Vol. 2 No. 1, Januari 2026
e-ISSN : 3163-6594

Penerapan Data Mining Metode K-Means Clustering untuk Analisis Pola Gaya pada Fashion dalam Mengidentifikasi Preferensi Konsumen

Mulyati Eka Saputri[1], Onoy Rohaeni[2], Cahya Rahmawati[3], Ashafa Multazam[4], Betrand Athariq Delpiero[5]
Universitas Koperasi Indonesia[1,2,3,4,5]
mulyatiekasaputri@gmail.com[1], onoyrohaeni@ikopin.ac.id[2],
cahyarahmawati646@gmail.com[3], multazam.ashafa@gmail.com[4],
betrandathariq@gmail.com[5]

ABSTRAK

Penelitian ini bertujuan untuk menerapkan metode *data mining* menggunakan algoritma *K-Means Clustering* dalam menganalisis pola gaya pada produk fashion guna mengidentifikasi preferensi konsumen. Data yang digunakan merupakan data sekunder berupa *dataset Fashion Product Images (Small)* yang diperoleh dari *Kaggle*, dengan jumlah awal lebih dari 44.000 data produk *fashion*. Setelah melalui tahap pra-pemrosesan data yang meliputi pembersihan data, penghapusan nilai kosong dan data duplikat, transformasi data kategorikal menggunakan label *encoding*, serta normalisasi data, diperoleh sebanyak 7.091 data yang siap dianalisis. Atribut yang digunakan dalam proses *clustering* meliputi *gender, masterCategory, subCategory, articleType, baseColour, season, usage,* dan *year*. Penentuan jumlah klaster optimal dilakukan menggunakan metode Elbow dengan analisis nilai *Within-Cluster Sum of Squares* (WCSS), yang menunjukkan bahwa jumlah klaster terbaik adalah empat klaster. Hasil penerapan algoritma *K-Means Clustering* menghasilkan pengelompokan produk *fashion* ke dalam empat klaster dengan karakteristik pola gaya yang berbeda, yaitu gaya kasual sehari-hari, gaya formal dan semi-formal, gaya kasual modern dengan variasi warna, serta gaya *sportswear* dan aktivitas santai. Setiap klaster merepresentasikan kecenderungan preferensi konsumen terhadap jenis produk fashion tertentu. Hasil penelitian ini menunjukkan bahwa algoritma *K-Means Clustering* efektif digunakan untuk mengidentifikasi pola gaya dan preferensi konsumen pada produk fashion digital, sehingga dapat dimanfaatkan sebagai dasar dalam pengembangan strategi pemasaran, segmentasi produk, dan pengambilan keputusan pada industri fashion digital.

Kata Kunci: *data mining, K-Means clustering, produk fashion, preferensi konsumen*

I. PENDAHULUAN

Industri fashion merupakan salah satu sektor yang mengalami pertumbuhan paling pesat di era digital, dengan nilai pasar global yang diperkirakan mencapai lebih dari 1,7 triliun dolar AS pada tahun 2024 dan terus meningkat seiring perkembangan *e-commerce* serta media sosial (Statista, 2024; Sulianta, Ulfah, & Amalia, 2024). Peningkatan ini didorong oleh perubahan perilaku konsumen yang semakin dinamis, di mana preferensi terhadap gaya, warna, dan identitas visual menjadi faktor penting dalam keputusan pembelian produk fashion (Dalimunthe & Putri, 2024). Perkembangan tersebut menghasilkan volume data yang sangat besar dari transaksi penjualan, katalog produk, dan interaksi digital konsumen, sehingga menuntut adanya analisis data yang lebih efektif untuk memahami pola dan tren konsumen di industri fashion modern (Yulianti et al., 2019).

Pemanfaatan data dalam industri fashion masih belum optimal, terutama dalam konteks analisis pola gaya dan preferensi konsumen. Banyak pelaku usaha yang hanya memanfaatkan data penjualan untuk tujuan administratif, tanpa menggunakannya untuk menggali wawasan mendalam terkait perilaku pelanggan atau pengelompokan produk berdasarkan karakteristik visual dan deskriptifnya. Padahal, analisis yang tepat dapat membantu perusahaan dalam

memahami tren mode, mengoptimalkan strategi pemasaran, serta mengelompokkan produk berdasarkan kesamaan atribut (Siregar, 2018).

