JITET (Jurnal Informatika dan Teknik Elektro Terapan)
Vol. 13 No. 2, pISSN: 2303-0577 eISSN: 2830-7062 http://dx.doi.org/10.23960/jitet.v13i2.6438

PEMANFAATAN ALGORITMA K-MEANS DALAM ANALISIS DATA PENJUALAN TOKO BUYUNG UPIK JS DI LAZADA

Devita Fitri Angraeni[1*], Nining Rahaningsih[2], Raditya Danar Dana[3], Cep Lukman Rohmat[4]

[1,2,3,4]STMIK IKMI Cirebon; Jl. Perjuangan No. 10B, Majasem, Cirebon, Jawa Barat 45135, Telp. (0231) 490480

Received: 9 Maret 2025
Accepted: 29 Maret 2025
Published: 14 April 2025

Keywords:
Clustering, Data Mining, Davies Bouldin Indeks, K-Means

Corespondent Email:
devitafitriaa@gmail.com

Abstrak. Banyaknya produk yang dijual oleh Toko Buyung Upik JS di Lazada menimbulkan kesulitan dalam menentukan produk yang laku dan kurang laku, sehingga terjadi ketidakseimbangan stok, seperti kelebihan pada produk yang kurang diminati dan kekurangan pada produk yang populer. Penelitian ini bertujuan mengelompokkan produk berdasarkan pola penjualan menggunakan teknik data mining untuk membantu strategi penjualan dan pengelolaan stok yang lebih efektif. Algoritma *K-Means* digunakan untuk *clustering* data penjualan, mencakup jumlah stok, transaksi, dan harga. Proses data mining meliputi tahapan *Selection*, *Preprocessing*, *Transformation*, *Data Mining*, dan *Interpretation/Evaluation*. Penentuan jumlah *cluster* optimal dilakukan dengan *Elbow Method*, sedangkan kualitas *clustering* dievaluasi menggunakan *Davies Bouldin Index (DBI)*. Hasil penelitian menunjukkan jumlah *cluster* optimal adalah empat*:* Cluster 0 (83 produk, penjualan stabil), *Cluster 1* (121 produk, penjualan tinggi), *Cluster 2* (14 produk, kurang diminati), dan *Cluster 3* (38 produk, penjualan moderat). Nilai rata-rata jarak dalam *cluster* adalah 54.941.560,812, dengan *DBI* sebesar 0,386 yang menunjukkan kualitas *clustering* cukup baik. Hasil ini memberikan wawasan bagi toko untuk memprioritaskan pengelolaan stok dan mengoptimalkan penjualan.

Abstract. *The large number of products sold by Toko Buyung Upik JS on Lazada creates challenges in identifying bestselling and less popular products, leading to stock imbalances, such as overstocking less popular items and understocking popular ones. This study aims to cluster products based on sales patterns using data mining techniques to support more effective sales strategies and inventory management. The K-Means algorithm was employed to cluster sales data, including stock levels, sales transactions, and prices. The data mining process involved the stages of Selection, Preprocessing, Transformation, Data Mining, and Interpretation/Evaluation. The optimal number of clusters was determined using the Elbow Method, and clustering quality was assessed using the Davies Bouldin Index (DBI). The results identified four optimal clusters: Cluster 0 (83 products, stable sales), Cluster 1 (121 products, high sales), Cluster 2 (14 products, less popular), and Cluster 3 (38 products, moderate sales). The average within-cluster distance was 54,941,560.812, with a DBI of 0.386, indicating good clustering quality. These findings provide insights for the store to prioritize inventory management and optimize sales by focusing on more popular products and reducing the buildup of less popular items.*

1. PENDAHULUAN

Perkembangan teknologi informasi dan komunikasi telah mendorong pertumbuhan *e-commerce* secara signifikan. Platform seperti Lazada memberikan kemudahan bagi penjual untuk menjangkau pelanggan secara luas dan meningkatkan penjualan. Namun, semakin banyaknya produk yang dijual sering kali menyebabkan permasalahan seperti ketidakseimbangan stok akibat kesulitan dalam mengidentifikasi produk yang laku dan tidak laku di pasaran. Hal ini dapat mengakibatkan kelebihan stok untuk produk yang kurang diminati dan kekurangan stok pada produk populer. Toko Buyung Upik JS di Lazada menghadapi tantangan serupa dalam menentukan pola penjualan produk mereka. Pengelolaan stok yang tidak efektif dapat menurunkan efisiensi operasional dan profitabilitas toko. Oleh karena itu, diperlukan analisis data yang mendalam untuk memahami pola penjualan guna menyusun strategi promosi yang lebih efektif dan manajemen stok yang optimal. Salah satu pendekatan yang dapat digunakan adalah teknik *data mining* dengan algoritma *K-Means Clustering*, yang memungkinkan pengelompokan produk berdasarkan pola penjualan. Data mining adalah proses mengekstraksi informasi, pengetahuan, dan pola dari sejumlah besar data[1]. *Data mining*, yang sering disebut sebagai *knowledge discovery in database (KDD)*, adalah proses yang mencakup pengumpulan dan penggunaan data historis untuk mengidentifikasi pola, hubungan, atau keteraturan dalam data berukuran besar[2]. *Clustering* merupakan metode statistik yang diterapkan untuk mengeklompokan sejumlah data atau objek besar ke dalam klaster berdasarkan karakteristik yang dimiliki oleh data atau objek tersebut[3]. Pentingnya *K-Means* dalam mengidentifikasi pola transaksi di toko pakaian, sehingga membantu toko tersebut dalam merencanakan strategi pemasaran yang lebih efektif berdasarkan analisis perilaku pelanggan[4].

