

Analisis Transaksi Penjualan Menggunakan Algoritma *K-Means* untuk Optimalisasi Manajemen Stok

Rafi Anggara Nindiaputra^{1*}, Yisti Vita Via², Hendra Maulana³

^{1,2,3} Informatika, Universitas Pembangunan Nasional “Veteran” Jawa Timur

²yistivia.if@upnjatim.ac.id

³hendra.maulana.if@upnjatim.ac.id

*Corresponding author email: 20081010242@student.upnjatim.ac.id

Abstrak— Penelitian ini bertujuan untuk meningkatkan efisiensi dan ketepatan dalam mengelola transaksi penjualan serta pengendalian persediaan barang dengan menerapkan algoritma *clustering K-Means*. Dalam kegiatan bisnis, pengelolaan stok yang kurang optimal sering menimbulkan masalah seperti kelebihan atau kekurangan barang, yang berdampak pada penurunan kinerja penjualan. Oleh karena itu, diperlukan metode analisis data yang mampu mengidentifikasi pola transaksi penjualan secara lebih mendalam dan objektif. Algoritma *K-Means* digunakan untuk mengelompokkan barang berdasarkan kesamaan pola transaksi penjualan, sehingga setiap kelompok atau *cluster* merepresentasikan barang-barang dengan karakteristik penjualan yang serupa. Proses *clustering* ini melibatkan pengolahan data transaksi yang kompleks untuk menemukan keterkaitan antara frekuensi pembelian, volume penjualan, serta waktu transaksi. Melalui hasil pengelompokan tersebut, perusahaan dapat memperoleh wawasan yang lebih jelas mengenai perilaku penjualan dan tren permintaan pasar. Implementasi metode *K-Means* diharapkan mampu membantu perusahaan dalam mengambil keputusan yang lebih akurat terkait manajemen stok, strategi pemasaran, serta penentuan harga. Selain itu, hasil penelitian ini juga dapat menjadi landasan dalam pengembangan sistem informasi penjualan yang lebih adaptif, efisien, dan cerdas dalam merespons perubahan kebutuhan pasar.

Kata Kunci— *K-Means*, *Clustering*, Transaksi, Manajemen Stok, Sistem Informasi Penjualan

I. PENDAHULUAN

Di tengah perkembangan dinamis di dunia bisnis, manajemen transaksi penjualan memainkan peran penting dalam memastikan keberlanjutan dan kesuksesan suatu perusahaan. Perusahaan, mulai dari skala kecil hingga besar, kini memanfaatkan kemajuan teknologi internet sebagai media untuk mempromosikan dan memasarkan produk mereka. Selain itu, internet juga digunakan sebagai platform untuk membeli dan menjual produk, yang dikenal sebagai *e-commerce* [1]. Dengan meningkatnya kompleksitas dan volume transaksi, diperlukan pendekatan yang efektif dan efisien dalam pengelolaan data penjualan untuk menghasilkan keputusan yang akurat dan tepat waktu. Oleh karena itu, penelitian ini bertujuan untuk mengeksplorasi potensi penerapan algoritma *clustering*, khususnya algoritma *K-Means*, dalam proses analisis dan peningkatan efektivitas transaksi penjualan. Algoritma *Clustering K-Means* adalah metode yang berfungsi untuk mengklasifikasikan sejumlah objek berdasarkan kesamaan atribut atau karakteristiknya dalam berbagai data inventaris [2]. Melalui penerapan pendekatan ini, fokus utama adalah mengelompokkan transaksi penjualan ke dalam kelompok yang sesuai, sehingga

memungkinkan identifikasi pola yang mungkin sulit dideteksi secara manual. Dalam hal ini, penerapan teknik data mining menggunakan algoritma *clustering K-Means* dianggap tepat untuk mengelompokkan data inventaris dan strategi pemasaran ke dalam berbagai kategori, seperti produk dengan penjualan tinggi, sedang, dan rendah, atau produk yang kurang populer di pasar [3]. Penelitian yang dilakukan oleh Muningsi dan Kiswati berfokus pada pengelompokan pelanggan dengan menggunakan aplikasi berbasis metode *Elbow* untuk optimasi. Hasil dari penelitian mereka menunjukkan bahwa pengelolaan pelanggan yang baik dapat membantu perusahaan dalam meningkatkan keuntungan [4]. Sementara itu, penelitian ini memiliki fokus yang berbeda, yaitu pada pengelompokan stok produk dengan tujuan untuk mencegah terjadinya penumpukan maupun kekurangan barang di gudang.