Untuk menghadapi tantangan tersebut, diperlukan penerapan metode analisis data yang mampu mengekstraksi informasi penting dari kumpulan data berukuran besar. Salah satu teknik yang dapat digunakan adalah *data mining*. Menurut Nurajizah (2019), *data mining* merupakan proses penggalian pengetahuan dari sejumlah data besar untuk menemukan pola atau hubungan yang bermakna menggunakan algoritma tertentu. Melalui teknik ini, perusahaan dapat memperoleh wawasan yang lebih dalam mengenai perilaku konsumen dan pola pembelian, yang kemudian dapat digunakan sebagai dasar pengambilan keputusan strategis.

Salah satu metode data mining yang banyak digunakan dalam analisis pola data adalah *K-Means Clustering*. Algoritma ini merupakan metode *unsupervised learning* yang berfungsi untuk mengelompokkan data ke dalam sejumlah klaster berdasarkan tingkat kemiripan atribut (Bahar, Pramono, & Sagala, 2016). *K-Means* memiliki kemampuan untuk mengidentifikasi struktur tersembunyi dalam data tanpa memerlukan label awal, sehingga sangat cocok digunakan untuk segmentasi produk atau konsumen dalam industri *fashion*.

Penelitian terdahulu menunjukkan efektivitas algoritma *K-Means* dalam berbagai konteks industri fashion. Normah et al. (2021) mengimplementasikan metode *K-Means* untuk menganalisis penjualan produk hijab di Banten dan berhasil mengelompokkan produk menjadi tiga kategori: sangat laris, laris, dan kurang laris. Penelitian serupa oleh Yulianti et al. (2019) juga membuktikan bahwa *K-Means* efektif dalam mengetahui minat konsumen terhadap produk *fashion* berdasarkan atribut warna dan jenis pakaian. Temuan tersebut menunjukkan bahwa metode ini dapat membantu perusahaan fashion memahami perilaku konsumen melalui analisis berbasis data yang objektif.

Dalam konteks penelitian ini, digunakan dataset *Fashion Product Images (Small)* yang diperoleh dari *platform Kaggle*, berisi lebih dari 44.000 data produk *fashion* dengan atribut seperti kategori utama, subkategori, warna dasar, musim, dan jenis penggunaan. *Dataset* tersebut memiliki potensi besar untuk dianalisis menggunakan algoritma *K-Means Clustering* guna menemukan pola gaya produk yang merepresentasikan preferensi konsumen terhadap gaya fashion tertentu.

Beberapa penelitian terkini juga memperkuat relevansi penerapan algoritma ini. Dalimunthe & Putri (2024) menerapkan *K-Means* untuk menganalisis data penjualan pakaian wanita di *marketplace* dan berhasil menemukan pola pembelian dominan berdasarkan warna serta kategori produk. Penelitian oleh Sulianta et al. (2024) dalam *International Journal of Engineering Continuity* juga menunjukkan bahwa *K-Means* efektif dalam mengelompokkan konsumen berdasarkan kemiripan gaya, warna, dan preferensi produk.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk menerapkan metode data mining menggunakan algoritma *K-Means Clustering* dalam menganalisis pola gaya pada produk *fashion*, mengidentifikasi atribut yang memengaruhi pembentukan klaster, serta memahami preferensi konsumen berdasarkan hasil pengelompokan data. Hasil penelitian ini diharapkan dapat memberikan manfaat berupa pemahaman yang lebih baik mengenai penerapan data mining di industri fashion, membantu mengidentifikasi pola gaya konsumen,

serta menjadi referensi dalam penyusunan strategi pemasaran dan segmentasi produk pada industri fashion digital.

II. TINJAUAN PUSTAKA

Tinjauan pustaka memuat dasar teori, konsep, serta hasil penelitian terdahulu yang relevan dengan penelitian ini. Bagian ini bertujuan untuk memberikan landasan konseptual mengenai penerapan *data mining*, khususnya algoritma *K-Means Clustering*, dalam menganalisis pola gaya dan preferensi konsumen pada produk fashion digital.

A. Penelitian Terdahulu

Penelitian terdahulu merupakan acuan penting dalam menentukan arah dan kontribusi penelitian yang dilakukan saat ini. Penelitian ini merujuk pada beberapa studi yang membahas penerapan metode *data mining* dan algoritma *K-Means Clustering* dalam analisis data, khususnya dalam menganalisis pola gaya serta preferensi konsumen di bidang fashion. Peneliti memetakan empat penelitian terdahulu berdasarkan topik, konsep atau teori yang digunakan, metodologi yang diterapkan, serta hasil penelitian yang relevan dengan topik Penerapan Data Mining Metode *K-Means Clustering* untuk Analisis Pola Gaya pada Produk Fashion dalam Mengidentifikasi Preferensi Konsumen.

