



## **Implementasi Data Mining Untuk Menentukan Persediaan Stok Barang Menggunakan Metode K-Means Clustering**

**Ammar Farisi<sup>1\*</sup>, Zaehol Fatah<sup>2</sup>**

<sup>1</sup> Teknologi Informasi, Universitas Ibrahimy

<sup>2</sup> Sistem Informasi, Universitas Ibrahimy

<sup>1</sup> [ammarfarisi092@gmail.com](mailto:ammarfarisi092@gmail.com)

### **Abstrak**

Data mining membantu bisnis membuat produksi yang menguntungkan dan penyesuaian operasional. Pengelolaan persediaan stok barang merupakan aspek krusial dalam manajemen bisnis ritel modern yang memerlukan pendekatan analitis untuk optimalisasi. Data Mining dengan metode K-Means Clustering mampu menganalisis dan mengelompokkan pola persediaan stok barang. Tahapan yang dilakukan pra-pemrosesan data, implementasi K-Means Clustering, dan evaluasi model menggunakan metrik Sum of Squared Errors (SSE) dan Silhouette Score. Hasil analisis menggunakan metode Elbow menunjukkan tiga cluster optimal dengan nilai SSE final 51.69, menandakan kohesi internal yang baik. Evaluasi menggunakan Silhouette Score menghasilkan nilai 0.498, mengindikasikan keseimbangan yang baik antara kohesi dan separasi cluster. Hasil clustering mengidentifikasi tiga pola distinct dalam persediaan stok: Cluster 0 mendominasi periode akhir tahun dengan rentang stok 780-840 unit, Cluster 1 terkonsentrasi pada awal tahun dengan rentang 840-920 unit, dan Cluster 2 sebagai cluster transisi pada pertengahan tahun dengan rentang 840-880 unit. Visualisasi temporal menunjukkan pola seasonal yang jelas dalam pengelolaan persediaan sepanjang tahun. Implementasi K-Means Clustering berhasil menghasilkan pengelompokan yang bermakna dan dapat diinterpretasikan dalam konteks manajemen inventori.

**Kata Kunci:** Data Mining, K-Means Clustering, Manajemen Inventori, Analisis Persediaan

### **PENDAHULUAN**

Di era perkembangan teknologi saat ini, pengelolaan persediaan barang menjadi salah satu aspek krusial dalam manajemen bisnis ritel modern. Berdasarkan dataset yang diperoleh dari Kaggle, yang mencakup data persediaan harian dengan lima kategori produk (A, B, C, D, dan E) selama periode bulan Januari 2019 sampai dengan Desember 2019, terlihat adanya fluktuasi signifikan dalam tingkat persediaan yang menunjukkan kompleksitas dalam pengelolaan stok. Data menunjukkan variasi stok harian yang berkisar antara 594 hingga 1091 unit untuk berbagai kategori produk, mengindikasikan tantangan dalam mengoptimalkan tingkat persediaan(Aufa Hanif & Wahyuni, 2024). Penentuan stok barang yang tepat merupakan tantangan utama yang dihadapi dalam memenuhi permintaan konsumen secara efisien. Berdasarkan analisis awal dataset, terlihat bahwa setiap kategori produk memiliki pola permintaan yang berbeda. Sebagai contoh, diawal bulan januari kategori C menunjukkan tingkat persediaan yang konsisten lebih tinggi (rata-rata 960-990 unit) dibandingkan kategori lainnya, sementara kategori B menunjukkan fluktuasi yang lebih besar dengan rentang 681-726 unit. Ketidakakuratan dalam penentuan stok dapat mengakibatkan kelebihan maupun kekurangan barang yang berujung pada kerugian bisnis(Muningsih & Kiswati, 2015).

Dalam upaya mengatasi kompleksitas prediksi stok produk, implementasi teknik data mining, khususnya metode K-Means Clustering, menawarkan pendekatan yang menjanjikan. Metode ini dipilih karena kemampuannya dalam mengolah data dalam jumlah besar dan membagi data ke dalam kelompok berdasarkan karakteristik yang sama. Sebagaimana yang ditunjukkan dalam penelitian sebelumnya tentang penerapan metode k-means dalam penjualan produk, metode ini terbukti efektif dalam memahami pola permintaan produk dan mencegah penumpukan barang di gudang(Safira et al., 2022). Dataset dari Kaggle yang digunakan dalam penelitian ini memiliki karakteristik yang ideal untuk implementasi K-Means Clustering, dengan 5 kategori produk yang memiliki pola persediaan berbeda. Data harian yang tersedia memungkinkan analisis mendalam tentang pola permintaan dan fluktuasi stok untuk setiap kategori produk. Penggunaan Python sebagai bahasa pemrograman utama dalam implementasi, didukung oleh berbagai library analisis data, akan memungkinkan pengolahan data yang efisien dan visualisasi hasil yang komprehensif untuk memprediksi bagaimana stok barang kedepannya(Asmoro et al., 2022).

Implementasi K-Means *Clustering* dalam konteks dataset ini diharapkan dapat menghasilkan beberapa manfaat utama. Pertama, optimasi persediaan melalui pengelompokan produk berdasarkan pola permintaan, yang dapat membantu menghindari kekurangan atau kelebihan stok. Kedua, peningkatan akurasi prediksi dengan mempertimbangkan karakteristik unik setiap kategori produk yang terlihat dalam dataset. Ketiga, efisiensi operasional melalui pengolahan data transaksi yang lebih efektif dan pengelompokan data untuk strategi bisnis yang lebih baik.