Penelitian sebelumnya menunjukkan bahwa algoritma *K-Means* efektif untuk mengelompokkan produk berdasarkan kinerja penjualan, sehingga membantu dalam pengelolaan stok dan strategi pemasaran[5]. Sementara itu, studi membuktikan bahwa *K-Means* dapat mengidentifikasi kategori produk populer dan kurang diminati dalam toko *fashion* hijab[6]. Penelitian lainnya juga menekankan pentingnya penerapan *K-Means* untuk menganalisis penjualan produk digital, di mana hasil segmentasi membantu toko dalam menyusun promosi yang lebih tepat sasaran[7]. Lalu pada penelitian selanjutnya berhasil mengkaji penerapan *K-Means* untuk mengidentifikasi produk terlaris di PT. Titian Nusantara Boga, dan menemukan bahwa segmentasi yang dihasilkan oleh algoritma ini membantu perusahaan dalam merumuskan strategi pemasaran yang lebih terarah dan efisien[8].

Dalam konteks penelitian ini, algoritma *K-Means* digunakan untuk mengelompokkan data penjualan Toko Buyung Upik JS berdasarkan atribut seperti jumlah stok, transaksi penjualan, dan harga produk. Penelitian lain juga berhasil mengimplementasikan *K-Means* untuk mengelompokkan produk di toko online menggunakan data penjualan dari platform *e-commerce*, yang memberikan wawasan tentang produk yang paling diminati dan kurang diminati oleh konsumen[9]. Untuk memastikan kualitas *clustering*, digunakan *Elbow method* untuk menentukan jumlah optimal *cluster* dan *Davies Bouldin Index* untuk mengevaluasi hasil *clustering*. Implementasi analisis dilakukan menggunakan perangkat lunak RapidMiner, yang dikenal mampu menyederhanakan proses *data mining*.

Melalui penelitian ini, diharapkan Toko Buyung Upik JS dapat memperoleh wawasan mendalam terkait pola penjualan dan karakteristik produk, sehingga dapat meningkatkan efektivitas strategi promosi dan pengelolaan stok.

2. TINJAUAN PUSTAKA

2.1. Data Mining

Data Mining merupakan proses di mana korelasi-korelasi dan pola-pola yang baru dan bermanfaat ditemukan melalui eksplorasi dalam repositori data yang besar. Teknologi pengenalan pola seperti teknik statistik dan matematika digunakan pada proses ini[3].

2.2. *K-Means*

K-Means adalah salah satu algoritma clustering yang paling populer dan sederhana. Algoritma ini bekerja dengan cara mengelompokkan data ke dalam jumlah klaster yang ditentukan sebelumnya. Proses K-Means

diawali dengan pemilihan jumlah klaster yang diinginkan(k) dan pemilihan titik awal yang digunakan sebagai centroid dari masing-masing klaster. Kemudian, setiap data akan diklasifikasikan ke dalam klaster yang centroidnya paling dekat dengan data tersebut. Setelah semua data diklasifikasikan, centroid dari setiap klaster akan diperbarui berdasarkan rata-rata dari semua data yang terklasifikasi ke dalam klaster tersebut. Proses ini akan diulangi beberapa kali hingga tidak ada perubahan lagi pada klasifikasi data[7].

2.3. *David Bouldin Index (DBI)*

Davies-Bouldin Index (DBI) adalah metrik yang digunakan untuk mengevaluasi kualitas hasil clustering dalam analisis data. DBI mengukur seberapa baik pemisahan antara cluster dan seberapa kompak data dalam setiap cluster. Nilai *DBI* yang lebih kecil menunjukkan clustering yang lebih baik, karena menunjukkan bahwa cluster lebih terpisah dan lebih homogen[10].

3. METODE PENELITIAN

Penelitian ini dilakukan untuk menganalisis data penjualan Toko Buyung Upik JS di Lazada. Dimulai dengan mengidentifikasi masalah atau persoalan yang belum dirumuskan solusinya, lalu dilakukan pengumpulan data untuk mendapatkan data yang diperlukan dalam penelitian ini, data tersebut berupa jumlah stok produk, stok akhir, data transaksi penjualan serta beberapa informasi data yang berhubungan dengan harga. Kemudian dilakukan analisa data supaya data tersebut dibersihkan untuk memastikan keakuratan data termasuk penghapusan data duplikat atau data yang tidak relevan. Setelah dilakukan proses analisa data kemudian dilakukan pengolahan data dengan *K-means* untuk mengelompokkan data penjualan dari Toko Buyung Upik JS.

Proses dimulai dengan menyiapkan data yang relevan, seperti jumlah penjualan dan frekuensi pembelian. Selanjutnya, data akan dinormalisasi agar semua fitur memiliki skala yang sama, mencegah bias dalam analisis. Setelah itu, jumlah *cluster* yang tepat akan ditentukan menggunakan metode *Elbow*, yang membantu menemukan jumlah *cluster* optimal. Lalu, melakukan pengujian data dengan menggunakan RapidMiner supaya hasil data tersebut bisa divisualisasikan dan dianalisis setiap segmen.

