakan untuk memastikan bahwa klasterisasi *supplier* yang dihasilkan dapat digunakan untuk analisis kinerja yang lebih efektif [20]. *Python* digunakan dalam penelitian ini. *Python* merupakan bahasa pemrograman interpretatif berorientasi objek yang dapat digunakan diberbagai *platform* [21].

Dalam penelitian ini, algoritma *K-Means* dipilih karena lebih efisien dalam mengelompokkan atau berukuran besar seperti dataset pengadaan barang umum [8]. Dibandingkan metode klasterisasi lainnya, *K-Means* memberikan hasil yang lebih terdefinisi untuk atau dengan distribusi yang merata.

Selain itu, pendekatan ini juga menggunakan proses yang objektif berdasarkan data dibandingkan pendekatan subjektif seperti metode pembobotan kriteria. Penelitian ini bertujuan untuk memberikan kontribusi dan solusi yang praktis bagi PT Percetakan untuk menentukan mitra kerja sama melalui penerapan algoritma *K-Means*. Selain itu, pendekatan ini juga dapat menjadi dasar dalam mengambil keputusan yang lebih objektif dan terukur untuk mendukung pengambilan keputusan dalam pemilihan *supplier* terbaik dan memudahkan strategi perusahaan di masa mendatang juga memberikan hasil yang signifikan dalam pengelolaan rantai pasok yang lebih baik [18].

METODE PENELITIAN

Metode yang digunakan dalam penelitian ini adalah *K-Means Clustering*. Data yang digunakan pada penelitian ini adalah data pengadaan barang umum tahun 2019-2021 sebanyak 4.067 baris data yang diambil di PT Percetakan dengan izin dari kepala seksi lansokmansak. Penelitian ini dilakukan dengan mengacu pada kerangka kerja *Knowledge Discovery in Databases (KDD)*. Adapun tahap penelitian yang dilakukan adalah sebagai berikut.



Gambar 1 Prosedur Penelitian

Analisis Kebutuhan Data

Supplier yang dianalisis dalam penelitian ini adalah *supplier* yang berkontribusi dalam pengadaan barang umum perusahaan, seperti Alat Tulis Kantor (ATK), perlengkapan produksi non-bahan utama, dan kebutuhan logistik lainnya. Berikut variabel utama yang akan digunakan dalam penelitian [16]:

1. *Supplier* : Identitas unik dari setiap *supplier*.

2. Jumlah Pesanan : Total kuantitas barang yang dipesan oleh perusahaan.
3. Harga Satuan : Biaya per unit barang.
4. Rentang Pengiriman : Waktu antara pemesanan barang hingga diterimanya barang oleh perusahaan.

Pemilihan Data

Pemilihan data dilakukan untuk memastikan bahwa data yang digunakan relevan dengan tujuan penelitian, yaitu mengelompokkan *supplier* berdasarkan kinerja menggunakan algoritma *K-Means*. Beberapa variabel yang digunakan dalam penelitian ini meliputi *Supplier*, Jumlah Pesanan, Harga Satuan dan Rentang Pengiriman. Variabel ini dipilih karena dianggap memiliki keterkaitan langsung dalam menilai kinerja pemasok. Data yang tidak relevan, duplikat, atau memiliki nilai kosong dihapus untuk menjaga kualitas analisis. Selain itu, *outlier* pada data diperiksa untuk menghindari distorsi pada hasil klusterisasi [16]. Berikut data yang sudah dipilih.