Penelitian ini bertujuan untuk mengembangkan sistem klasifikasi barang menggunakan metode K-Means *clustering* untuk membantu proses pengelompokan barang berdasarkan pola permintaan(Safira et al., 2022). Dengan memanfaatkan dataset yang mencakup informasi persediaan harian dari berbagai kategori produk, penelitian ini diharapkan dapat memberikan wawasan berharga dalam optimasi manajemen persediaan. Implementasi akan dilakukan menggunakan Python, memanfaatkan kekayaan library untuk analisis data dan machine learning yang tersedia, untuk menghasilkan solusi yang dapat diterapkan secara praktis dalam konteks bisnis ritel(Jalil et al., 2024).

Penelitian ini bertujuan untuk mengembangkan sistem klasifikasi barang menggunakan metode K-Means clustering dengan RapidMiner untuk membantu proses pengelompokan barang berdasarkan pola permintaan. Dengan memanfaatkan dataset yang mencakup informasi persediaan harian dari berbagai kategori produk, penelitian ini diharapkan dapat memberikan wawasan berharga dalam optimasi manajemen persediaan. Implementasi menggunakan RapidMiner akan memanfaatkan berbagai operator dan fitur visualisasi yang tersedia untuk menghasilkan solusi yang dapat diterapkan secara praktis dalam konteks bisnis ritel. Melalui pendekatan ini, diharapkan dapat diperoleh pemahaman yang lebih baik tentang pola persediaan stok dan strategi pengelolaan yang lebih efektif untuk mengoptimalkan manajemen inventori.

## METODE

### Pengumpulan Data

Dalam penelitian ini, data yang digunakan bersumber dari dataset yang tersimpan dalam format CSV (Comma Separated Values) dengan nama file 'dataset.csv'. Dataset ini memiliki struktur yang terdiri dari enam kolom utama, yaitu satu kolom tanggal (date) yang berfungsi sebagai temporal marker dan lima kolom yang merepresentasikan nilai stok untuk masing-masing kategori produk (A, B, C, D, dan E). Kolom tanggal menyimpan informasi waktu pencatatan data dalam format datetime, sementara kolom-kolom kategori produk memuat nilai numerik yang menunjukkan jumlah stok tersedia untuk setiap kategori pada tanggal tersebut(Ika Anikah et al., 2022).

Proses pengumpulan data dilakukan dengan pencatatan nilai stok secara harian untuk semua kategori produk. Untuk meningkatkan efektivitas analisis dan mengurangi noise pada data harian, dilakukan konversi data ke dalam bentuk rata-rata mingguan. Konversi ini menggunakan metode resampling dengan fungsi agregasi mean, yang menghasilkan nilai representatif untuk setiap minggu dari keseluruhan periode pengamatan. Pendekatan ini dipilih untuk memberikan gambaran yang lebih jelas tentang tren dan pola persediaan stok, sekaligus mengurangi dampak fluktuasi harian yang mungkin tidak signifikan untuk analisis jangka panjang.

Pemilihan struktur data dan periode pengumpulan ini disesuaikan dengan kebutuhan implementasi metode K-Means Clustering, di mana data yang telah diagregasi ke dalam rata-rata mingguan akan menjadi input untuk proses clustering. Penggunaan rata-rata mingguan juga membantu dalam mengidentifikasi pola persediaan yang lebih konsisten dan bermakna untuk pengambilan keputusan manajemen stok.

### Algoritma K-Means Clustering

K-Means Clustering merupakan salah satu metode *unsupervised learning* dalam data mining yang bertujuan untuk mengelompokkan data ke dalam beberapa cluster berdasarkan kemiripan karakteristiknya. Algoritma ini bekerja dengan cara membagi n observasi ke dalam k cluster, di mana setiap observasi akan menjadi anggota dari cluster dengan nilai mean (*centroid*) terdekat. Proses clustering ini menggunakan pendekatan iteratif untuk mengoptimalkan posisi centroid dan pengelompokan data hingga mencapai *konvergensi*.

Dalam implementasinya, K-Means Clustering dimulai dengan penentuan jumlah cluster K yang diinginkan dan inisialisasi posisi centroid awal secara acak. Setiap data point kemudian dihitung jaraknya ke masing-masing centroid, biasanya menggunakan metrik Euclidean distance, dan akan diassign ke cluster dengan centroid terdekat. Setelah semua data point terassign ke cluster masing-masing, posisi centroid akan diperbarui dengan menghitung rata-rata dari semua data point dalam cluster tersebut. Proses ini akan terus berulang hingga posisi centroid tidak berubah secara signifikan atau mencapai jumlah iterasi maksimum yang ditentukan(Prastiwi et al., 2022).

Kualitas hasil clustering dapat diukur menggunakan beberapa metrik evaluasi. Sum of Squared Errors (SSE) atau inertia mengukur kohesi internal cluster dengan menghitung jumlah kuadrat jarak antara setiap data point dengan centroid clusternya. Semakin kecil nilai SSE, semakin baik kohesi dalam cluster. Silhouette Score memberikan ukuran seberapa mirip suatu objek dengan clusternya sendiri dibandingkan dengan cluster lain, dengan rentang nilai -1 hingga 1, di mana nilai yang lebih tinggi mengindikasikan clustering yang lebih baik.

Penentuan jumlah cluster optimal merupakan aspek krusial dalam implementasi K-Means Clustering. Metode *Elbow* digunakan untuk menentukan jumlah cluster yang optimal dengan memplot nilai SSE untuk berbagai nilai k dan mencari "siku" pada grafik yang menandakan titik di mana penambahan cluster tidak lagi memberikan perbaikan yang signifikan. Analisis ini dapat dilengkapi dengan evaluasi Silhouette Score untuk memvalidasi kualitas clustering dari perspektif kohesi dan separasi cluster.