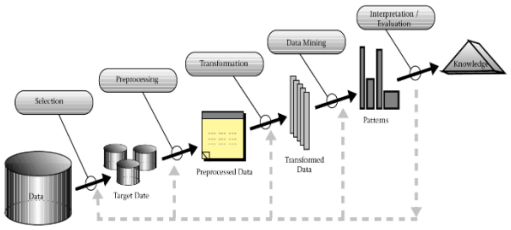
3.1. Metode Pengumpulan Data

Metode pengumpulan data dalam penelitian ini dilakukan dengan menggunakan pendekatan dokumentasi, yaitu dengan mengumpulkan data historis yang tersedia di *platform e-commerce* Lazada melalui sistem administrasi penjualan Toko Buyung Upik JS. Data yang digunakan dalam penelitian ini meliputi tiga atribut utama, yaitu jumlah stok produk, transaksi penjualan, dan harga produk. Atribut-atribut ini dipilih karena merupakan faktor penting yang mempengaruhi analisis pola penjualan dan pengelolaan stok produk.

Jumlah stok produk merujuk pada data yang menunjukkan jumlah unit barang yang tersedia di gudang atau dalam proses pengiriman pada setiap produk yang dijual di Lazada. Transaksi penjualan menunjukkan banyaknya unit produk yang terjual dalam periode tertentu, sedangkan harga produk mencatat harga jual setiap produk yang terdaftar di platform. Data tersebut diperoleh dengan cara mengunduh laporan transaksi penjualan yang tersedia dalam sistem administrasi Lazada yang disediakan untuk penjual. Data yang diunduh dalam *format spreadsheet (Excel)* kemudian diproses lebih lanjut untuk dianalisis.

Pengumpulan data ini dilakukan dalam periode waktu tertentu yang dianggap *representatif*, dengan tujuan untuk mencakup fluktuasi tren penjualan yang relevan. Periode pengumpulan data yang digunakan adalah selama 6 bulan terakhir, yang mencakup data harian, mingguan, atau bulanan, tergantung pada format yang disediakan oleh platform Lazada. Setelah data dikumpulkan, dilakukan verifikasi untuk memastikan kelengkapan dan kualitas data yang akan digunakan dalam analisis.

3.2. Tahapan Penelitian



Gambar 3. 1 Tahapan Penelitian

Teknik yang dilakukan dalam penelitian ini adalah dengan proses *Knowledge Discovery in Databases (KDD)*. *KDD (Knowledge Discovery in Databases)* adalah suatu proses yang sistematis untuk menemukan informasi berharga atau pola-pola menarik dari kumpulan data yang besar. Knowledge Discovery in Database (KDD) merupakan pendekatan sistematis yang dirancang untuk mengidentifikasi pola atau informasi penting dari data mentah. Metode ini melibatkan serangkaian tahapan yang saling berkesinambungan, mulai dari pengumpulan data, proses transformasi, analisis mendalam, hingga interpretasi hasil[11]. Proses ini melibatkan beberapa tahap, mulai dari pembersihan data, integrasi data, hingga penggunaan algoritma data mining untuk mengungkap pola tersembunyi. Tujuan utama *KDD* adalah untuk mengubah data mentah menjadi pengetahuan yang berguna bagi pengambilan keputusan. Proses dalam *KDD* terdapat 5 tahapan yaitu seleksi data dari data sumber ke data target, tahap *pre-processing*, transformasi, *data mining* dan tahap evaluasi[12].

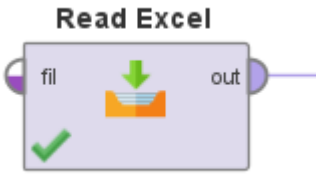
4. HASIL DAN PEMBAHASAN

4.1. Data

Datasheet yang menjadi dasar penelitian ini berasal dari data yang dikumpulkan dari Toko Buyung Upik JS, dengan periode pengambilan data mulai Januari 2023 hingga September 2024. Dataset ini terdiri dari 256 entri dan mencakup 8 atribut, yaitu: nomor, kategori, nama barang, jumlah produk terjual, harga per unit, stok awal, stok akhir, serta total penjualan. Data tersebut diperoleh dalam bentuk dokumen *Microsoft Excel* dengan format .xlsx.