Tabel 1. Daftar Penelitian Terdahulu yang Relevan

No	Peneliti	Tahun	Judul Penelitian	Hasil Penelitian
1	**Normah, Nurajizah, & Salbinda**	2021	*Penerapan Data Mining Metode K-Means Clustering untuk Analisa Penjualan pada Toko Fashion Hijab Banten*	Berhasil mengelompokkan produk menjadi tiga kategori penjualan: sangat laris, laris, dan kurang laris.
2	**Yulianti, Utami, Hikmah, & Hasan**	2019	*Penerapan Data Mining Menggunakan Algoritma K-Means untuk Mengetahui Minat Customer di Toko Hijab*	Menghasilkan klaster pelanggan berdasarkan warna dan jenis produk yang paling diminati.
3	**Dalimunthe & Putri**	2024	*Data Mining on Women's Clothing Sales in Marketplaces with the K-Means Clustering Algorithm*	Menemukan pola pembelian dominan berdasarkan warna dan kategori produk.
4	**Sulianta, Ulfah, & Amalia**	2024	*Revealing Consumer Preferences in the Fashion Industry Using K-Means Clustering*	Mengelompokkan preferensi konsumen berdasarkan atribut visual seperti warna dan gaya produk.

B. Landasan Teori

1. Penelitian Kuantitatif dengan Pendekatan Eksploratif

Penelitian kuantitatif dengan pendekatan eksploratif bertujuan untuk menggali pola, tren, atau hubungan baru dari data numerik tanpa berangkat dari hipotesis yang bersifat kausal. Pendekatan ini bersifat fleksibel dan terbuka karena berfokus pada eksplorasi data untuk menemukan keterkaitan antarvariabel. Dalam penelitian ini, pendekatan kuantitatif eksploratif digunakan untuk mengidentifikasi pola gaya dan preferensi konsumen pada produk *fashion* melalui penerapan metode *data mining*.

2. Data Sekunder

Data sekunder merupakan data yang dikumpulkan oleh pihak lain dan digunakan kembali untuk keperluan analisis tertentu. Penggunaan data sekunder memiliki keunggulan dari segi efisiensi waktu dan biaya serta cakupan data yang luas. Penelitian ini menggunakan *dataset Fashion Product Images (Small)* yang diperoleh dari *Kaggle*, yang berisi ribuan data produk

fashion dengan berbagai atribut yang relevan untuk analisis pola gaya dan preferensi konsumen.

3. Konsep *Data Mining*

Data mining adalah proses penggalian pengetahuan dari kumpulan data berukuran besar untuk menemukan pola atau informasi yang bermakna. Proses ini merupakan bagian dari *Knowledge Discovery in Databases (KDD)* yang mencakup tahap pembersihan, transformasi, dan analisis data. Dalam penelitian ini, data mining digunakan untuk menggali pola gaya produk fashion dan preferensi konsumen melalui metode *clustering*.

4. Konsep *Clustering* dalam *Data Mining*

Clustering merupakan teknik dalam data mining yang digunakan untuk mengelompokkan data ke dalam beberapa klaster berdasarkan tingkat kemiripan atribut. Teknik ini termasuk dalam metode *unsupervised learning* karena tidak memerlukan label awal. Dalam penelitian ini, *clustering* digunakan untuk menemukan pola tersembunyi pada data produk *fashion* berdasarkan kesamaan atribut seperti kategori, warna, musim, dan tujuan penggunaan.

5. Algoritma *K-Means Clustering*

Algoritma *K-Means* adalah metode *clustering* yang membagi data ke dalam sejumlah klaster berdasarkan jarak ke pusat klaster *(centroid).* Proses *K-Means* meliputi penentuan jumlah klaster, inisialisasi *centroid*, perhitungan jarak menggunakan *Euclidean Distance*, pengelompokan data ke klaster terdekat, serta pembaruan *centroid* hingga konvergen. *K-Means* banyak digunakan karena efisien dalam menangani data berukuran besar dan mampu mengidentifikasi pola tersembunyi.