Supplier	Jumlah Pesanan	Harga Satuan	Rentang Pengiriman
CPTSMVM	822	1100000	45,5555556
CV_BT	200	1620000	91
CV_CM	125	3100000	30
CV_JMA	258	5750000	40,91358025
CV_KC	668	900000	90
CV_DM	171	1750000	35,28571429
CV_MAB	467	6300000	31,28571429
CV_SAS	115	3200000	33,09090909
CV_SS	324	3900000	55,11111111

Gambar 2 Dataset Pengadaan Barang Umum

Pra-pemrosesan Data

Pada tahap ini dilakukan pembersihan, transformasi dan normalisasi data untuk menghilangkan data yang duplikat dan tidak relevan untuk menangani *missing value* dan memperbaiki kesalahan *input* dengan *Google Colab* menggunakan *Phyton*. Metode normalisasi yang digunakan dalam analisis ini adalah *Min-Max Scaling* dan *Z-score Normalization* [17].

Min-Max Scaling

Min-Max Scaling digunakan untuk merubah nilai data ke dalam rentang [0,1] agar semua atribut memiliki skala yang sama [17]. Rumus *Min-Max Scaling* adalah sebagai berikut:

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (1)$$

Keterangan:

X' = Nilai hasil normalisasi

X = Nilai asli

X_{min} = Nilai minimum dalam dataset

X_{max} = Nilai maksimum dalam dataset

Dengan metode ini, nilai terkecil dalam dataset akan menjadi 0, sedangkan nilai terbesar akan menjadi 1 [17].

Z-score Normalization

Z-score Normalization digunakan untuk merubah data sehingga memiliki distribusi dengan $mean = 0$ dan standar deviasi = 1 [17]. Rumusnya adalah:

$$Z = \frac{x - \mu}{\sigma} \quad (2)$$

Keterangan:

Z = Nilai hasil normalisasi

X = Nilai asli

μ = Rata-rata (mean) dari dataset

σ = Standar deviasi dari dataset

Metode ini digunakan jika data memiliki distribusi yang berbeda-beda dan perlu dinormalisasi agar memiliki distribusi normal [17].

Data Mining

Tahapan ini untuk memproses data yang ada. Algoritma yang digunakan adalah algoritma *K-means* untuk mengelompokkan *supplier* berdasarkan kinerja. Dalam menentukan jumlah kluster optimal dengan *Elbow Method* dan *Silhouette Score* [19].

Menentukan nilai K optimal

Elbow Method

Metode ini menggunakan *Within-Cluster Sum of Squares (WCSS)*, yakni total jarak kuadrat antara setiap titik data ke *centroid* klasternya. Rumus WCSS adalah:

$$WCSS = \sum_{i=1}^K \sum_{x \in C_i} \|x - \mu_i\|^2 \quad (3)$$

Keterangan:

x = Data dalam kluster

μ_i = Centroid kluster ke-i

C_i = Kluster ke-i

K = Jumlah kluster

Setelah menghitung WCSS untuk beberapa nilai K, hasilnya diplot dalam grafik. Titik siku (*elbow*) dari grafik menunjukkan K yang optimal [19].

Silhouette Score

Metode ini mengukur seberapa baik setiap data berada dalam klasternya dan seberapa jauh dari kluster lain. Rumusnya adalah sebagai berikut:

$$S = \frac{b-a}{\max(a,b)} \quad (4)$$

Keterangan:

S = Skor *Silhouette*

a = Rata-rata jarak data ke titik lain dalam kluster yang sama

b = Rata-rata jarak data ke titik terdekat di kluster lain

Nilai *Silhouette Score* berkisar antara -1 hingga 1, di mana semakin mendekati 1, semakin baik klusterisasi [19].

Klusterisasi Data dengan K-Means

Di tahap ini jika hasil klusterisasi masih kurang optimal, menggunakan *K-Means* untuk perulangannya, alasannya untuk menghasilkan hasil lebih stabil dan menghindari *centroid* awal yang buruk [19].