4.2. Data Selection

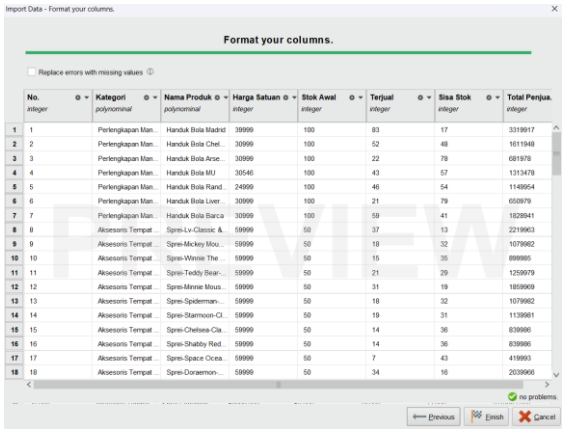
Pada tahap penelitian ini, data yang digunakan berasal dari Toko Buyung Upik JS dengan periode penjualan dari Januari 2023 hingga September 2024. Dataset ini memiliki 256 entri dan terdiri dari 8 atribut, antara lain nomor urut, kategori produk, nama produk, harga per unit, stok awal, jumlah produk terjual, sisa stok, serta total penjualan. Untuk memilih atribut yang akan digunakan, langkah yang paling tepat adalah menggunakan operator



Gambar 4. 1 Operator *Read Excel*

"Select Attributes" di RapidMiner. Namun, sebelum itu, *operator "Read Excel"* perlu digunakan terlebih dahulu untuk mengimpor dataset yang tersimpan dalam *format Excel* (.xlsx atau .xls) ke dalam proses *data mining*. *Operator "Read Excel"* berfungsi membaca data dari *file Excel* dan memasukkannya ke dalam alur proses *data mining* di RapidMiner.

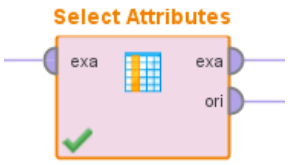
Kemudian, setelah dilakukannya proses *"Read Excel"* menggunakan RapidMiner, hasil dataset yang sudah diimport ditampilkan dalam *format columns* atau bentuk tabel.



Gambar 4. 2 Hasil *Import Data*

Langkah selanjutnya adalah proses seleksi data, yang dilakukan dengan menggunakan operator *"Select Attributes"* untuk memilih atribut-

atribut yang tidak relevan atau tidak dibutuhkan. Dalam dataset penjualan Toko Buyung Upik JS, atribut yang dianggap tidak diperlukan adalah "Nomor urut" dan "Kategori". Atribut "Nomor urut" dianggap sebagai gangguan atau *noise* dalam data, sedangkan atribut "Kategori" dapat membuat model analisis menjadi lebih kompleks dan sulit untuk diinterpretasikan jika terdapat terlalu banyak kategori produk yang unik.



Gambar 4. 4 Operator *Select Attributes*

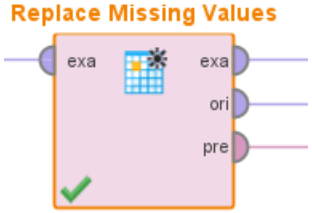
Hasil penggunaan dari *Operator "Select Attributes"* yang semula berjumlah 8 atribut pada dataset Toko Buyung Upik JS hanya 6 atribut yang akan dipilih dalam proses *filtering* ini yaitu Nama Produk, Harga Satuan, Stok Awal, Terjual, Sisa Stok, dan Total Penjualan.

Tabel 4. 1 Hasil Operator *Select Attributes*

Parameter	Isi
Type	*Include Attributes*
Attribute filter type	*A Subset*
Select Attributes	Nama Produk, Harga Satuan, Stok Awal, Terjual, Sisa Stok, Total Penjualan

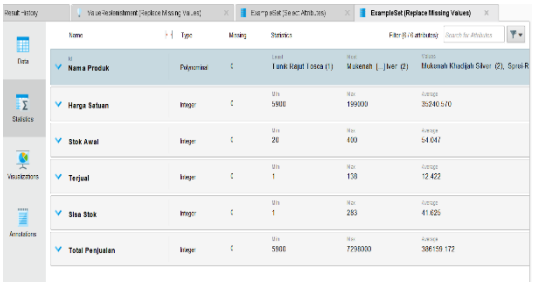
4.3. *Preprocessing* Data

Pada tahap *Preprocessing Data*, *missing value* dalam data perlu dihilangkan untuk menjaga integritas dataset dan memastikan bahwa analisis yang dilakukan tidak terpengaruh oleh *missing value*. *Missing value* dapat mengganggu proses *modeling* data mining selanjutnya, sehingga perlu ditangani dengan tepat[13].



Gambar 4. 6 Operator *Replace Missing Value*

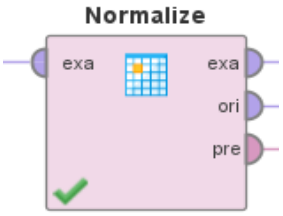
Hasil penggunaan operator *Replace Missing Values*, dari data penjualan Toko Buyung Upik JS tidak ada dataset yang memiliki missing values. Semua dataset yang ditampilkan memiliki nilai lengkap, tidak ada nilai yang hilang atau kosong.



Gambar 4. 3 Hasil Penggunaan Operator

4.4. Data Transformations

Pada proses data transformasi ini, digunakan operator *Normalize* untuk menormalkan atau mengubah skala data sehingga nilainya berada dalam rentang tertentu. *Operator Normalize* pada dataset ini di gunakan untuk menormalkan nilai *DBI* sebelum diclusterkan dengan menggunakan algoritma *K-Means* [14].