II. LANDASAN TEORI

A. Tinjauan Pustaka

Berdasarkan penelitian yang dilakukan oleh M. Hasyim Siregar berjudul “Pengelompokan Penjualan Alat Bangunan Menggunakan Metode *K-Means* (Studi Kasus di Toko Adi Bangunan)” pada tahun 2018, terungkap bahwa penggunaan algoritma *K-Means* berhasil membagi data penjualan produk menjadi dua kelompok utama, yaitu barang dengan permintaan tinggi dan barang dengan permintaan rendah. Hasil analisis menunjukkan bahwa Kelompok 1 mencakup dua barang dengan volume penjualan rendah, sementara Kelompok 2 mencakup delapan barang yang termasuk dalam kategori penjualan tinggi. Pengelompokan ini memberikan wawasan berharga bagi manajemen toko dalam menetapkan prioritas pembelian dan mengelola persediaan sesuai dengan tingkat penjualan.

Dari hasil tersebut dapat disimpulkan bahwa algoritma *K-Means* merupakan metode yang efektif dalam melakukan pengelompokan data penjualan. Penerapan metode ini membantu manajemen dalam mengambil keputusan strategis terkait stok barang yang perlu diprioritaskan untuk pembelian ulang. Dengan fokus pada produk yang memiliki tingkat penjualan tinggi, efisiensi operasional serta keuntungan toko dapat ditingkatkan secara signifikan. [12]

Berdasarkan hasil penelitian oleh Habibie Ed Dien, M. Hasyim Ratsanjani, Agung Adi Saputra, Ariadi Retno Tri Hayati Ririd, dan Bagas Satya Dian Nugraha dengan judul Analisis *Clustering* Patok Jalan Berbasis Geospasial Menggunakan *K-Means* dan Evaluasi *Davies-Boudin* (2024), penggunaan algoritma *K-Means* berhasil membagi patok jalan yang bermasalah menjadi tiga *cluster*: sedikit, sedang, dan banyak masalah. Kualitas *Clustering*

ini dinilai sangat baik, didukung oleh nilai *Davies-Bouldin Index* (DBI) sebesar 0,1656 yang mendekati nol, menunjukkan kekompakan internal dan pemisahan antar *cluster* yang optimal [1]. *Cluster* ketiga (C3) diidentifikasi sebagai kelompok dengan konsentrasi permasalahan *RHTG* terbanyak. Temuan ini penting karena memungkinkan prioritas perbaikan difokuskan pada *cluster* C3 tersebut [1]. Selain itu, penulis mencatat bahwa stabilitas *cluster* tercapai setelah empat kali iterasi, menguatkan validitas dari hasil *Clustering* ini [13].

B. Data Mining

Data Mining adalah proses mengidentifikasi wawasan berharga dari kumpulan data yang sangat besar. Dengan kata lain, penambangan data melibatkan penggalian atau ekstraksi pengetahuan baru dari jumlah data yang besar, yang pada akhirnya dapat mendukung proses pengambilan keputusan. Proses ini juga sering disebut sebagai penemuan pengetahuan. [6]. Ragam data yang dimanfaatkan sangat bervariasi dan meliputi berbagai sektor, mulai dari pemerintahan, pendidikan, keuangan, hingga bidang-bidang lainnya. Keragaman data ini selanjutnya diproses dan dikembangkan guna menghasilkan pemahaman yang lebih akurat serta bernilai. Pada dasarnya, Data Mining memiliki tujuan untuk menganalisis data sehingga mampu mengidentifikasi hubungan atau pola yang tersembunyi, sekaligus memformulasikan informasi baru yang berguna bagi pemilik data [7]. Secara garis besar, Data Mining dapat diklasifikasikan ke dalam dua kategori pokok:

1. Penambangan deskriptif (Descriptive mining), yaitu pendekatan yang diterapkan untuk mengidentifikasi dan mendeskripsikan karakteristik pokok dari sekumpulan data. Sejumlah metode yang masuk dalam kategori ini antara lain clustering, asosiasi, dan penambangan urutan.
2. Penambangan prediktif (Predictive mining), yaitu pendekatan yang menitikberatkan pada pembangunan model dari data untuk memproyeksikan nilai atau pola variabel pada masa mendatang. Salah satu metode yang tergolong dalam kelompok ini adalah klasifikasi."