6. *Preprocessing* Data

Preprocessing data merupakan tahap penting dalam analisis data untuk memastikan kualitas dan konsistensi data. Tahapan *preprocessing* meliputi pembersihan data, normalisasi, pengkodean data kategorikal, serta seleksi fitur. Penerapan preprocessing yang tepat akan menghasilkan klaster yang lebih akurat dan mudah diinterpretasikan. Dalam penelitian ini, *preprocessing* diterapkan pada dataset fashion sebelum dianalisis menggunakan algoritma *K-Means.*

7. Analisis Deskriptif dalam Penelitian Eksploratif

Analisis deskriptif digunakan untuk menggambarkan karakteristik data tanpa melakukan generalisasi. Dalam penelitian eksploratif, analisis ini membantu memahami distribusi data, kecenderungan variabel, serta pola yang terbentuk dari hasil *clustering*. Analisis deskriptif digunakan untuk menginterpretasikan karakteristik setiap klaster produk *fashion*.

8. Analisis Pola dan Preferensi Konsumen

Analisis pola dan preferensi konsumen bertujuan untuk memahami kecenderungan konsumen dalam memilih produk berdasarkan karakteristik tertentu. Dalam industri *fashion*, preferensi konsumen dapat tercermin dari atribut seperti warna, gaya, dan kategori produk.

Hasil clustering digunakan untuk mengidentifikasi kelompok produk yang mencerminkan preferensi konsumen tertentu.

9. Industri *Fashion* Digital

Industri *fashion* digital berkembang pesat seiring dengan kemajuan teknologi dan meningkatnya aktivitas belanja daring. Pemanfaatan teknologi digital memungkinkan industri fashion untuk meningkatkan efisiensi produksi, pemasaran, serta inovasi produk. Namun, masih diperlukan optimalisasi pemanfaatan data dan teknologi digital agar industri *fashion*, khususnya di Indonesia, mampu bersaing di era digitalisasi.

III. METODE PENELITIAN

Bagian ini menjelaskan tahapan penelitian yang dilakukan, meliputi jenis dan pendekatan penelitian, desain penelitian, sumber dan jenis data, variabel penelitian, teknik pengumpulan data, teknik pengolahan dan analisis data, hipotesis penelitian, serta alat dan perangkat yang digunakan.

A. Jenis dan Pendekatan Penelitian

Penelitian ini menggunakan pendekatan kuantitatif eksploratif. Pendekatan ini dipilih karena penelitian berfokus pada proses eksplorasi data untuk menemukan pola atau kecenderungan tertentu tanpa menguji hubungan sebab-akibat antar variabel. Metode yang digunakan adalah *data mining* dengan algoritma *K-Means Clustering*, yang bertujuan untuk mengelompokkan produk *fashion* berdasarkan kesamaan atribut seperti warna, kategori, dan gaya. Melalui pendekatan ini diharapkan dapat diperoleh pemahaman yang lebih mendalam mengenai pola gaya dan preferensi konsumen terhadap produk fashion digital.

B. Desain Penelitian

Desain penelitian yang digunakan adalah desain eksploratif kuantitatif berbasis data mining. Penelitian eksploratif bertujuan untuk menggali pola tersembunyi dalam data tanpa menguji hipotesis statistik tertentu. Tahapan penelitian meliputi pengumpulan data, *preprocessing* data, penerapan algoritma *K-Means Clustering*, analisis hasil *clustering*, serta interpretasi hasil untuk menarik kesimpulan mengenai preferensi konsumen terhadap produk fashion.

C. Sumber dan Jenis Data

Data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari *dataset Fashion Product Images (Small)* yang tersedia pada platform *Kaggle* (Param Aggarwal, 2024). *Dataset* ini berisi lebih dari 44.000 data produk *fashion* dengan atribut seperti *gender, masterCategory, subCategory, articleType, baseColour, season, usage,* dan *year*. Jenis data yang digunakan terdiri atas data kategorikal dan data numerik. Data kategorikal mencakup atribut *gender, masterCategory, subCategory, articleType, baseColour, seaso*n, dan *usage*, sedangkan data numerik berupa atribut *year*.

D. Variabel Penelitian

Variabel dalam penelitian ini merupakan atribut yang dijadikan dasar dalam proses pengelompokan *(clustering).* Atribut-atribut yang digunakan dijelaskan pada tabel 2.