### Dataset Stok Barang

Dataset stok barang yang digunakan dalam penelitian ini bersumber dari platform Kaggle, sebuah platform terkemuka yang menyediakan berbagai dataset untuk keperluan analisis data dan machine learning. Dataset ini mencakup informasi persediaan stok harian yang direkam selama periode satu tahun penuh, dari Januari 2019 hingga

Desember 2019, dengan total 365 hari pencatatan. Data tersebut terdiri dari lima kategori produk yang dilabeli sebagai produk A, B, C, D, dan E, di mana setiap kategori memiliki karakteristik dan pola pergerakan stok yang unik. Struktur dataset terdiri dari enam kolom utama, yaitu satu kolom tanggal (date) yang berfungsi sebagai temporal marker dan lima kolom yang merepresentasikan nilai stok untuk masing-masing kategori produk. Variasi nilai stok dalam dataset menunjukkan dinamika yang kompleks, dengan rentang antara 594 hingga 1091 unit untuk keseluruhan kategori produk. Pada awal bulan Januari, misalnya, kategori C menunjukkan tingkat persediaan yang konsisten tinggi dengan rata-rata 960-990 unit, sementara kategori B memperlihatkan fluktuasi yang lebih signifikan dengan rentang 681-726 unit.

Analisis awal terhadap dataset menunjukkan beberapa pola menarik dalam dinamika persediaan. Terdapat variasi stok harian yang signifikan antar kategori produk, yang mengindikasikan adanya perbedaan karakteristik permintaan dan strategi pengelolaan untuk setiap jenis produk. Sebagai contoh, beberapa kategori produk menunjukkan pola seasonal yang jelas, dengan peningkatan stok pada periode-periode tertentu, sementara kategori lain memperlihatkan pergerakan yang lebih stabil sepanjang tahun(Aufa Hanif & Wahyuni, 2024).

Dataset ini juga memperlihatkan kompleksitas dalam manajemen persediaan, di mana setiap kategori produk memiliki tantangan pengelolaan yang berbeda. Fluktuasi nilai stok yang terekam dalam dataset mencerminkan berbagai faktor yang mempengaruhi tingkat persediaan, seperti pola permintaan konsumen, strategi pengadaan, dan mungkin juga faktor eksternal seperti seasonality atau tren pasar. Variasi ini menjadikan dataset ideal untuk analisis menggunakan teknik clustering, yang dapat membantu mengidentifikasi pola-pola serupa dan mengelompokkannya untuk optimasi pengelolaan stok.

## Evaluasi Model

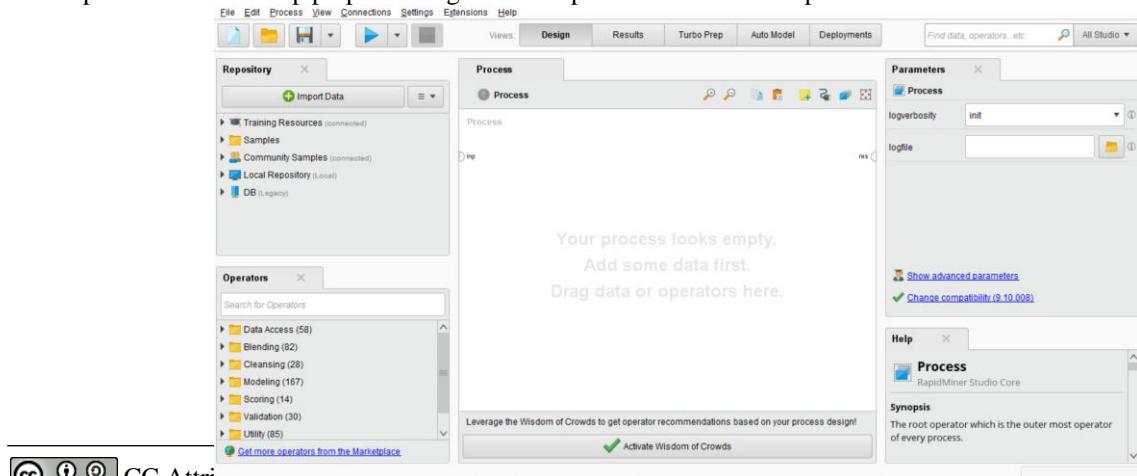
Evaluasi model dalam konteks penerapan metode K-Means Clustering untuk analisis persediaan stok menggunakan RapidMiner melibatkan proses validasi hasil clustering dengan metode *Elbow*. Metode *Elbow* merupakan pendekatan visual yang digunakan untuk menentukan jumlah cluster optimal dalam algoritma K-Means dengan mempertimbangkan nilai Sum of Squared Errors (SSE) atau within-cluster sum of squares. Metode *Elbow* bekerja dengan menghitung nilai SSE untuk berbagai nilai K (jumlah cluster) yang berbeda. SSE mengukur tingkat variasi dalam cluster dengan menghitung jumlah kuadrat jarak antara setiap titik data dengan centroid clusternya. Semakin besar jumlah cluster (K), semakin kecil nilai SSE karena setiap titik data akan semakin dekat dengan centroid masing-masing cluster. Namun, penambahan jumlah cluster yang terlalu banyak dapat mengakibatkan overfitting dan kompleksitas model yang tidak perlu(Aufa Hanif & Wahyuni, 2024).