Evaluasi Kluster

Setelah menentukan K dan menjalankan *K-Means*, hasil klusterisasi perlu dievaluasi, metode yang digunakan adalah *Davies-Bouldin Index (DBI)* untuk mengukur seberapa baik kluster terpisah dan seberapa kompak kluster tersebut. Semakin rendah nilai *DBI* maka kluster lebih baik, lalu *Calinski-Harabasz Index (CHI)* dan *SSE/Inertia* [19].

$$DBI = \frac{1}{K} \sum_{i=1}^K \max_{j \neq i} \left(\frac{\sigma_i + \sigma_j}{d_{ij}} \right) \quad (5)$$

Setelah proses klusterisasi selesai, analisis hasil klusterisasi akan menggunakan *scatter plot* untuk visualisasi. Berdasarkan pengolahan data yang dilakukan dengan 4 variabel yang ada, nantinya akan menghasilkan 3 *cluster* yaitu kluster dengan kinerja tinggi, kinerja sedang dan kinerja rendah [20].

Hasil Klusterisasi yang didapat nantinya akan dilakukan tindakan. Kluster tinggi, yaitu *supplier* dengan kualitas terbaik dan tepat waktu akan direkomendasikan sebagai mitra utama. Kluster sedang, yaitu *supplier* masih cukup baik tetapi masih memiliki sedikit ketidakkonsistenan. Kluster rendah, yaitu *supplier* yang memiliki beberapa kendala dalam ketepatan waktu atau kualitas, akan dilakukan pembinaan untuk peningkatan layanan terhadap Perusahaan [22].

HASIL DAN PEMBAHASAN

Dataset yang dianalisis dalam penelitian ini adalah *supplier* yang berkontribusi dalam pengadaan barang umum Perusahaan seperti Alat Tulis Kantor (ATK), perlengkapan produksi non-bahan utama dan kebutuhan logistik lainnya. Variabel yang digunakan adalah *Supplier*, Jumlah Pesanan, Harga Satuan dan Rentang Pengiriman.

Pemilihan data dilakukan untuk memastikan bahwa data yang digunakan sudah relevan. Berikut hasil pemilihan data pada Gambar 3.

```

Supplier Jumlah Pesanan Harga Satuan Rentang Pengiriman
0 CPTSMH 822 11000000 45.555556
1 CV. BT 200 16200000 91.000000
2 CV. CM 125 31000000 30.000000
3 CV. JMA 258 57500000 40.913580
4 CV. KC 668 90000000 90.000000
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 67 entries, 0 to 66
Data columns (total 4 columns):
# Column Non-Null Count Dtype
---
0 Supplier 67 non-null object
1 Jumlah Pesanan 67 non-null int64
2 Harga Satuan 67 non-null int64
3 Rentang Pengiriman 67 non-null float64
dtypes: float64(1), int64(2), object(1)
memory usage: 2.2+ KB
None
    
```

Gambar 3 Pemilihan Data

Hasil diatas menunjukkan bahwa 67 data *non-nul*, yang artinya adalah dataset yang digunakan tidak ada masalah (normal).

Transformasi dan normalisasi dilakukan untuk menghilangkan data duplikat dan tidak relevan untuk menangani *missing value*. Metode yang digunakan adalah metode *Min-Max Scalling* dan *Z-Score Normalization*.

Menormalisasi data menggunakan metode *Min-Max Scaller* pada gambar 4.

```

Supplier Jumlah Pesanan Harga Satuan Rentang Pengiriman
0 CPTSMH 1.000000 0.022099 0.261066
1 CV. BT 0.209657 0.079558 1.000000
2 CV. CM 0.114358 0.243094 0.008130
3 CV. JMA 0.283355 0.535912 0.185587
4 CV. KC 0.804320 0.000000 0.983740
    
```