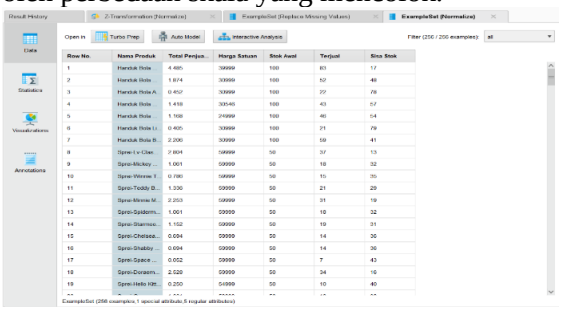
Gambar 4. 5 Operator *Normalize*

Untuk penggunaan *operator Normalize* atribut yang diubah skalanya adalah atribut Total Penjualan, dikarenakan atribut Total Penjualan memiliki nilai numerik yang dapat dinormalisasi.

Tabel 4. 2 Hasil Operator *Normalize*

Parameter	Isi
Attribute filter type	*Subset* :
	Total Penjualan
Method	*Z-transformation*

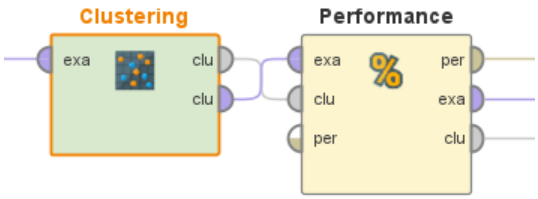
Kemudian, untuk tampilan dari penggunaan operator *Normalize* pada atribut Total Penjualan sudah diubah menjadi nilai yang berada dalam skala lebih kecil (antara -1 dan 1). Tujuan dari transformasi ini adalah agar skala Total Penjualan tidak terlalu besar atau kecil dibandingkan atribut lainnya, sehingga model analisis atau prediksi tidak terpengaruh oleh perbedaan skala yang mencolok.



Gambar 4. 8 Hasil Penggunaan Operator

4.5. Data Mining

Pada tahap *data mining*, teknik *clustering* yang diimplementasikan adalah algoritma *K-Means Clustering* menggunakan operator *Clustering*. *Operator* ini merupakan *operator* utama pemodelan untuk menghasilkan pengklasteran *dataset* [13]**.** *Operator* yang digunakan dalam tahapan tersebut yaitu *operator clustering K-means* dan juga operator *cluster distance performance* untuk dengan metode evaluasi nilai *DBI* [14].

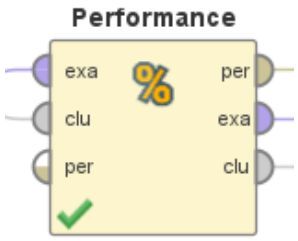


Gambar 4. 9 Operator Clustering

Pada penggunaan *operator Clustering K-Means* terdapat parameter yang harus disesuaikan. Parameter yang akan digunakan ini dapat dilihat pada tabel 4.3.

Tabel 4. 3 Paramater *Clustering K-Means*

Parameter	Value
K	2-9
Measure Types	*Mixed Measures*
Max Optimalization	*100%*



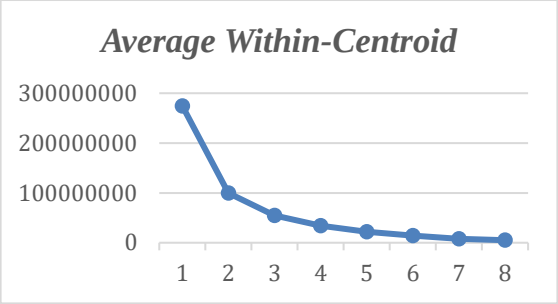
Gambar 4. 7 Operator *Performance*

Kemudian, untuk menambahkan *operator Performance* digunakan untuk mengevaluasi hasil dari algoritma *clustering* berdasarkan jarak antar *cluster* dan distribusi anggota dalam *cluster*. *Operator Performance* untuk mengukur *Cluster Distance Performance* dengan metode *Davies Bouldin Index (DBI)*. *Operator* ini bertujuan untuk mengetahui nilai DBI dari proses *clustering* yang dilakukan [13].

Untuk menentukan jumlah cluster yang optimal, salah satu cara yang dapat digunakan adalah dengan *metode Elbow*. Salah satu metrik yang umum diterapkan dalam metode ini adalah *Average Within-Centroid Distance*, yang mengukur rata-rata jarak antara data dalam sebuah *cluster* dengan *centroid-nya*. Tujuan dari metrik ini adalah untuk membantu menemukan titik optimal jumlah *cluster* dengan mempertimbangkan keseimbangan antara jumlah *cluster* dan kedekatan data, serta memberikan gambaran yang jelas mengenai kepadatan setiap *cluster*.