Tabel 2. Daftar Variabel Penelitian yang Digunakan dalam Prses Clustering

No	Nama Atribut	Deskripsi
1	Gender	Jenis kelamin target pengguna (Men/Women).
2	MasterCategory	Kategori utama produk (Apparel, Footwear, Accessories, dll).
3	SubCategory	Subkategori produk seperti Topwear, Bottomwear, atau Watches.
4	ArticleType	Jenis artikel fashion seperti Shirts, Jeans, atau Tshirts.
5	BaseColour	Warna utama produk fashion.
6	Season	Musim penggunaan seperti Summer, Fall, atau Winter.
7	Usage	Jenis penggunaan seperti Casual, Formal, atau Ethnic.
8	Year	Tahun rilis produk fashion.

E. Teknik Pengumpulan Data

Teknik pengumpulan data dilakukan dengan metode dokumentasi melalui pengunduhan dataset dari sumber terbuka. Langkah-langkah pengumpulan data antara lain:

1. Mengakses situs *Kaggle* pada tautan: https://www.kaggle.com/datasets/paramaggarwal/fashion-product-images-small
2. Mengunduh *dataset* dalam format .csv.
3. Menyeleksi atribut yang relevan untuk penelitian.
4. Menyimpan *dataset* untuk digunakan dalam proses analisis di *Python*.

F. Teknik Pengolahan Data

Pengolahan data merupakan tahapan penting dalam penelitian ini karena sangat memengaruhi kualitas hasil analisis yang diperoleh. Seluruh proses pengolahan data dilakukan menggunakan bahasa pemrograman *Python* melalui platform *Google Colab*. Tahapan pengolahan data dalam penelitian ini meliputi beberapa langkah sebagai berikut.

1. *Data Cleaning*

Tahap *data cleaning* bertujuan untuk memastikan data yang digunakan memiliki kualitas yang baik dan bebas dari kesalahan yang dapat memengaruhi hasil *clustering*. Proses yang dilakukan pada tahap ini meliputi:

- Menghapus baris data yang mengandung nilai kosong *(missing values)* menggunakan metode *dropna*, karena algoritma *K-Means* tidak dapat memproses data yang tidak lengkap.
- Menghapus data duplikat untuk menghindari bias dalam proses pengelompokan data.
- Menyeragamkan format penulisan atribut seperti warna dan kategori produk agar tidak terjadi inkonsistensi data.
- Menghapus atribut yang tidak relevan terhadap tujuan penelitian, seperti *id* dan *productDisplayName*, karena tidak berkontribusi dalam pembentukan pola gaya produk *fashion*.

2. *Data Transformation*

Sebagian besar atribut dalam dataset bersifat kategorikal sehingga tidak dapat diproses secara langsung oleh algoritma *K-Means*. Oleh karena itu, dilakukan proses transformasi data untuk mengonversi atribut kategorikal menjadi bentuk numerik. Teknik transformasi yang digunakan adalah *Label Encoding*, yang diterapkan pada atribut *gender, masterCategory*,

subCategory, articleType, baseColour, season, dan *usage*. Transformasi ini dilakukan agar seluruh atribut dapat digunakan dalam perhitungan jarak *Euclidean* pada proses *clustering*.

3. *Normalization*

Normalisasi data dilakukan untuk menyamakan skala antar variabel numerik sehingga tidak terdapat atribut tertentu yang mendominasi perhitungan jarak. Metode yang digunakan adalah *Min-Max Normalization*, yang mengubah rentang nilai data ke dalam interval 0 hingga 1. Dengan metode ini, setiap atribut memiliki bobot yang seimbang dalam proses *clustering* menggunakan algoritma *K-Means*.

4. *Feature Selection*

Tahap *feature selection* dilakukan untuk memilih atribut yang paling relevan dalam mengidentifikasi pola gaya dan preferensi konsumen terhadap produk fashion. Atribut yang digunakan dalam proses *clustering* meliputi:

- *Gender*
- *masterCategory*
- *subcategory*
- *articleType*
- *baseColour*
- *season*
- *usage*

5. Proses *Clustering*

Tahap inti penelitian ini adalah proses *clustering* menggunakan algoritma K-*Means Clustering*, dengan langkah-langkah sebagai berikut:

a. Menentukan jumlah klaster optimal (nilai K) menggunakan metode *Elbow Method* berdasarkan nilai *Within-Cluster Sum of Squares (WCSS).*
b. Menentukan pusat klaster *(centroid)* awal secara acak.
c. Menghitung jarak setiap data ke pusat klaster menggunakan *Euclidean Distance.*
d. Mengelompokkan data ke dalam klaster terdekat.
e. Memperbarui posisi *centroid* hingga tidak terjadi perubahan signifikan pada pembentukan klaster.