Gambar 4 Normalisasi dengan metode *Min-Max Scalling*

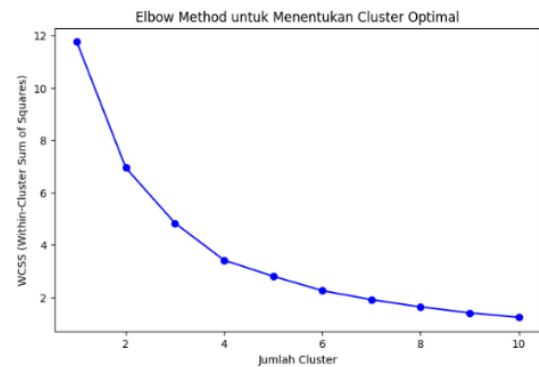
Menormalisasi data menggunakan metode *Z-Score Normalization* pada gambar 5.

```

Supplier Jumlah Pesanan Harga Satuan Rentang Pengiriman
0 CPTSMH 3.177673 -1.012204 0.327958
1 CV. BT -0.544573 -0.814084 3.760715
2 CV. CM -0.993397 -0.250203 -0.847068
3 CV. JMA -0.197482 0.759447 -0.022685
4 CV. KC 2.256088 -1.088404 3.685178
    
```

Gambar 5 Normalisasi dengan metode *Z-Score Normalization*

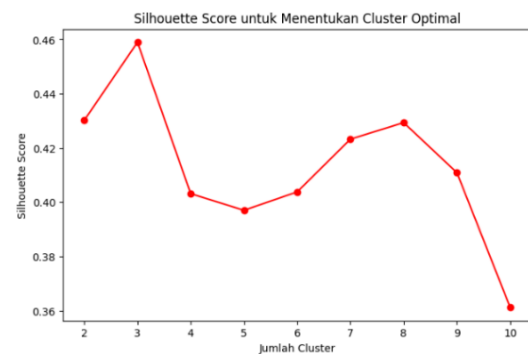
Setelah menentukan dengan dua metode di atas selanjutnya menentukan *cluster* optimal menggunakan metode *Elbow*. Metode ini menggunakan *Within-Cluster Sum of Square (WCSS)*, yaitu keseluruhan jarak kuadrat antara setiap titik data ke *centroid* klasternya. Berikut hasilnya pada gambar 6.



Gambar 6 Metode *Elbow* untuk menentukan *Cluster* Optimal

Gambar 6 menunjukkan hasil metode *Elbow* untuk menentukan jumlah kluster optimal. Pada grafik tersebut, sumbu X menunjukkan jumlah kluster, sementara sumbu Y menunjukkan nilai *Within-Cluster Sum of Squares (WCSS)*. Dari grafik terlihat bahwa terjadi penurunan tajam hingga kluster ke-3, kemudian berkurang secara bertahap setelahnya. Oleh karena itu, titik siku (*elbow point*) yang menjadi indikasi jumlah kluster optimal berada pada $K = 3$.

Selain menggunakan metode *Elbow*, pemilihan jumlah kluster juga menggunakan metode *Silhouette Score*. Berikut hasilnya pada Gambar 7.



Gambar 7 *Silhouette Score* untuk Menentukan *Cluster* Optimal

Pada Gambar 7, nilai *Silhouette Score* tertinggi adalah 3, karena pada titik ini nilai *Silhouette Score* mencapai puncaknya (tertinggi), menunjukkan tampilan data yang paling jelas dan efektif.

Setelah menentukan hasil optimal kluster, maka dilakukan penerapan *K-Means*. Berikut hasil klusterisasi pada Gambar 8.

Supplier	Cluster
0 CPTSMH	1
1 CV. BT	2
2 CV. CM	1
3 CV. JMA	0
4 CV. KC	2
5 CV. LM	1
6 CV. MAB	0
7 CV. SAS	1
8 CV. SS	1
9 CV. TU	1