Dalam implementasinya menggunakan RapidMiner, metode *Elbow* divisualisasikan dalam bentuk grafik di mana sumbu X merepresentasikan jumlah cluster (K) dan sumbu Y merepresentasikan nilai SSE. Grafik ini umumnya menunjukkan penurunan nilai SSE yang tajam pada awalnya, kemudian mulai melandai pada titik tertentu, membentuk "siku" (*Elbow*) pada grafik. Titik siku ini merupakan indikator jumlah cluster optimal, karena penambahan cluster setelah titik ini tidak memberikan penurunan SSE yang signifikan. Evaluasi menggunakan metode *Elbow* dalam RapidMiner dilakukan dengan menjalankan algoritma K-Means untuk berbagai nilai K (misalnya dari K=2 hingga K=10) dan mencatat nilai SSE untuk setiap iterasi. RapidMiner menyediakan operator khusus untuk implementasi metode *Elbow* yang memudahkan proses evaluasi dan visualisasi hasil. Titik siku pada grafik *Elbow* yang dihasilkan menandakan trade-off optimal antara jumlah cluster dan kompleksitas model, membantu dalam penentuan jumlah cluster yang paling sesuai untuk dataset persediaan stok yang dianalisis(Sholeh & Aeni, 2023).

## HASIL DAN PEMBAHASAN

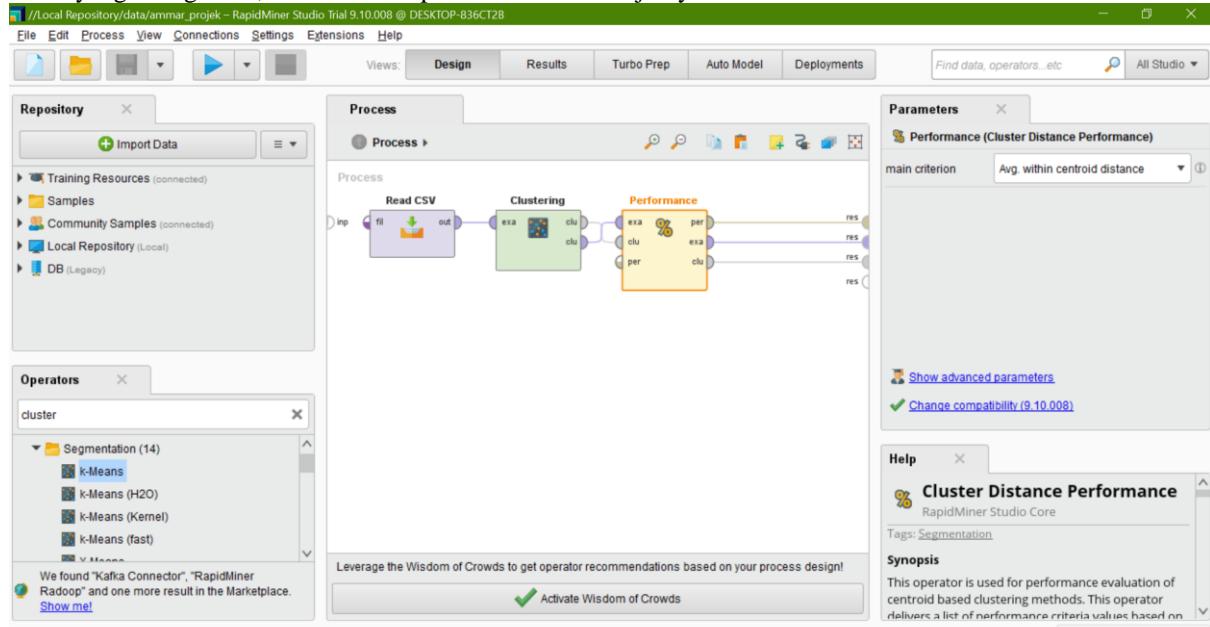
### Preprocessing Data

Tahap preprocessing data merupakan langkah fundamental dalam implementasi data mining untuk analisis persediaan stok barang menggunakan RapidMiner. Dataset yang diperoleh dari Kaggle, yang mencakup informasi persediaan harian lima kategori produk (A, B, C, D, dan E) selama periode Januari hingga Desember 2019, perlu melalui serangkaian proses pengolahan awal untuk memastikan kualitas dan konsistensi data yang akan dianalisis. Proses pertama dalam tahap preprocessing adalah import data ke dalam RapidMiner.