Tabel 4. 4 Hasil Perhitungan *Average*

K	*Average Within-Centroid*
2	274455085.403
3	100014147.455
4	54941560.812
5	34232346.684
6	22070925.284
7	14377921.802
8	7731256.945
9	5095218.039



Gambar 4. 10 Gravik Metode *Elbow*

Berdasarkan perhitungan Average Within-Centroid Distance di atas, analisis menggunakan metode Elbow menunjukkan bahwa jarak rata-rata antar data dalam cluster menurun seiring dengan peningkatan jumlah cluster (K). Penurunan jarak ini mengindikasikan bahwa data dalam cluster semakin homogen ketika jumlah cluster bertambah. Namun, penurunan ini mulai melambat setelah mencapai titik tertentu, yang dapat dijadikan sebagai jumlah cluster optimal. Pada K=2, nilai Average Within-Centroid Distance tercatat sebesar 274.455.085,403, yang kemudian mengalami penurunan signifikan ke K=3 dengan nilai 100.014.147,455. Penurunan ini sebesar 174.440.937,948, menunjukkan bahwa penambahan satu cluster membuat data terbagi dengan lebih baik ke dalam kelompok yang lebih homogen. Penurunan signifikan lainnya terjadi antara K=3 dan K=4, dengan selisih 45.072.586,643, yang menghasilkan nilai rata-rata jarak sebesar 54.941.560,812. Setelah K=4, laju penurunan mulai melambat. Dari K=4 ke K=5, penurunan hanya sebesar 20.709.214,128, dan antara K=6 ke K=7, jarak rata-rata hanya turun sebesar 7.193.003,482. Penurunan ini terus semakin kecil pada nilai K yang lebih tinggi. Berdasarkan grafik dan pola tersebut, titik "elbow" yang terlihat jelas berada pada K=4, yang menunjukkan jumlah cluster optimal. Pada titik ini, model berhasil membagi data dengan baik tanpa memerlukan jumlah cluster yang terlalu banyak, sehingga lebih efisien untuk analisis dan interpretasi lebih lanjut.

Tabel 4. 5 Hasil Perhitungan Operator *Clustering*

Cluster	***Measure Types***	***Davies Bouldin Index***
2	*Mixed Measures*	0.256
3		0.408
4		0.386
5		0.398
6		0.450
7		0.394
8		0.322
9		0.301

Hasil pembacaan *operator K-Means clustering* dengan menggunakan parameter *Measure Types Mixed Measures*, serta pembacaan operator *Cluster Distance Performance*, lalu pada bagian *Main Criterion* pilih menggunakan *Davies Bouldin Index (DBI)* [13]. Berdasarkan hasil perhitungan *Davies Bouldin Index (DBI)* untuk jumlah *cluster* (K) antara 2 hingga 9 menunjukkan variasi nilai *DBI* dengan pola tertentu. Pada K=2, nilai *DBI* terendah tercatat sebesar 0.256, yang mengindikasikan kualitas *clustering* yang sangat baik dengan jarak antar cluster yang cukup besar. Namun, seiring bertambahnya jumlah *cluster*, nilai *DBI* sedikit meningkat. Pada K=3 dan K=4, nilai *DBI* masing-masing adalah 0.408 dan 0.386, yang masih menunjukkan kualitas *clustering* yang baik meskipun ada penurunan kecil. Selanjutnya, pada K=5 dan K=6, nilai *DBI* meningkat menjadi 0.398 dan 0.450, yang menunjukkan bahwa *clustering* dengan jumlah *cluster* ini mulai mengalami penurunan kualitas segmentasi. Namun, pada K=7 dan K=8, nilai *DBI* kembali menurun menjadi 0.394 dan 0.322, yang menunjukkan adanya perbaikan kualitas clustering. Pada K=9, nilai *DBI* tercatat sebesar 0.301, yang juga menunjukkan clustering yang cukup baik, meskipun tidak sebaik pada K=2. Berdasarkan temuan ini dan dengan memperhitungkan hasil dari metode *Elbow* yang menunjukkan K=4 sebagai jumlah *cluster* optimal, K=4 tetap dipilih sebagai jumlah *cluster* terbaik karena memberikan keseimbangan yang optimal antara kualitas *clustering* dan kompleksitas model.

4.6. Evaluasi

Berdasarkan hasil perhitungan menggunakan metode *Elbow* dan *Davies Bouldin Index (DBI)*, dapat disimpulkan bahwa K=4 adalah jumlah *cluster* yang optimal. *Davies Bouldin Index (DBI)* adalah satu-satunya metode yang digunakan untuk mengurangi validitas cluster dalam metode pengumpulan data apapun[15]. Setelah proses *clustering* selesai, langkah selanjutnya adalah mengevaluasi kualitas hasil pengelompokan dengan menggunakan *operator Performance* untuk menghasilkan *Performance Vector*. Hasil yang diperoleh menunjukkan nilai *Avg. within centroid distance* sebesar 54.941.560,812, dengan rincian sebagai berikut: *Avg. within centroid distance_cluster_0*: 35.108.500,057, *Avg. within centroid distance_cluster_1*: 25.465.650,961, *Avg. within centroid distance_cluster_2*: 400.959.277,196, *Avg. within centroid distance_cluster_3*: 64.638.432,002, dan *Davies Bouldin Index* sebesar 0.386.