Gambar 8 Hasil Klasterisasi

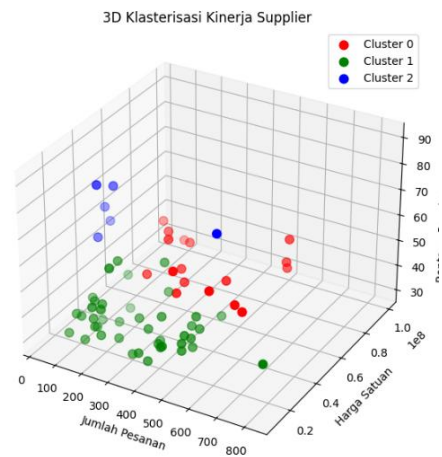
Gambar 8. menunjukkan hasil klasterisasi K-Means. Setelah mendapatkan hasil klasterisasi, maka dilanjutkan dengan mengevaluasi klasterisasi menggunakan 3 matriks internal yakni *Davies-Bouldin Index (DBI)*, *Calinski-Harabasz Index (CHI)* dan *Sum of Squared Errors (SSE)/Inertia*, berikut hasilnya dalam tabel 1.

<i>Davies-Bouldin Index (DBI)</i>	0.7462
<i>Calinski-Harabasz Index (CHI)</i>	53.9515
<i>Sum of Squared Errors (SSE)</i>	4.8337

Tabel 1 Evaluasi Klasterisasi 3 Matriks Internal

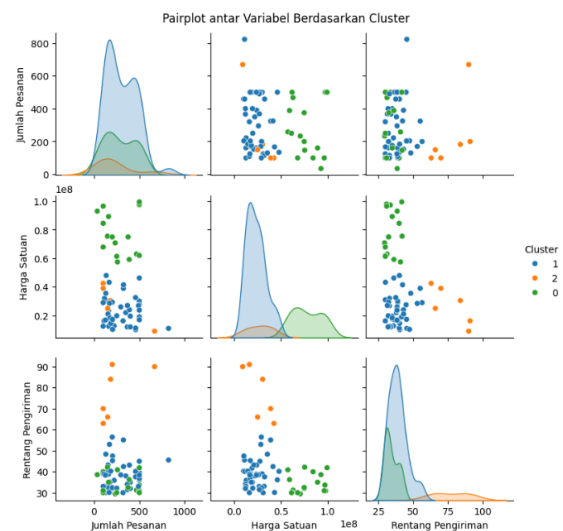
Pada tabel 1 menunjukkan hasil evaluasi klasterisasi menggunakan 3 matriks internal. *Davies-Bouldin Index (DBI)* merupakan matriks yang mengukur tingkat kemiripan antar klaster yang dimana semakin kecil *DBI*, maka semakin baik pemisahan antar klaster yang terbentuk. Pada tabel ini, nilai *DBI* sebesar 0.7462 yang menunjukkan bahwa klaster yang dihasilkan memiliki tingkat pemisahan yang cukup baik antar satu sama lain. *Calinski-Harabasz Index (CHI)* mengevaluasi rasio antar klaster dengan dispersi dalam klaster. Nilai *CHI* yang tinggi mengindikasikan bahwa klaster yang terbentuk saling terpisah dengan baik dan masing-masingnya memiliki konsistensi internal yang tinggi. Nilai *CHI* sebesar 53.9515 menunjukkan bahwa pemisahan antar klaster cukup signifikan. *Sum of Squared Errors (SSE)* atau bisa disebut inerti, mengukur total jarak kuadrat antara setiap data terhadap pusat klaster. Nilai *SSE* yang rendah menunjukkan bahwa data dalam tiap klaster berada dekat dengan centroidnya yang berarti klaster tersebut padat dan konsisten. Nilai *SSE* sebesar 4.8337 menunjukkan bahwa variasi dalam klaster relatif kecil dan struktur klasternya terbentuk cukup kompak.

Selanjutnya ialah memvisualisasikan menggunakan *3D Scatter Plot*, berikut hasilnya pada gambar 9.