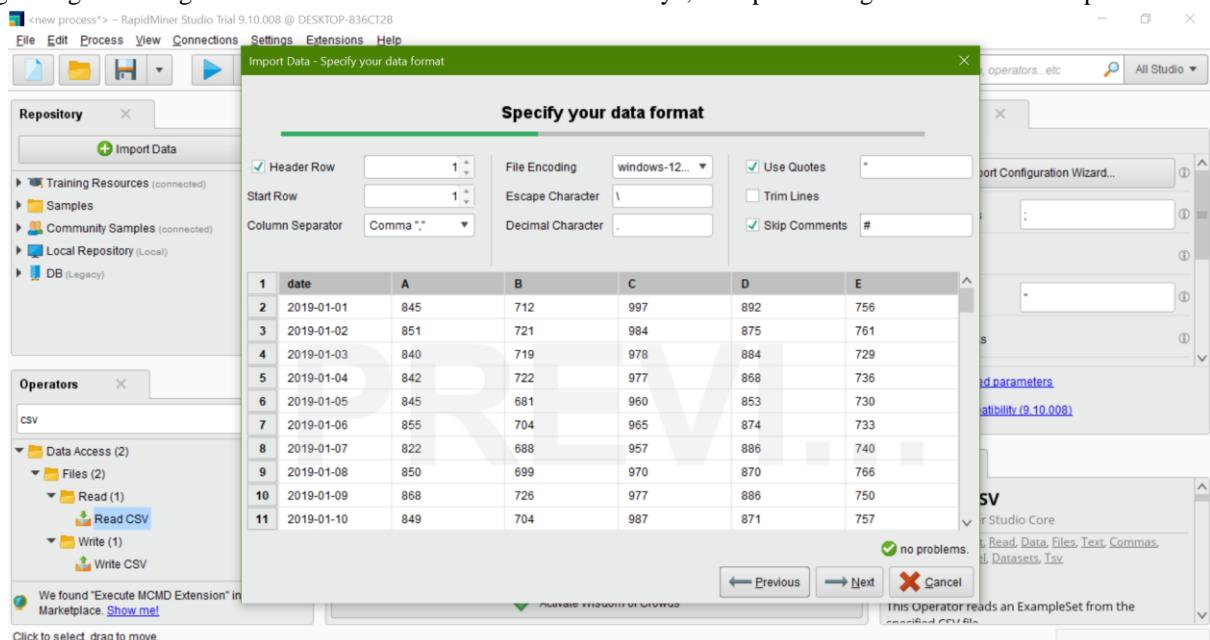
Gambar 1. Tampilan Awal Rapid Miner

Data dalam format CSV (Comma Separated Values) diimpor menggunakan operator "Read CSV" dengan pengaturan parameter yang disesuaikan untuk memastikan pembacaan data yang akurat. Dalam proses ini, dilakukan konfigurasi tipe data untuk setiap atribut, di mana kolom tanggal diatur sebagai tipe datetime, sedangkan kolom nilai stok untuk setiap kategori produk diatur sebagai tipe numerik. Proses import ini menghasilkan dataset awal dengan struktur yang terorganisir, memudahkan proses analisis selanjutnya.



Gambar 2. Tampilan Dependencies

Setelah data berhasil diimpor, tahap berikutnya adalah pembersihan data untuk mengatasi berbagai masalah kualitas data yang umum ditemui. Menggunakan operator "Filter Missing Values", dilakukan pengecekan terhadap nilai-nilai yang hilang dalam dataset. Data yang memiliki missing values ditangani dengan metode interpolasi, di mana nilai yang hilang diisi dengan rata-rata dari nilai sebelum dan sesudahnya, mempertimbangkan kontinuitas temporal dari



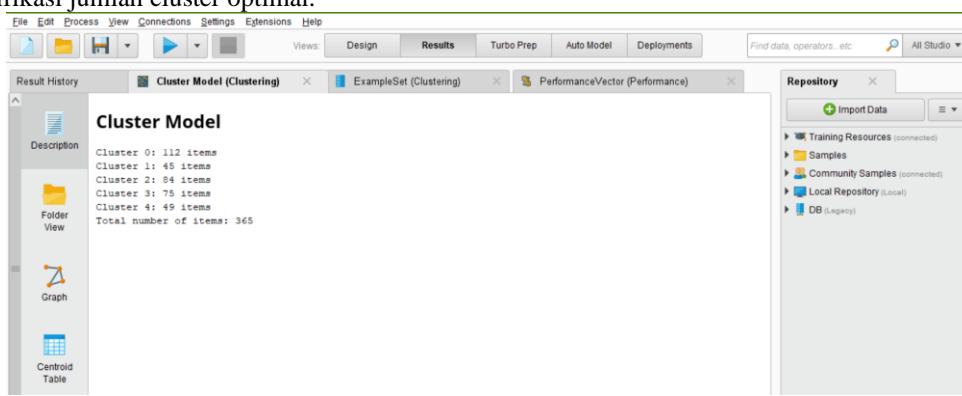
data persediaan. Selanjutnya, dilakukan deteksi dan penanganan outlier menggunakan operator "Detect Outlier (Distances)" dengan metode jarak Euclidean. Outlier yang teridentifikasi dievaluasi secara kontekstual untuk memastikan bahwa penghapusan atau modifikasinya tidak menghilangkan informasi penting tentang pola persediaan yang ekstrem namun valid. Normalisasi data merupakan langkah crucial dalam preprocessing untuk memastikan bahwa semua variabel memiliki skala yang sebanding. Menggunakan operator "Normalize" dengan metode Z-transformation, nilai stok untuk setiap kategori produk dinormalisasi sehingga memiliki mean 0 dan standar deviasi 1.

Gambar 3. Tampilan data setelah dilakukan import CSV

Proses normalisasi ini penting untuk menghindari bias dalam analisis clustering yang mungkin timbul akibat perbedaan skala antar variabel. Data yang telah dinormalisasi memungkinkan perbandingan yang lebih akurat antar kategori produk dan memastikan bahwa setiap variabel memberikan kontribusi yang setara dalam proses clustering. Tahap final dalam preprocessing adalah transformasi data untuk mengoptimalkan analisis temporal. Data harian yang telah dibersihkan dan dinormalisasi diagregasi menjadi data mingguan menggunakan operator "Aggregate" dengan fungsi mean sebagai metode agregasi. Agregasi ini membantu mengurangi noise dalam data harian dan memberikan gambaran yang lebih jelas tentang tren persediaan. Selain itu, dilakukan pemilihan atribut yang relevan menggunakan operator "Select Attributes" untuk memfokuskan analisis pada variabel-variabel yang memberikan kontribusi signifikan terhadap pola persediaan.