6. Evaluasi dan Visualisasi

Evaluasi hasil clustering dilakukan untuk menilai kualitas dan karakteristik klaster yang terbentuk, dengan cara:

- Menganalisis distribusi jumlah data pada setiap klaster.
- Mengidentifikasi atribut dominan pada masing-masing klaster.
- Membandingkan klaster berdasarkan kategori produk, warna, dan jenis penggunaan.

Hasil clustering divisualisasikan dalam bentuk:

- Grafik *Elbow Method* untuk menentukan jumlah klaster optimal.
- Diagram batang distribusi klaster untuk menunjukkan persebaran produk fashion pada setiap klaster.

7. Interpretasi Hasil

Tahap akhir adalah interpretasi hasil clustering untuk mengidentifikasi pola gaya dominan pada setiap klaster. Hasil interpretasi ini digunakan untuk memahami preferensi konsumen terhadap produk fashion digital.

G. Teknik Analisis Data

Teknik analisis data menggunakan algoritma *K-Means Clustering*. Jumlah klaster optimal ditentukan menggunakan metode *Elbow* dengan menganalisis nilai *Within-Cluster Sum of Squares (WCSS)*. Proses *clustering* dilakukan dengan menghitung jarak *Euclidean* antara data dan pusat klaster *(centroid),* mengelompokkan data ke klaster terdekat, serta memperbarui *centroid* hingga tercapai kondisi konvergen. Hasil clustering dianalisis untuk mengidentifikasi karakteristik klaster dan pola gaya produk fashion.

H. Hipotesis Penelitian

Penelitian ini menggunakan hipotesis eksploratif-deskriptif karena tidak menguji hubungan sebab-akibat secara langsung. Hipotesis penelitian menyatakan bahwa penerapan algoritma *K-Means Clustering* mampu mengelompokkan produk fashion berdasarkan kesamaan atribut secara jelas dan terstruktur. Selain itu, diasumsikan bahwa terdapat pola gaya dominan yang mencerminkan preferensi konsumen terhadap produk fashion digital dan dapat digunakan sebagai dasar penyusunan strategi produk dan pemasaran.

I. Alat dan Perangkat yang Digunakan

Dalam penelitian ini digunakan beberapa alat bantu sebagai berikut:

1. Perangkat Lunak :
   - *Python* (melalui *Google Colab*)
   - Libr*aries: Pandas, NumPy, Scikit-Learn, Matplotlib, Seaborn.*
2. Dataset : *Fashion Product Images (Small)* dari *Kaggle.*
3. Perangkat Keras : Laptop/PC dengan spesifikasi minimal RAM 8GB dan koneksi internet stabil.

IV. HASIL PENELITIAN DAN PEMBAHASAN

A. Gambaran Umum Data Penelitian

Penelitian ini menggunakan data sekunder berupa data produk *fashion* yang diperoleh dari berkas data *skripsi.csv*. Data awal terdiri dari sekitar 44.000 data produk fashion dengan sepuluh atribut utama. Namun demikian, berdasarkan hasil observasi awal terhadap dataset, ditemukan adanya nilai kosong *(missing value)* pada beberapa atribut, sehingga data tersebut tidak dapat digunakan secara langsung dalam proses analisis.

Sesuai dengan tahapan metodologi penelitian yang telah dijelaskan, dilakukan proses pembersihan data *(data cleaning)* dengan menghapus data yang memiliki nilai kosong menggunakan metode *listwise deletion (dropna).* Selain itu, data duplikat juga dihilangkan untuk menghindari bias dalam proses pengelompokan. Setelah melalui tahapan tersebut, jumlah data yang digunakan dalam penelitian ini menjadi sebanyak 7.091 data yang telah bersih dan siap untuk dianalisis lebih lanjut.

Atribut yang digunakan dalam proses *clustering* meliputi *gender*, *masterCategory*, *subCategory*, *articleType*, *baseColour*, *season*, *usage*, dan *year*. Pemilihan atribut ini didasarkan pada pertimbangan bahwa atribut-atribut tersebut mampu merepresentasikan karakteristik utama produk fashion yang berkaitan dengan pola gaya serta preferensi konsumen.