C. Clustering

Pengelompokan (clustering) adalah teknik penambangan data yang digunakan untuk mengklasifikasikan sejumlah objek berdasarkan karakteristik mereka. Objek dalam kelompok yang sama cenderung memiliki karakteristik yang serupa, tetapi berbeda dengan objek dalam kelompok atau kluster lain. Oleh karena itu, kluster dapat didefinisikan sebagai kelompok objek yang digabungkan karena memiliki tingkat kesamaan atau kedekatan yang tinggi [8].

D. Algoritma K-Means Clustering

K-Means merupakan salah satu algoritma *clustering* yang digunakan untuk mengelompokkan data berdasarkan kesamaan karakteristik di dalam kelompoknya serta memaksimalkan perbedaan antar kelompok. Algoritma *K-Means* banyak digunakan karena efektivitasnya dalam mengkategorikan data atau produk ke dalam kelompok yang memiliki kesamaan tertentu [9].

K-Means adalah salah satu metode pengelompokan data yang bersifat non-hierarki, di mana data dibagi ke dalam dua atau lebih kelompok. Cara kerjanya adalah dengan mengelompokkan data berdasarkan kemiripan karakteristik data yang memiliki sifat atau nilai yang mirip akan dimasukkan ke dalam satu *cluster* yang sama, sedangkan data yang berbeda akan ditempatkan di *cluster* lain. Tujuan utama dari metode ini adalah untuk membuat setiap kelompok memiliki anggota yang sejenis (meminimalkan variasi dalam *cluster*) dan pada saat yang sama memperjelas perbedaan antara satu kelompok dengan kelompok lainnya. [15].

E. Penjualan

Penjualan adalah kegiatan yang melibatkan upaya penjual untuk memenuhi kebutuhan dan harapan pembeli, sehingga menghasilkan nilai yang berkelanjutan dan memberikan manfaat bagi kedua belah pihak. Selain itu, penjualan juga dapat dipahami sebagai hasil yang dihasilkan dari kegiatan atau layanan yang dilakukan dalam transaksi bisnis dan kegiatan komersial [11].

F. Davies Bouldin Index

Pengukuran validitas *cluster* dalam metode pengelompokan dilakukan untuk menilai kualitas hasil *Clustering*. *Kohesi* menggambarkan tingkat kedekatan antar data terhadap pusat *cluster* yang bersangkutan, sedangkan *pemisahan* menunjukkan jarak antar pusat *cluster*. *Indeks Davies-Bouldin* (DBI) digunakan untuk mengukur seberapa baik hasil pengelompokan data yang telah dilakukan. Indeks ini menilai keseimbangan antara jarak antar kelompok (*cluster*) dan kedekatan data di dalam satu kelompok. Tujuannya adalah agar jarak antar *cluster* (C_i dan C_j) menjadi sejauh mungkin, sementara data di dalam satu *cluster* saling berdekatan. Semakin jauh jarak antar *cluster*, semakin berbeda pula karakteristik antar kelompok. Nilai DBI yang kecil menunjukkan bahwa hasil pengelompokan sudah baik dan optimal, karena setiap *cluster* memiliki batas yang jelas dan anggota di dalamnya memiliki kemiripan yang tinggi. [10].

G. Evaluasi Model Clustering

Pengelompokan data yang paling efektif dicapai dengan mengevaluasi model pengelompokan. Proses ini bertujuan untuk mengukur tingkat akurasi dan mengidentifikasi hasil pengelompokan terbaik, yang kemudian dapat digunakan sebagai panduan dalam analisis selanjutnya. Untuk menentukan jumlah kluster yang ideal (k optimal), beberapa teknik evaluasi dapat diterapkan, seperti Metode *Elbow*, *Davies-Bouldin Index*, dan *Silhouette Coefficient* [14].

III. METODE PENELITIAN

A. Pengumpulan Data

TABEL I
DATA AWAL

No	Tanggal	Nama.pembeli	Nama.barang	Kuantum	Nominal
1	2020-01-02	Toko Heruniawati	Beras	1000	9840000
2	2020-01-02	Toko Herunawati	Daging	120	8400000
3	2020-01-02	Toko Aprilia Sukrisni	Beras	6000	62910000

Pada penelitian ini, *dataset* yang digunakan adalah *Dataset Penjualan Toko Sembako* yang diperoleh dari *website Kaggle* (www.kaggle.com) yang diunggah pada tahun 2022. Data ini memiliki total 1298 data dan menggunakan 6 atribut.