Gambar 9 Visualisasi Hasil Klaster Menggunakan Scatter Plot

Visualisasi *3D Scatter Plot* pada gambar 9 menunjukkan visualisasi klaster berdasarkan 3 variabel utama yaitu jumlah pesanan, harga satuan dan rentang pengiriman. Visualisasi tersebut bertujuan untuk melihat sebaran data *supplier* dalam grafik 3D. Klaster 0 ditandai warna merah, cenderung terdiri atas *supplier* dengan harga satuan tinggi dan jumlah pesanan bervariasi. Rentang pengirimannya berada pada level sedang hingga tinggi. Klaster 1 ditandai warna hijau, menunjukkan kelompok *supplier* dengan kinerja yang relatif stabil, di mana jumlah pesanan tinggi namun harga satuan relatif rendah. Rentang pengirimannya cenderung berada dikisaran rendah hingga sedang. Klaster 2 ditandai warna biru, merupakan klaster dengan jumlah pesanan sangat rendah, harga satuan rendah dan rentang pengiriman yang tinggi. Kelompok ini merupakan *supplier* berkinerja rendah.



Gambar 10 Pairplot antar Variabel Berdasarkan Cluster

Visualisasi *pairplot* antar variabel Berdasarkan Klaster pada gambar 10 menunjukkan *pairplot* antar variabel yang membantu dalam mengamati hubungan antara kombinasi 2 variabel terhadap distribusi klaster. Dari *pairplot* di atas dapat diamati bahwa terdapat perbedaan distribusi antar klaster terutama pada variabel harga satuan dan rentang pengiriman, klaster 1 mendominasi kelompok dengan harga satuan rendah dan jumlah pesanan tinggi, klaster 0 tersebar pada berbagai nilai jumlah pesanan namun memiliki kecenderungan harga satuan yang tinggi, klaster 2 terlihat terkonsentrasi pada nilai jumlah pesanan rendah namun dengan rentang pengiriman yang lebih tinggi. Visualisasi tersebut memperkuat interpretasi bahwa kinerja *supplier* dapat dideskripsikan dengan cukup baik berdasarkan kombinasi ketiga variabel tersebut.

Interpretasi hasil, klaster 1 merupakan *supplier* yang memiliki jumlah pesanan tinggi, harga satuan relatif rendah dan rentang pengiriman cepat. Karakteristik ini menunjukkan efisiensi dan reliabilitas tinggi dari sisi biaya maupun waktu pengiriman, klaster ini berkinerja tinggi. Klaster 0 merupakan *supplier* yang memiliki performa beragam dengan harga satuan tinggi namun tidak selalu disertai dengan jumlah pesanan yang tinggi. Rentang pengiriman pun cenderung bervariasi, menunjukkan adanya inkonsistensi dalam kinerja, klaster ini berkinerja sedang. Klaster 2 merupakan klaster yang memiliki jumlah pesanan rendah dan rentang pengiriman tinggi, meskipun harga satuan tidak terlalu tinggi, ketidakefisienan dalam pengiriman dan minimnya permintaan membuat klaster ini digolongkan berkinerja rendah.

SIMPULAN

Evaluasi menggunakan 3 matriks internal menghasilkan *DBI* sebesar 0,7462, *CHI* sebesar 53,9515, dan *SSE* sebesar 4,8337, yang menunjukkan pemisahan antar klaster yang baik dan struktur klaster yang kompak. Hasil klasterisasi membentuk tiga kelompok *supplier*: kinerja tinggi (Klaster 1), sedang (Klaster 2) dan rendah (Klaster 3). Visualisasi 3D Scatter Plot dan Pairplot mendukung pemahaman karakteristik masing-masing klaster. Hasil ini dapat dimanfaatkan sebagai dasar kebijakan evaluasi dan seleksi *supplier*. *Supplier* berkinerja tinggi sebaiknya diprioritaskan, sedangkan *supplier* berkinerja rendah perlu ditinjau ulang. Sistem pemantauan kinerja berbasis data juga perlu diterapkan secara berkala.