PerformanceVector

```
PerformanceVector:
Avg. within centroid distance: 54941560.812
Avg. within centroid distance_cluster_0: 35108500.057
Avg. within centroid distance_cluster_1: 25465650.961
Avg. within centroid distance_cluster_2: 400959277.196
Avg. within centroid distance_cluster_3: 64638432.002
Davies Bouldin: 0.386
```

Gambar 4. 11 Hasil *PerformanceVector*

Hasil *PerformanceVector* menunjukkan bahwa rata-rata jarak data dalam *cluster* ke *centroid(Avg. within centroid distance)* adalah 54.941.560,812, dengan perbedaan yang signifikan antar *cluster*. *Cluster 0* memiliki jarak terkecil (35.108.500,057), yang mengindikasikan adanya kesamaan yang tinggi antar data, sementara *Cluster 2* memiliki jarak yang sangat besar (400.959.277,196), yang menunjukkan penyebaran data yang kurang terkelompok. Nilai *Davies Bouldin Index (DBI)* sebesar 0.386 menunjukkan kualitas clustering yang baik, meskipun terdapat ketidakseimbangan dalam distribusi data. Pada *Cluster Model*, distribusi data terlihat dengan Cluster 0 berisi 83 item, Cluster 1 berisi 121 item, *Cluster 2* hanya berisi 14 item, dan *Cluster 3* berisi 38 item dari total 256 item. Sebagian besar data terkumpul di *Cluster 1*, sementara *Cluster 2* memiliki jumlah item yang sangat sedikit, yang mungkin disebabkan oleh distribusi data yang tidak merata. Secara keseluruhan, meskipun ada ketidakseimbangan, *model clustering* ini tetap menunjukkan performa yang baik dengan nilai *DBI* yang relatif rendah.

4.7. Interpretasi Hasil

Interpretasi *clustering* dijelaskan berdasarkan hasil pengelompokan yang telah dilakukan menggunakan algoritma *K-Means*. Interpretasi hasil cluster adalah proses menganalisis dan memberikan makna pada kelompok-kelompok (cluster) yang terbentuk setelah data dianalisis menggunakan Algoritma K-Means[16]. Pada penelitian ini menekankan bahwa pendekatan berbasis data, seperti *clustering*, mampu memberikan wawasan mendalam untuk pengelolaan stok dan promosi[17]. Berikut adalah penjelasan rinci dari masing-masing cluster:

1. *Cluster 0 (*Penjualan Stabil): Produk dalam *cluster* ini memiliki karakteristik penjualan yang relatif konsisten, menunjukkan permintaan yang stabil dari pelanggan.
2. *Cluster 1* (Penjualan Tinggi): Produk dalam *cluster* ini memiliki tingkat penjualan yang sangat tinggi, sehingga memerlukan perhatian lebih dalam pengelolaan stok dan strategi promosi.
3. *Cluster 2* (Produk Kurang Diminati): Produk dalam cluster ini menunjukkan permintaan yang rendah di pasar. Hal ini dapat menjadi dasar untuk evaluasi ulang strategi penjualan, seperti pemberian diskon atau penghapusan produk dari katalog.
4. *Cluster 3* (Penjualan Moderat): Produk dalam *cluster* ini memiliki tingkat penjualan sedang. Strategi promosi tambahan dapat membantu meningkatkan penjualan produk dalam kategori ini.

5. KESIMPULAN

Penelitian ini bertujuan untuk meningkatkan pengelompokan data penjualan Toko Buyung Upik JS di Lazada menggunakan algoritma *K-Means Clustering*. Proses analisis dilakukan melalui tahapan *Knowledge Discovery in Databases (KDD)*, meliputi seleksi data, pra-pemrosesan, transformasi, data mining, evaluasi, dan interpretasi.

Hasil penelitian menunjukkan bahwa algoritma K-Means berhasil mengelompokkan produk penjualan ke dalam empat *cluster* berdasarkan pola penjualan yaitu *Cluster 0* merupakan produk dengan penjualan stabil, yang menunjukkan konsistensi permintaan di pasar, *Cluster 1* adalah produk dengan penjualan tinggi, yang membutuhkan stok yang memadai dan promosi untuk mempertahankan performa, *Cluster 2* merupakan produk yang kurang diminati, memerlukan evaluasi strategi pemasaran seperti diskon atau promosi khusus, kemudian *Cluster 3* yaitu produk dengan penjualan moderat, yang memiliki potensi untuk ditingkatkan melalui strategi pemasaran tambahan.

Evaluasi *clustering* menggunakan *Davies-Bouldin Index (DBI)* memberikan nilai 0.386, yang menunjukkan kualitas pengelompokan yang cukup baik. Selain itu, jumlah *cluster* optimal (K=4) ditentukan melalui *Elbow Method*.

Penelitian ini memberikan kontribusi signifikan dalam memahami pola penjualan dan membantu toko dalam merancang strategi yang lebih efektif untuk pengelolaan stok, promosi, dan inovasi produk. Dengan pendekatan berbasis data, Toko Buyung Upik JS dapat mengoptimalkan sumber daya mereka untuk meningkatkan daya saing di pasar *e-commerce* yang kompetitif.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada pihak-pihak terkait yang telah memberi dukungan terhadap penelitian ini.

DAFTAR PUSTAKA

[1] A. A. Alya Putri and S. A. Rahmah, "Implementasi Data Mining Dengan Algoritma K-Means Clustering Untuk Analisis Bisnis Pada Perusahaan Asuransi," *Djtechno J. Teknol. Inf.*, vol. 5, no. 1, pp. 139–152, 2024, doi: 10.46576/djtechno.v5i1.4537.

[2] S. Pujiono, R. Astuti, and F. Muhamad Basysyar, "Implementasi Data Mining Untuk Menentukan Pola Penjualan Produk Menggunakan Algoritma K-Means Clustering," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 1, pp. 615–620, 2024, doi: 10.36040/jati.v8i1.8360.