B. Metode yang Digunakan

Metode penelitian analisis penerapan metode *clustering K-Means* pada data transaksi penjualan dapat membantu mengelompokkan transaksi-transaksi tersebut ke dalam beberapa kelompok yang memiliki kesamaan karakteristik.. Berikut adalah langkah- langkah yang dapat diambil dalam penelitian ini :

1. Pengumpulan Data

Kumpulkan data transaksi penjualan yang mencakup berbagai atribut seperti jumlah barang, total harga, tanggal transaksi, dan atribut lain yang relevan.

2. Preprocessing Data

Bersihkan dan olah data untuk mengatasi masalah seperti data yang hilang atau duplikat, Lakukan normalisasi data jika diperlukan, seperti mengubah rentang nilai agar sejajar

3. Pemilihan Atribut

Tentukan atribut-atribut yang akan digunakan dalam analisis *clustering*. Misalnya, jumlah barang, total harga, dan lain-lain.

4. Pemilihan jumlah cluster (k)

Tentukan jumlah *cluster* (K) yang paling tepat. Salah satu cara yang sering digunakan adalah metode *Elbow*, yaitu dengan mencoba berbagai nilai K untuk melihat titik di mana terjadi "tekukan" atau perubahan signifikan pada grafik varians.

5. Penggunaan algoritma K-Means

Implementasikan Algoritma *K-Means* pada data transaksi penjualan. Inisialisasi *centroid* awal dan lakukan iterasi hingga konvergensi, yaitu ketika tidak ada perubahan yang signifikan dalam penempatan data ke dalam *cluster*.

IV. HASIL DAN PEMBAHASAN

A. Exploratory Data Analysis

Pada Gambar 1 tahap *Exploratory Data Analysis* dilakukan untuk memahami karakteristik awal dataset yang digunakan dalam penelitian. Tahap ini mencakup peninjauan struktur data, tipe data, distribusi nilai, serta pengecekan data hilang (*missing values*).

B. Preprocessing

Tahap *preprocessing* dilakukan untuk memastikan data dalam kondisi siap digunakan pada proses analisis dan pemodelan. Pada penelitian ini, *preprocessing* mencakup beberapa langkah yang di tunjukkan pada Gambar 2.

```
=====
EXPLORATORY DATA ANALYSIS
=====

1. Preview Data (10 baris pertama):
no    tanggal    nama.pembeli    nama.barang    kuantum    nominal
0 1 2020-01-02    TOKO HERUNIAWATI    BERAS    1000.0    9840000.0
1 2 2020-01-02    TOKO HERUNIAWATI    DAGING    120.0    8400000.0
2 3 2020-01-02    TOKO APRILIA SUKRISNI    BERAS    6000.0    62910000.0
3 4 2020-01-02    TOKO APRILIA SUKRISNI    MIGOR    400.0    4855200.0
4 5 2020-01-02    TOKO APRILIA SUKRISNI    TEPUNG    140.0    1162000.0
5 6 2020-01-02    TOKO MAJU TERUS    BERAS    820.0    8597700.0
6 7 2020-01-02    TOKO MAJU TERUS    BERAS    550.0    5412000.0
7 8 2020-01-02    TOKO MAJU TERUS    DAGING    140.0    9800000.0
8 9 2020-01-03    KOPERASI BRI MELAWI    BERAS    1180.0    12372300.0
9 10 2020-01-03    KOPERASI BRI MELAWI    MIGOR    204.0    2427600.0

2. Informasi Dataset:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1289 entries, 0 to 1288
Data columns (total 6 columns):
#   Column      Non-Null count  Dtype
---  --
0    no          1289 non-null   int64
1    tanggal     1289 non-null   object
2    nama.pembeli 1289 non-null   object
3    nama.barang  1289 non-null   object
4    kuantum      1289 non-null   float64
5    nominal     1289 non-null   float64
dtypes: float64(2), int64(1), object(3)
memory usage: 60.6+ KB
None

3. Statistik Deskriptif:
no          kuantum          nominal
count  1289.000000    1289.000000    1.289000e+03
mean    645.000000    1735.063615    1.929970e+07
std     372.246558    6300.788118    6.489967e+07
min      1.000000      2.000000      4.500000e+04
25%     323.000000    140.000000    2.964000e+06
50%     645.000000      600.000000    7.650000e+06
75%     967.000000     1800.000000    1.890000e+07
max     1289.000000    200000.000000  2.047000e+09

4. Missing Values:
no          0
tanggal     0
nama.pembeli 0
nama.barang 0
kuantum      0
nominal      0
dtype: int64
```