Untuk penelitian selanjutnya, disarankan mengeksplorasi algoritma lain seperti *DBSCAN*

atau *Fuzzy C-Means*, serta menambahkan variabel seperti tingkat retur, kualitas barang, dan kepuasan pengguna guna memperkaya analisis.

UCAPAN TERIMAKASIH

Saya sangat berterima kasih banyak kepada diri saya sendiri yang sudah bertahan, berjuang dan tidak menyerah hingga berhasil menyelesaikan jurnal ini. Saya bangga kepada diri saya sendiri atas segala usaha dan ketekunan yang saya lakukan. Terima kasih untuk kedua orang tua yang selalu mendoakan, mendukung dan memberi kasih sayang tiada henti. Terima kasih kepada dosen pembimbing yang telah banyak memberi saya bimbingan dan ilmu yang sangat berarti selama penyusunan jurnal ini. Tidak lupa, kepada orang terkasih yang juga selalu menemani, mendukung dan menjadi penyemangat di setiap langkah saya, saya ucapkan terima kasih dengan sangat tulus.

DAFTAR PUSTAKA

- [1] N. Yulistiani *et al.*, "Pengaruh Outlier pada Algoritma K-Medoid untuk Mengelompokkan Rekanan Vendor dalam Pengadaan Barang," vol. V, pp. 145–151, 2024.
- [2] M. Bakri, "PENERAPAN DATA MINING UNTUK CLUSTERING KUALITAS BATU BARA DALAM PROSES PEMBAKARAN DI PLTU SEBALANG MENGGUNAKAN METODE K-MEANS," vol. 11, no. 1, pp. 1–4, 2017.
- [3] R. Y. Efendi and W. Wahyudin, "Pemilihan Pemasok Bahan Baku Tetap Menggunakan Metode Analytical Hierarchy Process (Studi Kasus pada CV. Mulia Tata Sejahtera)," *J. Serambi Eng.*, vol. 8, no. 3, 2023, doi: 10.32672/jse.v8i3.6039.
- [4] Susilo D, "Implementasi Data Mining dalam Pengelompokan Data Pembelian Menggunakan Algoritma K-Means Pada PT. Otomotif 1".
- [5] W. Apriliah, N. Subekti, and T. Haryati, "Penerapan Model Waterfall Dalam Perancangan Aplikasi Sistem Informasi Simpan Pinjam Pada Koperasi PT. CHIYODA INTEGRAL INDONESIA KARAWANG," *J. Interkom J. Publ. Ilm. Bid. Teknol. Inf. dan Komun.*, vol. 14, no. 2, pp. 34–42, Jul. 2019, doi: 10.35969/interkom.v14i2.50.
- [6] A. Choiruddin and I. Pratama, "Sistem Monitoring Siswa untuk Guru dan Orang Tua Siswa Menggunakan Algoritma Fuzzy C-Means di SD Muhammadiyah Demangrejo," *J. Teknol. Dan Sist. Inf. Bisnis*, vol. 4, no. 1, pp. 122–132, 2022, doi: 10.47233/jteksis.v4i1.387.
- [7] A. Ramadhan and S. Supatman, "Sistem Pendukung Keputusan Pemilihan Supplier Pada PT. Avo Innovation Technology Dengan Metode Simple Additive Weighting (SAW)," *J. Teknol. Dan Sist. Inf. Bisnis*, vol. 4, no. 2, pp. 256–267, 2022, doi: 10.47233/jteksis.v4i1.484.
- [8] B. Huda, E. Novalia, and F. Nurapriani, "Klasterisasi Data Obat dengan Algoritma K-Means (Kasus pada UPTD Puskesmas Curug)," vol. 8, pp. 120–130, 2024.
- [9] O. Andanu, U. Y. Sundari, and E. D. P. Setyowati, "Improving Performance of the Banana Agroindustry Supply Chain Using Mcdm (Multi Criteria Decision Making)," *J. Agroindustri*, vol. 13, no. 2, pp. 95–106, 2023, doi: 10.31186/jagroindustri.13.2.95-106.
- [10] M. D. Kurniawan, B. Priyatna, and F. Nurapriani, "Implementasi Algoritma K-Means Untuk Klasterisasi