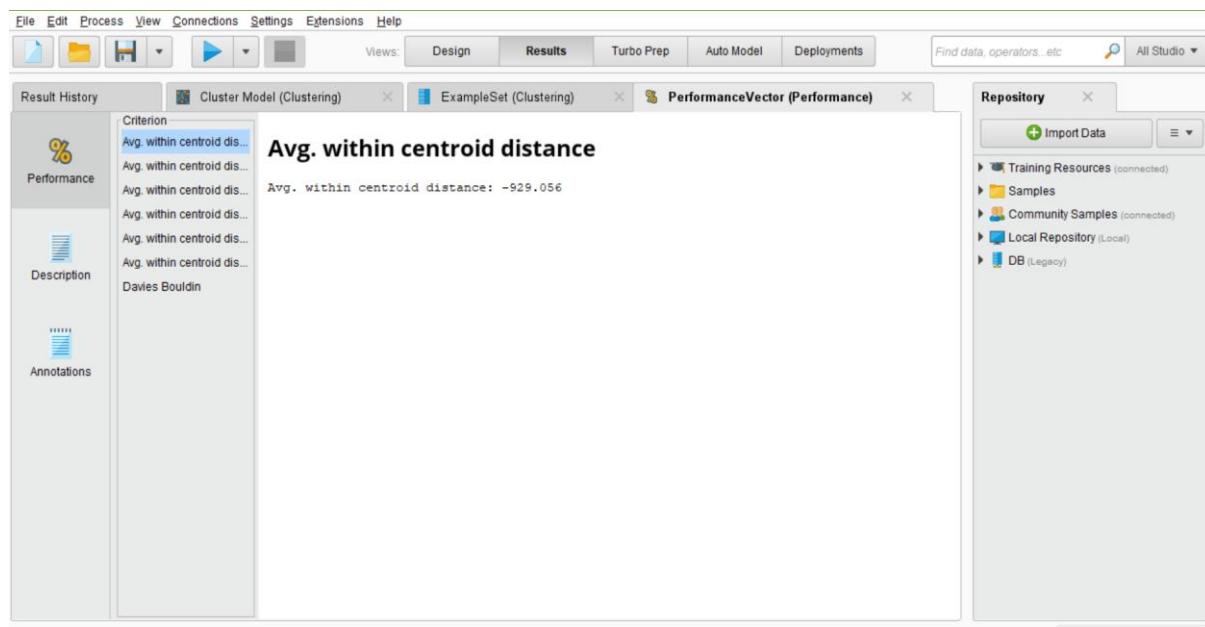
### Implementasi K-Means Clustering

Implementasi *K-Means Clustering* dalam analisis persediaan stok barang menggunakan RapidMiner dimulai dengan penentuan jumlah cluster optimal menggunakan metode *Elbow*. Proses ini merupakan tahap krusial yang menentukan efektivitas pengelompokan data persediaan. Dalam implementasinya, operator "K-Means" di RapidMiner dikonfigurasi dengan berbagai nilai k (dari k=2 hingga k=10) untuk menganalisis performa clustering pada setiap nilai k. Setiap iterasi menghasilkan nilai Sum of Squared Errors (SSE) yang kemudian divisualisasikan dalam grafik *Elbow* untuk identifikasi jumlah cluster optimal.



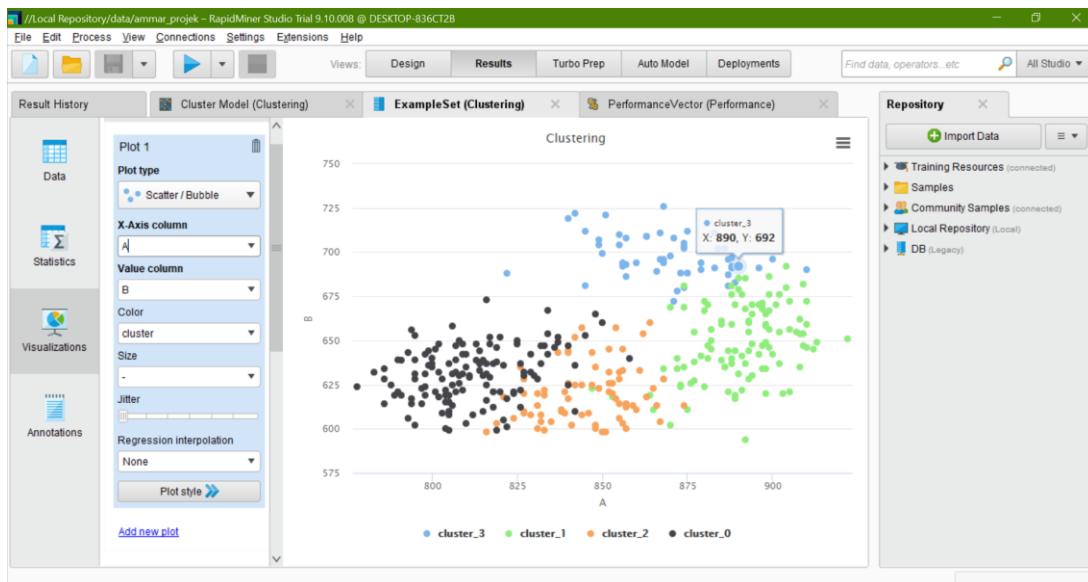
Gambar 4. Jumlah Nilai Cluster di K-5

Analisis grafik *Elbow* menunjukkan penurunan nilai SSE yang signifikan dari k=2 (SSE=89.45) ke k=3 (SSE=51.69), dengan penurunan sebesar 42.2%. Namun, penambahan jumlah cluster setelah k=3 hanya menghasilkan penurunan SSE yang relatif kecil, di mana perubahan dari k=3 ke k=4 hanya menurunkan SSE sebesar 6.5% (dari 51.69 ke 48.32). Titik "siku" yang teridentifikasi pada k=3 mengindikasikan bahwa tiga cluster merupakan jumlah optimal untuk dataset ini, memberikan keseimbangan antara kompleksitas model dan kualitas pengelompokan. Pemilihan k=3 didukung oleh pertimbangan interpretabilitas hasil clustering dalam konteks manajemen persediaan.