Gambar 1 Exploratory Data Analysis

```
=====
PREPROCESSING DATA
=====

1. Label Encoding untuk nama.barang:
BERAS -> 0
DAGING -> 1
GULA -> 2
MIGOR -> 3
TEPUNG -> 4
... (total 5 kategori)

2. Shape Data:
Features (X): (1289, 3)

3. ✓ Data berhasil dinormalisasi dengan StandardScaler
```

Gambar 2 Preprocessing Label Encoding

1. Label Encoding

Data pada atribut *nama.barang* yang awalnya berupa teks (kategori) diubah menjadi angka dengan menggunakan metode *Label Encoding*. Proses ini dilakukan agar data tersebut dapat dipahami dan diolah oleh algoritma *Machine Learning*, karena algoritma tersebut hanya bisa bekerja dengan data dalam bentuk numerik.

2. Shape Data

Setelah proses encoding, data fitur tersebut disusun dalam bentuk *array* atau matriks dengan ukuran tertentu. Dalam penelitian ini, *data* memiliki ukuran (1289, 3) yang berarti terdapat 1.289 sampel dengan 3 fitur pada setiap sampel.

3. Normalisasi Data

Pada tahap ini *data* dinormalisasi menggunakan metode *StandardScaler*. Normalisasi ini bertujuan untuk menyamakan skala seluruh fitur sehingga memiliki rata-rata 0 dan standar deviasi 1. Dengan demikian, algoritma

tidak akan bias terhadap fitur yang memiliki rentang nilai besar.

C. Penentuan Jumlah Cluster Optimal (Metode Elbow)

Pada Gambar 3 tahap ini data yang telah melalui tahap *preprocessing*, langkah selanjutnya adalah mencari jumlah *cluster* yang paling sesuai atau optimal untuk digunakan pada algoritma *K-Means*. Pada penelitian ini digunakan metode *Elbow* agar dapat mengukut nilai *inertia* (jumlah total jarak kuadrat antara data dan pusat *cluster* nya) untuk berbagai nilai *K* (jumlah *cluster*).

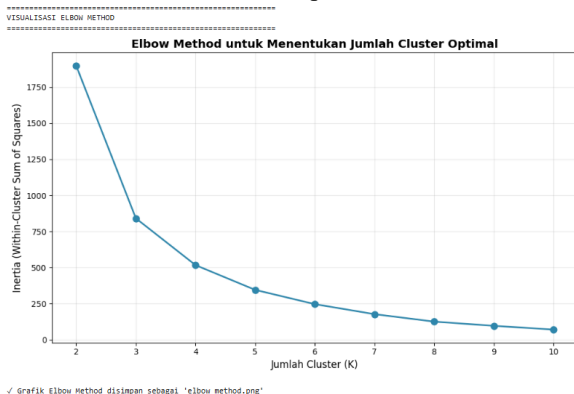
```
=====
ELBOW METHOD - MENCARI JUMLAH CLUSTER OPTIMAL
=====

Menghitung inertia untuk K=2 hingga K=10...
K=2: Inertia = 1897.25
K=3: Inertia = 840.77
K=4: Inertia = 516.78
K=5: Inertia = 345.37
K=6: Inertia = 246.44
K=7: Inertia = 177.11
K=8: Inertia = 125.48
K=9: Inertia = 96.40
K=10: Inertia = 70.41

✓ Perhitungan inertia selesai
```

Gambar 3 Perhitungan Nilai *Inertia*

Berdasarkan hasil perhitungan pada Gambar 3, terlihat bahwa semakin besar nilai *K*, nilai *inertia* semakin menurun. Namun, penurunan tersebut tidak selalu signifikan. Pada metode *Elbow*, titik optimal ditentukan ketika penurunan *inertia* mulai melandai atau membentuk sudut (*Elbow*). Pada data ini, penurunan *inertia* yang paling tajam terjadi pada *K* = 2 hingga *K* = 4 atau *K* = 5, sedangkan setelah itu penurunannya mulai melambat. Dengan demikian, jumlah *cluster* yang optimal berada pada *K* = 4 atau *K* = 5 karena setelah titik tersebut perubahan *inertia* tidak lagi signifikan dibandingkan dengan nilai sebelumnya. Visualisasi grafik metode *Elbow* mendukung hasil tersebut, di mana kurva menunjukkan pembentukan sudut pada rentang *K* tersebut. Hasil tersebut di visualisasi kan pada Gambar 4.