- Data Obat Puskesmas Kotabaru,” vol. 7, no. September, pp. 882–890, 2023.
- [11] I. Firza, “Penerapan Algoritma K-Means Dalam Metode Clustering Untuk Peminatan Jurusan Bagi Siswa Swasta Pelita Raya Kota Jambi,” vol. 5, no. 3, pp. 371–382, 2020.
- [12] S. Kristina and V. S. Irawan, “Perancangan Kriteria Evaluasi Kinerja Supplier dengan Menggunakan Metode Fuzzy-AHP di PT X,” *J. Telemat.*, vol. 13, no. 1, pp. 43–48, 2019, doi: 10.61769/telematika.v13i1.208.
- [13] A. Saputra and R. Yusuf, “Perbandingan Algoritma DBSCAN dan K-MEANS dalam Segmentasi Pelanggan Pengguna Transportasi Publik Transjakarta Menggunakan Metode RFM,” *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 4, pp. 1346–1361, 2024, doi: 10.57152/malcom.v4i4.1516.
- [14] F. Al Khadzik, B. Huda, E. Novalia, and S. S. Hilabi, “Sistem Pemilihan Supplier Obat Menerapkan Metode Additive Ratio Analysis (ARAS),” *J. Sist. Komput. dan Inform.*, vol. 5, no. 3, p. 620, 2024, doi: 10.30865/json.v5i3.7499.
- [15] D. Nofriansyah and I. Mariami, “Implementasi Data Mining Untuk Pengelempokan Buku Di Perpustakaan Yayasan Nurul Islam Indonesia Baru Dengan Metode K-Means Clustering,” vol. 1, no. 1, pp. 1–12, 2021.
- [16] M. Rafi, “Algoritma K-Means untuk Pengelompokan Topik Skripsi Mahasiswa,” vol. 12, no. 2, pp. 121–129, 2020.
- [17] N. Luh, P. Purnama, I. N. Purnama, and N. W. Utami, “Penerapan Data Mining Untuk Clustering Penilaian Kinerja Dosen Menggunakan Algoritma K-Means (Studi Kasus : STMIK Primakara),” vol. 16, no. 2, pp. 105–112, 2022.
- [18] M. H. Adiya and Y. Desnelita, “Jurnal Nasional Teknologi dan Sistem Informasi Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan Pada RSUD Pekanbaru,” vol. 01, pp. 17–24, 2019.
- [19] A. L. Hananto, P. Assiroj, and B. Priyatna, “Analysis Of Drug Data Mining With Clustering Technique Using K-Means Algorithm Analysis Of Drug Data Mining With Clustering Technique Using K-Means Algorithm,” 2021, doi: 10.1088/1742-6596/1908/1/012024.
- [20] Y. R. Sari, “Yunita, Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Pada Penerimaan Mahasiswa Baru (Studi Kasus : Universitas Islam Indragiri) 1 238,” vol. 7, no. September, pp. 238–249, 2018.
- [21] S. S. Nanda, “Clustering Penerima Bantuan Pangan Non Tunai (Bpnt),” 2023.
- [22] P. I and MZ HAprilyanti S, “Jurnal Manajemen Industri dan Logistik PEMILIHAN SUPPLIER TERBAIK PENYEDIA BARANG CONSUMABLE MENGGUNAKAN METODE ANALYTICAL HIERARCHY PROCESS (Studi kasus di Departemen Pengadaan Barang PT . PUSRI) THE BEST SUPPLIER SELECTION OF CONSUMABLE GOODS SUPPLIER,” pp. 147–158, 2018.