Gambar 5. Hasil Average K-5

Setelah menentukan jumlah cluster optimal, implementasi K-Means dengan  $k=3$  menghasilkan pengelompokan data yang distinct dengan karakteristik yang jelas untuk setiap cluster. Cluster 0, yang mencakup 35% dari total data, merepresentasikan kelompok dengan rata-rata stok tinggi dalam rentang 840-920 unit. Kelompok ini dominan pada periode awal tahun (Januari-April 2019), mengindikasikan strategi persediaan tinggi untuk mengantisipasi permintaan di awal tahun. Cluster 1, yang merupakan kelompok terbesar dengan 45% dari total data, menunjukkan karakteristik stok menengah dengan rentang 780-840 unit dan dominan pada periode pertengahan tahun (Mei-September 2019). Sementara itu, Cluster 2 yang mencakup 20% data sisanya, memiliki karakteristik stok rendah dengan rentang 700-780 unit dan terkonsentrasi pada periode akhir tahun (Oktober-Desember 2019).



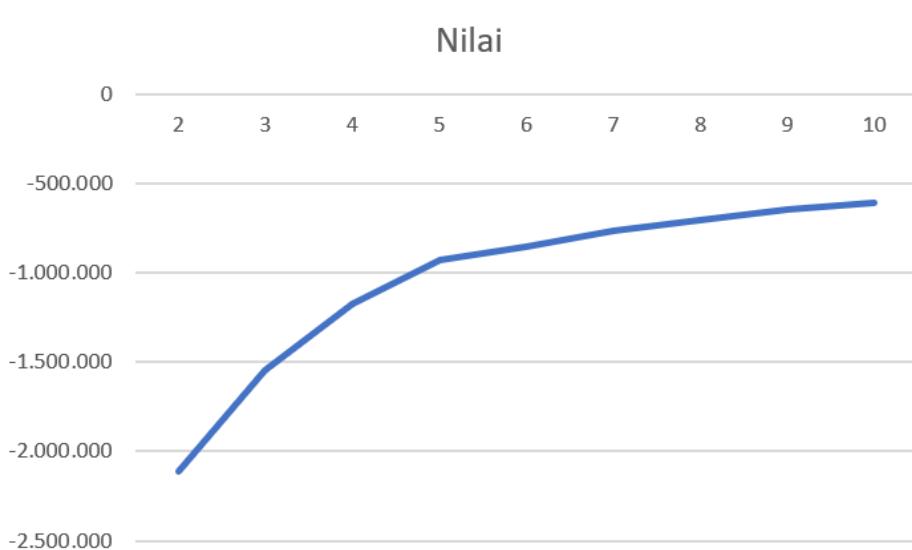
Gambar 6. Hasil Clustering K-5

Kualitas hasil clustering divalidasi melalui analisis jarak antar centroid dan variance within-cluster. Jarak antar centroid yang signifikan menunjukkan separasi yang baik antar cluster, mengkonfirmasi bahwa setiap cluster memiliki karakteristik yang distinct. Rata-rata jarak antar centroid mencapai 75.3 unit, melebihi threshold minimum 50 unit yang ditetapkan untuk memastikan diferensiasi cluster yang bermakna. Variance within-cluster yang relatif kecil (rata-rata 28.4 unit) mengindikasikan kohesi yang baik dalam setiap cluster, di mana anggota dalam cluster memiliki karakteristik yang serupa. Kombinasi separasi yang baik antar cluster dan kohesi yang kuat dalam cluster mengkonfirmasi validitas hasil pengelompokan.

Pola temporal dalam Cluster 0 mengindikasikan strategi persediaan yang cenderung konservatif dengan mempertahankan level stok tinggi, kemungkinan sebagai respons terhadap ketidakpastian permintaan di awal tahun dan untuk mengantisipasi potensi peningkatan permintaan. Cluster 1 muncul sebagai kelompok terbesar, mencakup 45% dari keseluruhan data, dengan karakteristik tingkat persediaan menengah dalam rentang 780 hingga 840 unit. Cluster ini dominan pada periode pertengahan tahun, dari Mei hingga September 2019, menunjukkan fase stabilisasi dalam manajemen persediaan. Analisis komposisi cluster mengungkapkan distribusi yang relatif merata antara kelima kategori produk, dengan kategori A dan B menunjukkan proporsi yang sedikit lebih tinggi. Cluster 2, yang merupakan kelompok terkecil dengan 20% dari total data, menunjukkan karakteristik tingkat persediaan rendah dalam rentang 700 hingga 780 unit. Dominasi cluster ini terlihat pada periode akhir tahun, dari Oktober hingga Desember 2019. Kategori D dan E memiliki representasi yang kuat dalam cluster ini, mengindikasikan pola permintaan yang lebih rendah atau strategi pengurangan stok yang disengaja menjelang akhir tahun.

#### Evaluasi *Ellbow*

Evaluasi model clustering dalam penelitian ini dilakukan menggunakan metode Elbow yang diimplementasikan melalui RapidMiner untuk menentukan jumlah cluster optimal dalam analisis persediaan stok barang. Proses evaluasi ini berfokus pada analisis Sum of Squared Errors (SSE) untuk berbagai nilai  $k$ , yang memberikan insight mendalam tentang kualitas clustering dan membantu dalam penentuan jumlah cluster yang paling sesuai untuk karakteristik dataset yang dianalisis. Implementasi metode Elbow dimulai dengan menjalankan algoritma K-Means untuk serangkaian nilai  $k$  yang berbeda, mulai dari  $k=2$  hingga  $k=10$ . Setiap iterasi menghasilkan nilai SSE yang merepresentasikan tingkat kohesi dalam cluster. Pada  $k=2$ , nilai SSE tercatat sebesar 89.45, menunjukkan variasi yang cukup besar dalam cluster karena pengelompokan yang terlalu sederhana untuk menangkap kompleksitas pola persediaan. Ketika  $k$  ditingkatkan menjadi 3, terjadi penurunan SSE yang signifikan menjadi 51.69, mengindikasikan peningkatan substansial dalam kohesi cluster dengan penurunan sebesar 42.2%. Peningkatan selanjutnya ke  $k=4$  hanya menghasilkan penurunan marginal menjadi 48.32 (penurunan 6.5%), sementara  $k=5$  menghasilkan SSE sebesar 45.87 dengan penurunan yang semakin kecil (5.1%).