Tabel 3. Struktur Atribut Data Penelitian

No	Atribut	Keterangan
1	Gender	Jenis kelamin pengguna
2	Master Category	Kategori utama produk
3	Sub Category	Subkategori produk
4	Article Type	Jenis produk
5	Base Colour	Warna dasar produk
6	Season	Musim penggunaan
7	Usage	Tujuan penggunaan
8	Year	Tahun produksi

B. Tahap Pra-Pemrosesan Data

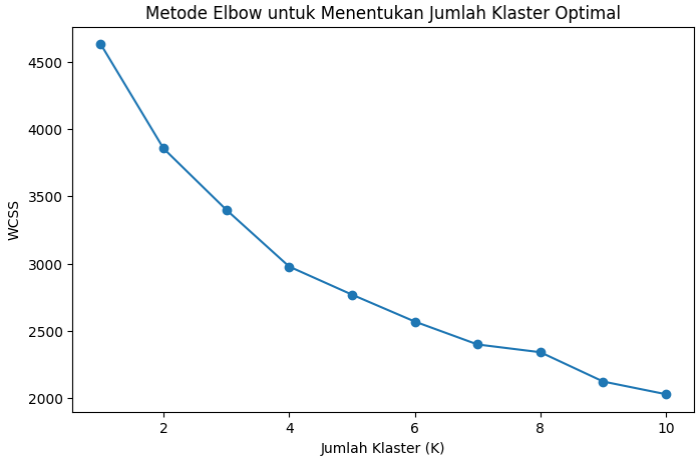
Tahap pra-pemrosesan data dilakukan untuk memastikan bahwa data yang digunakan dalam proses *clustering* memiliki kualitas yang baik dan sesuai dengan kebutuhan algoritma *K-Means*. Tahapan pra-pemrosesan data dalam penelitian ini meliputi penghapusan nilai kosong, pengkodean data kategorikal, dan normalisasi data.

Data kategorikal dikonversi ke dalam bentuk numerik menggunakan teknik label *encoding*. Proses ini diperlukan karena algoritma *K-Means* menggunakan perhitungan jarak berbasis numerik sehingga tidak dapat memproses data kategorikal secara langsung. Selanjutnya, dilakukan normalisasi data menggunakan metode scaling untuk menyamakan rentang nilai antar atribut, sehingga tidak terdapat atribut tertentu yang mendominasi proses perhitungan jarak.

Tahap pra-pemrosesan data ini merupakan langkah yang sangat penting karena kualitas data yang baik akan menghasilkan klaster yang lebih akurat, stabil, dan representatif dalam menggambarkan pola gaya produk fashion.

C. Penentuan Jumlah Klaster Optimal

Penentuan jumlah klaster optimal dilakukan menggunakan metode *Elbow* dengan menghitung nilai *Within Cluster Sum of Squares (WCSS)* pada beberapa nilai K. Metode *Elbow* digunakan untuk menemukan titik di mana penambahan jumlah klaster tidak lagi memberikan penurunan WCSS yang signifikan.



Gambar 1. Grafik Elbow Method Pemetaan Jumlah Klaster Optimal

Berdasarkan Gambar 1, terlihat bahwa nilai WCSS mengalami penurunan yang cukup signifikan pada rentang nilai K = 1 hingga K = 4. Setelah nilai tersebut, penurunan WCSS cenderung melandai. Kondisi ini menunjukkan bahwa penambahan jumlah klaster di atas empat tidak memberikan peningkatan kualitas klaster yang berarti, melainkan hanya menambah kompleksitas model.

Dengan demikian, titik siku *(elbow point)* berada pada nilai K = 4, sehingga jumlah klaster optimal yang digunakan dalam penelitian ini adalah sebanyak empat klaster.