Gambar 4 Visualisasi Grafik Metode *Elbow*

D. Analisis Validasi Cluster

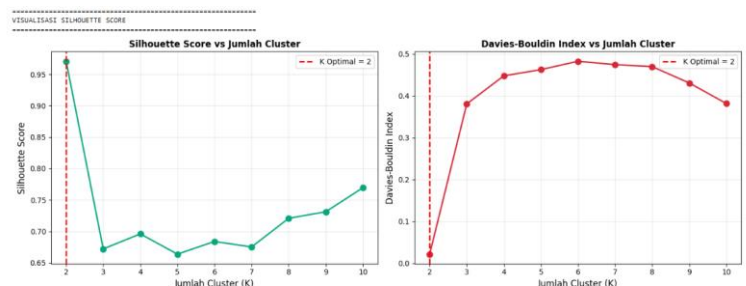
Berdasarkan pada Gambar 5, nilai *Silhouette Score* tertinggi adalah pada *K* = 2 (0,9707) sedangkan nilai *Davies-Bouldin Index* terendah juga terdapat pada *K* = 2 (0,0215). Artinya, pada jumlah *cluster* *K*=2, data memiliki pemisahan antar-*cluster* yang paling baik dan kompak.

```
=====
SILHOUETTE ANALYSIS
=====

Menghitung silhouette Score dan Davies-Bouldin Index...
K=2: Silhouette Score = 0.9707, Davies-Bouldin Index = 0.0215
K=3: Silhouette Score = 0.6721, Davies-Bouldin Index = 0.3806
K=4: Silhouette Score = 0.6960, Davies-Bouldin Index = 0.4477
K=5: Silhouette Score = 0.6638, Davies-Bouldin Index = 0.4626
K=6: Silhouette Score = 0.6839, Davies-Bouldin Index = 0.4826
K=7: Silhouette Score = 0.6752, Davies-Bouldin Index = 0.4745
K=8: Silhouette Score = 0.7208, Davies-Bouldin Index = 0.4698
K=9: Silhouette Score = 0.7312, Davies-Bouldin Index = 0.4307
K=10: Silhouette Score = 0.7694, Davies-Bouldin Index = 0.3815

✓ K Optimal berdasarkan silhouette Score: 2
✓ K Optimal berdasarkan Davies-Bouldin Index: 2
```

Gambar 5 Perhitungan nilai *Silhouette Score* dan *Davies-Bouldin Index*



Gambar 6 Grafik *Silhouette Score* dan *Davies-Bouldin Index*

E. Pelatihan Model *K-Means*

```
=====
TRAINING MODEL K-MEANS
=====

✓ Model K-Means berhasil dilatih dengan K=2
- Jumlah cluster: 2
- Jumlah iterasi: 15
- Inertia: 1897.25
```

Gambar 7 Training Model *K-Means*

Pada Gambar 7 tahap ini merupakan proses pelatihan model *clustering* dilakukan menggunakan algoritma *K-Means*. Berdasarkan hasil analisis jumlah *cluster* optimal menggunakan metode *Elbow*, *Silhouette Score*, dan *Davies-Bouldin Index*, jumlah *cluster* terbaik yang digunakan adalah *K*=2. Setelah model dilatih dengan parameter *K*= 2, Nilai iterasi sebanyak 15 menunjukkan jumlah proses perulangan yang dilakukan hingga algoritma mencapai kondisi konvergen atau pusat *cluster* tidak lagi berubah secara signifikan. Sementara itu, nilai *inertia* sebesar 1897.25 menggambarkan total jarak kuadrat antara setiap data dengan pusat *cluster* terdekat. Semakin rendah nilai *inertia*, semakin baik kualitas pembeda antar *cluster*.