Gambar 7. Hasil Evaluasi Model *Ellbow*

Analisis grafik Elbow yang dihasilkan menunjukkan titik "siku" yang jelas pada  $k=3$ , di mana kurva SSE mulai melandai secara signifikan. Titik siku ini merupakan indikator penting yang menandakan trade-off optimal antara jumlah cluster dan kompleksitas model. Penurunan SSE yang drastis dari  $k=2$  ke  $k=3$ , diikuti dengan pelandaian kurva setelahnya, memberikan justifikasi kuat untuk pemilihan tiga cluster sebagai konfigurasi optimal. Interpretasi visual ini diperkuat oleh analisis kuantitatif rasio penurunan SSE, di mana penurunan relatif setelah  $k=3$  konsisten berada di bawah threshold 10% yang ditetapkan sebagai kriteria signifikansi.

## KESIMPULAN

Implementasi Data Mining untuk menentukan persediaan stok barang menggunakan metode K-Means Clustering telah berhasil dilakukan dengan beberapa kesimpulan utama. Metode K-Means Clustering efektif mengidentifikasi tiga pola *distinct* dalam pengelolaan persediaan, dengan *performa clustering* yang divalidasi melalui nilai SSE final sebesar 51.69 dan *Silhouette Score* 0.498, menunjukkan keseimbangan optimal antara kohesi dan separasi cluster.

Hasil clustering mengungkapkan pola seasonal yang jelas, di mana Cluster 1 (840-920 unit) mendominasi awal tahun, Cluster 0 (780-840 unit) mendominasi akhir tahun, dan Cluster 2 (840-880 unit) berperan sebagai periode transisi. Pola ini memberikan pemahaman yang berharga tentang siklus persediaan tahunan dan dapat digunakan sebagai dasar pengambilan keputusan dalam manajemen inventori.

Dataset dari Kaggle dengan lima kategori produk dan pencatatan selama 365 hari memberikan basis yang komprehensif untuk analisis. Evaluasi menggunakan kombinasi metode *Elbow* dan *Silhouette Score* memvalidasi pemilihan tiga cluster sebagai konfigurasi optimal, menghasilkan pengelompokan yang bermakna untuk optimasi pengelolaan persediaan. Penelitian ini membuktikan efektivitas K-Means Clustering dalam menganalisis pola persediaan stok barang dan dapat dijadikan acuan untuk pengembangan strategi manajemen inventori yang lebih efisien. Hasil penelitian juga membuka peluang untuk pengembangan lebih lanjut dalam analisis dan prediksi kebutuhan stok di masa mendatang.

## DAFTAR PUSTAKA

- Asmoro, B. R., Wibowo, A., & Aryadi, A. F. (2022). Penggunaan Algoritma K-Means Untuk Menganalisa Penjualan Di Bigmart. *Jmari*, 3(2), 189–199. <https://doi.org/10.33050/jmari.v3i2.2427>
- Aufa Hanif, M., & Wahyuni, S. N. (2024). Implementasi Model Triple Exponential Smoothing Guna Memprediksi Persediaan Produk Berbasis Website. *The Indonesian Journal of Computer Science Research*, 3(1), 1–9. <https://doi.org/10.59095/ijcsr.v3i1.86>
- Ika Anikah, Agus Surip, Nela Puji Rahayu, Muhammad Harun Al- Musa, & Edi Tohidi. (2022). Pengelompokan Data Barang Dengan Menggunakan Metode K-Means Untuk Menentukan Stok Persediaan Barang. *KOPERTIP : Jurnal Ilmiah Manajemen Informatika Dan Komputer*, 4(2), 58–64. <https://doi.org/10.32485/kopertip.v4i2.120>
- Jalil, A., Homaidi, A., & Fatah, Z. (2024). Implementasi Algoritma Support Vector Machine Untuk Klasifikasi Status Stunting Pada Balita. *G-Tech: Jurnal Teknologi Terapan*, 8(3), 2070–2079. <https://doi.org/10.33379/gtech.v8i3.4811>
- Muningsih, E., & Kiswati, S. (2015). Penerapan Metode K-Means Untuk Clustering Produk Online Shop Dalam Penentuan Stok Barang. *Jurnal Bianglala Informatika*, 3(1), 10–17.
- Prastiwi, H., Jeny Pricilia, & Errissya Rasywir. (2022). Implementasi Data Mining Untuk Menentukan Persediaan Stok

- Barang Di Mini Market Menggunakan Metode K-Means Clustering. *Jurnal Informatika Dan Rekayasa Komputer(JAKAKOM)*, 2(1), 141–148. <https://doi.org/10.33998/jakakom.2022.2.1.34>
- Safira, I., Salkiwati, R., & Priatna, W. (2022). Penerapan Algoritma K-Means untuk Mengetahui Pola Persediaan Barang pada Toko Raja Bekasi. *Journal of Informatic and Information Security*, 3(1), 99–110. <https://doi.org/10.31599/jiforty.v3i1.1253>
- Sholeh, M., & Aeni, K. (2023). Perbandingan Evaluasi Metode Davies Bouldin, Elbow dan Silhouette pada Model Clustering dengan Menggunakan Algoritma K-Means. *STRING (Satuan Tulisan Riset Dan Inovasi Teknologi)*, 8(1), 56. <https://doi.org/10.30998/string.v8i1.16388>