[3] L. Awaliyah, N. Rahaningsih, and R. Danar Dana, "Implementasi Algoritma K-Means Dalam Analisis Cluster Korban Kekerasan Di Provinsi Jawa Barat," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 1, pp. 188–195, 2024, doi: 10.36040/jati.v8i1.8332.

[4] H. Rosika *et al.*, "PAKAIAN MENGGUNAKAN METODE K-MEANS Pada era digital perkembangan teknologi semakin berkembang khusunya pengelolaan data penjualan menjadi semakin krusial bagi bisnis untuk memahami perilaku konsumen , mengedentifikasi beberapa kelompok atau cluster memili," vol. 5, pp. 221–231, 2024.

[5] N. Afiasari, N. Suarna, and N. Rahaningsi, "Implementasi Data Mining Transaksi Penjualan Menggunakan Algoritma Clustering dengan Metode K-Means," *J. SAINTEKOM*, vol. 13, no. 1, pp. 100–110, 2023, doi: 10.33020/saintekom.v13i1.402.

[6] Normah, B. Rifai, S. Vambudi, and R. Maulana, "Analisa Sentimen Perkembangan Vtuber Dengan Metode Support Vector Machine Berbasis SMOTE," *J. Tek. Komput. AMIK BSI*, vol. 8, no. 2, pp. 174–180, 2022, doi: 10.31294/jtk.v4i2.

[7] M. Rafi Nahjan, Nono Heryana, and Apriade Voutama, "Implementasi Rapidminer Dengan Metode Clustering K-Means Untuk Analisa Penjualan Pada Toko Oj Cell," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 1, pp. 101–104, 2023, doi: 10.36040/jati.v7i1.6094.

[8] M. Rizki and M. Mulyawan, "Penerapan Metode K-Means Clustering Pada Data Penjualan Optik Chantika," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 2, pp. 1303–1307, 2023, doi: 10.36040/jati.v7i2.6562.

[9] I. Pii, N. Suarna, and N. Rahaningsih, "Penerapan Data Mining Pada Penjualan Produk Pakaian Dameyra Fashion Menggunakan Metode K-Means Clustering," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 1, pp. 423–430, 2023, doi: 10.36040/jati.v7i1.6336.

[10] M. Wahyudi, S. Solikhun, and L. Pujiastuti, "Komparasi K-Means Clustering dan K-Medoids Clustering dalam Mengelompokkan Produksi Susu Segar di Indonesia Berdasarkan Nilai DBI," *J. Bumigora Inf. Technol.*, vol. 4, no. 2, pp. 243–254, 2022, doi: 10.30812/bite.v4i2.2104.

[11] N. A. Hidayatullah and W. Prihartono, "CLUSTERING ALGORITMA K-MEANS UNTUK MENINGKATKAN EFEKTIVITAS PROGRAM SOSIAL DI KOTA / KABUPATEN CIREBON," vol. 13, no. 1, pp. 629–636, 2025.

[12] F. P. Dewi, P. S. Aryni, and Y. Umaidah, "Implementasi Algoritma K-Means Clustering Seleksi Siswa Berprestasi Berdasarkan Keaktifan dalam Proses Pembelajaran," *JISKA*

(Jurnal Inform. Sunan Kalijaga), vol. 7, no. 2, pp. 111–121, 2022, doi: 10.14421/jiska.2022.7.2.111-121.

[13] A. P. Bagustio, A. I. Purnamasari, and Irfan Ali, "Analisis Data Penjualan Menggunakan Algoritma K-Means Clustering Pada Toko Kecantikan Putri," *PROSISKO J. Pengemb. Ris. dan Obs. Sist. Komput.*, vol. 11, no. 2, pp. 159–167, 2024, doi: 10.30656/prosisko.v11i2.7928.

[14] P. Apriyani, A. R. Dikananda, and I. Ali, "Penerapan Algoritma K-Means dalam Klasterisasi Kasus Stunting Balita Desa Tegalwangi," *Hello World J. Ilmu Komput.*, vol. 2, no. 1, pp. 20–33, 2023, doi: 10.56211/helloworld.v2i1.230.

[15] N. Novitasari, N. D. Nuris, and R. Herdiana, "Penerapan Algoritma K-Means untuk Clustering Data Jumlah Penduduk Miskin Berdasarkan Kota/Kabupaten di Jawabarat menggunakan Rapidminer," *J. Inform. Terpadu*, vol. 9, no. 1, pp. 68–73, 2023, doi: 10.54914/jit.v9i1.660.

[16] V. No, J. Hal, L. Mayola, M. Hafizh, and H. Syahputra, "Klasterisasi Rumah Sakit berdasarkan Kunjungan Pasien menggunakan Algoritma K-Means : Data 2019-2023," vol. 7, no. 1, pp. 15–21, 2025.

[17] S. Handoko, F. Fauziah, and E. T. E. Handayani, "Implementasi Data Mining Untuk Menentukan Tingkat Penjualan Paket Data Telkomsel Menggunakan Metode K-Means Clustering," *J. Ilm. Teknol. dan Rekayasa*, vol. 25, no. 1, pp. 76–88, 2020, doi: 10.35760/tr.2020.v25i1.2677.