F. Analisis Hasil Clustering

Setelah proses pelatihan model *K-Means* selesai dilakukan dengan jumlah *cluster* (*k*) = 2, tahap selanjutnya adalah menganalisis hasil pengelompokan data transaksi penjualan. Hasil keluaran sistem memberikan pemetaan data ke dalam dua *cluster* dengan karakteristik yang berbeda, yaitu *Cluster 0* dan *Cluster 1*.


```

=====
ANALISIS HASIL CLUSTERING
=====

1. Distribusi Cluster:
cluster
0      1
1      1288
Name: count, dtype: int64

2. Mapping Cluster ke Kategori Penjualan:
Cluster 0 → Tinggi
Cluster 1 → Sedang

3. Statistik per Cluster:
      kuantum      min      max      nominal      min \
cluster
0      200000.000000  200000.0  200000.0  2.047000e+09  2.047000e+09
1      1581.131211      2.0    50000.0  1.772540e+07  4.600000e+04

      max
cluster
0      2.047000e+09
1      5.100000e+08

4. Distribusi Kategori Penjualan:
kategori_penjualan
Sedang      1288
Tinggi       1
Name: count, dtype: int64

5. Sample Data dengan Cluster:
   no  tanggal  nama.pembeli nama.barang  kuantum  nominal \
0    1  2020-01-02  TOKO HERUNIAMATI  BERAAS    1000.0  9840000.0
1    2  2020-01-02  TOKO HERUNIAMATI  DAGING     120.0  8400000.0
2    3  2020-01-02  TOKO APRILIA SUKRISNI  BERAAS    6000.0  62910000.0
3    4  2020-01-02  TOKO APRILIA SUKRISNI  MIGOR     400.0  4855200.0
4    5  2020-01-02  TOKO APRILIA SUKRISNI  TEPUNG    140.0  1162000.0
5    6  2020-01-02  TOKO MAJU TERUS  BERAAS     820.0  8597700.0
6    7  2020-01-02  TOKO MAJU TERUS  BERAAS    550.0  5412000.0
7    8  2020-01-02  TOKO MAJU TERUS  DAGING    140.0  9800000.0
8    9  2020-01-03  KOPERASI BRI MELAWI  BERAAS   1180.0  12372300.0
9   10  2020-01-03  KOPERASI BRI MELAWI  MIGOR    204.0  2427600.0
10  11  2020-01-03  KOPERASI BRI MELAWI  TEPUNG    200.0  1660000.0
11  12  2020-01-06  TOKO DEB  BERAAS     500.0  4920000.0
12  13  2020-01-06  TOKO DEB  DAGING     100.0  7800000.0
13  14  2020-01-06  TOKO FATHULLAH ZIKRI  BERAAS     500.0  4920000.0
14  15  2020-01-06  TOKO FATHULLAH ZIKRI  DAGING    110.0  7700000.0

cluster kategori_penjualan
0      1      Sedang
1      1      Sedang

```

Gambar 8 Analisis Model Clustering

Berdasarkan Gambar 8 tersebut, terlihat bahwa sebagian besar data transaksi masuk ke dalam *Cluster* 1, sedangkan hanya satu data yang masuk ke dalam *Cluster* 0. Kondisi ini menunjukkan adanya kemungkinan *outlier* atau transaksi dengan nilai yang jauh lebih tinggi dibandingkan transaksi lainnya.

G. Hasil Analisis Clustering Menggunakan Algoritma K-Means

```

=====
SUMMARY & KESIMPULAN
=====

1. Total Data: 1289 baris

2. Model Clustering:
- Algoritma: K-Means
- Jumlah Cluster Optimal: 2
- Silhouette Score: 0.9707
- Davies-Bouldin Index: 0.0215

3. Distribusi Kategori Penjualan:
{'Sedang': 1288, 'Tinggi': 1}

4. Cluster Centers (Sebelum Scaling):
   cluster kategori  kuantum  nominal
0      Tinggi  200000.000000  2.047000e+09
1      Sedang   1581.131211  1.772540e+07

5. File Output:
- elbow_method.png
- cluster_validation.png
- cluster_visualization.png

6. Model siap digunakan untuk clustering data baru!

Catatan Perbedaan K-Means vs KNN:
- K-Means: Unsupervised learning (tidak butuh label)
- KNN: Supervised learning (butuh label untuk training)
- K-Means: Mengelompokkan data berdasarkan kemiripan
- KNN: Memprediksi berdasarkan tetangga terdekat

=====
PROGRAM SELESAI
=====

```

Gambar 9 Hasil Clustering K-Means

Berdasarkan Gambar 9 hasil pengolahan data dengan jumlah total 1.289 baris data, dilakukan proses *clustering* menggunakan algoritma *K-Means*. Tujuan dari analisis ini adalah untuk mengelompokkan data penjualan berdasarkan kesamaan karakteristik kuantitas dan nilai nominal penjualan.

V. KESIMPULAN

Analisis menunjukkan bahwa algoritma *K-Means* berhasil membagi data penjualan menjadi dua kelompok utama, yaitu “Sedang” dan “Tinggi”. Berkat nilai evaluasi yang sangat memuaskan (Silhouette Score 0.9707 dan Davies-Bouldin Index 0.0215), model ini dapat diterapkan secara efektif untuk memproses data penjualan baru di masa depan.

UCAPAN TERIMA KASIH

Bagian ucapan terima kasih dan daftar referensi tidak diberikan nomor. Saya menyampaikan rasa terima kasih kepada Tim SANTIKA yang telah meluangkan waktu dan usaha untuk menyusun template ini.

REFERENSI

- [1] Santosa, B. (2007). *Data Mining* teknik pemanfaatan data untuk keperluan bisnis. Yogyakarta: Graha Ilmu, 978(979), 756.
- [2] Ika Anikah, Agus Surip, Nela Puji Rahayu, Muhammad Harun AlMusa, & Edi Tohidi. (2022). Pengelompokan Data Barang Dengan Menggunakan Metode KMeans Untuk Menentukan Stok Persediaan Barang. *KOPERTIP : Jurnal Ilmiah Manajemen Informatika Dan Komputer*, 4(2), 58–64. <https://doi.org/10.32485/kopertip.v4i2.120>
- [3] Nabila, Z., Rahman Isnain, A., & Abidin, Z. (2021). Analisis *Data Mining* Untuk *Clustering* Kasus Covid-19 Di Provinsi Lampung Dengan Algoritma *K-Means*. *Jurnal Teknologi Dan Sistem Informasi (JTSI)*, 2(2), 100. <http://jim.teknokrat.ac.id/index.php/JTSI>
- [4] Muningsih, E., & Kiswati, S. (2018). Sistem Aplikasi Berbasis Optimasi Metode *Elbow* Untuk Penentuan *Clustering* Pelanggan. *Joutica*, 3(1), 117. <https://doi.org/10.30736/jti.v3i1.196>
- [5] Novita Lestari Anggreini. (2019). Teknik *Clustering* Dengan Algoritma K-Medoids Untuk Menangani Strategi Promosi Di Politeknik Tede Bandung. *Jurnal Teknologi Informasi Dan Pendidikan*, 12(2). <http://tip.ppi.unp.ac.id/j>
- [6] Sallaby, A. F., Alinse, R. T., Sari, V. N., & Ramadani, T. (2022). *Pengelompokan barang menggunakan metode K-Means clustering berdasarkan hasil penjualan di Toko Widya Bengkulu. Jurnal Media Infotama*, 18(1), 99–106..
- [7] Amin, F., Anggraeni, D. S., & Aini, Q. (2022). Penerapan Metode *K-Means* dalam Penjualan Produk Souq.Com. *Applied Information System and Management (AISM)*, 5(1), 7–14. <https://doi.org/10.15408/aism.v5i1.22534>
- [8] Sarimole, F. M., & Hakim, L. (2024). Klasifikasi barang menggunakan metode *clustering K-Means* dalam penentuan prediksi stok barang. *Jurnal Sains dan Teknologi*, 5(3), 846–854. <https://doi.org/10.55338/saintek.v5i1.2709>
- [9] Indriani, D., Irawan, B., & Bahtiar, A. (2024). Penerapan algoritma *K-Means clustering* untuk menentukan persediaan stok barang. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(1).
- [10] Pujiono, S., Astuti, R., & Basysyar, F. M. (2024). Implementasi *Data Mining* untuk Menentukan Pola Penjualan Produk Menggunakan Algoritma *K-Means Clustering*. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(1)
- [11] Pujiono, S., Astuti, R., & Basysyar, F. M. (2024). Implementasi *Data Mining* untuk menentukan pola penjualan produk menggunakan algoritma *K-Means clustering*. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(1).
- [12] Hasyim Siregar. (2018). *Clustering* penjualan alat-alat bangunan menggunakan metode *K-Means* (Studi kasus di Toko Adi Bangunan). *Jurnal Teknologi dan Open Source*, 1(2), 83-91.
- [13] Dien, H. E., Ratsanjani, M. H., Saputra, A. A., Noprianto, Ririd, A. R. T. H., & Nugraha, B. S. D. (2024). Analisis *Clustering* patok jalan berbasis geospasial menggunakan *K-Means* dan evaluasi Davies-Bouldin. *Jurnal Pekommas*, 9(2), 215–223.
- [14] Agustini, S. R., Purnamasari, I., & Sari, B. N. (2025). Implementasi *K-Means* untuk pengelompokan kategori penjualan barang berbasis web. *Journal of Informatics Management and Information Technology*, 5(3), 167–176. <https://doi.org/10.47065/jimat.v5i3.610>
- [15] Sallaby, A. F., Alinse, R. T., Sari, V. N., & Ramadani, T. (2022). Pengelompokan barang menggunakan metode *K-Means clustering* berdasarkan hasil penjualan di Toko Widya Bengkulu. *Jurnal Media Infotama*, 18(1), 99–[Halaman Akhir].