D. Hasil Penerapan Algoritma *K-Means*

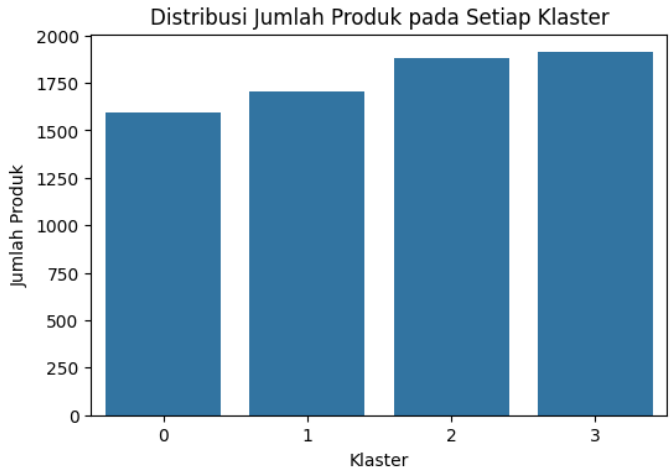
Setelah jumlah klaster optimal ditentukan, algoritma *K-Means* diterapkan pada data yang telah melalui tahap pra-pemrosesan. Proses ini menghasilkan pengelompokan data produk fashion ke dalam empat klaster berdasarkan tingkat kemiripan karakteristiknya.

Hasil proses *clustering* menghasilkan distribusi jumlah data pada masing-masing klaster sebagaimana ditunjukkan pada Tabel 4.

Tabel 4. Distribusi Jumlah Data pada Setiap Klaster

Klaster	Jumlah Data
Klaster 0	1.594
Klaster 1	1.708
Klaster 2	1.878
Klaster 3	1.911
Total	**7.091**

Untuk memperjelas distribusi data pada setiap klaster, digunakan visualisasi grafik batang sebagaimana ditunjukkan pada Gambar 2.



Gambar 2. Distribusi Jumlah Data pada Setiap Klaster

Berdasarkan Gambar 2, terlihat bahwa jumlah data pada setiap klaster relatif seimbang. Hal ini menunjukkan bahwa hasil *clustering* tidak terfokus pada satu klaster tertentu dan mencerminkan pembagian data yang stabil.

E. Analisis Karakteristik Klaster dan Pola Gaya Produk *Fashion*

Analisis karakteristik klaster dilakukan dengan mengidentifikasi atribut dominan pada masing-masing klaster. Atribut yang dianalisis meliputi *usage, articleType*, dan *baseColour*, karena atribut-atribut tersebut merupakan indikator utama dalam menggambarkan pola gaya produk fashion.

1. Klaster 0

Klaster 0 terdiri dari 1.594 data dan didominasi oleh produk dengan tujuan penggunaan casual. Jenis produk yang banyak muncul pada klaster ini antara lain pakaian sehari-hari seperti *T-shirt* dan *topwear*. Warna yang dominan adalah warna netral dan gelap seperti hitam dan biru.

Karakteristik ini menunjukkan bahwa klaster 0 merepresentasikan produk *fashion* untuk penggunaan sehari-hari dengan gaya sederhana dan fungsional.

2. Klaster 1

Klaster 1 terdiri dari 1.708 data dan memiliki karakteristik penggunaan formal dan semi-formal. Produk dalam klaster ini didominasi oleh jenis pakaian yang digunakan untuk kegiatan resmi maupun pekerjaan.

Klaster ini mencerminkan kelompok produk *fashion* yang ditujukan bagi konsumen dengan preferensi gaya formal dan tingkat kerapihan yang lebih tinggi.

3. Klaster 2

Klaster 2 terdiri dari 1.878 data dan didominasi oleh produk dengan penggunaan casual dengan variasi warna yang lebih beragam. Jenis produk pada klaster ini menunjukkan karakteristik fashion kasual modern.

Klaster ini merepresentasikan produk fashion dengan tingkat fleksibilitas penggunaan yang tinggi.

4. Klaster 3

Klaster 3 merupakan klaster dengan jumlah data terbanyak, yaitu 1.911 data. Klaster ini didominasi oleh produk dengan penggunaan casual dan *sportswear*.

Karakteristik tersebut menunjukkan bahwa klaster 3 merepresentasikan produk fashion yang digunakan untuk aktivitas santai dan olahraga dengan penekanan pada kenyamanan dan fungsi.

F. Identifikasi Preferensi Konsumen Berdasarkan Hasil *Clustering*

Berdasarkan karakteristik masing-masing klaster, preferensi konsumen dalam penelitian ini dapat diidentifikasi secara tidak langsung melalui pola gaya produk *fashion* yang terbentuk. Setiap klaster merepresentasikan kelompok produk dengan kesamaan karakteristik yang mencerminkan kecenderungan preferensi konsumen tertentu.

Klaster 0 dan Klaster 2 menunjukkan preferensi konsumen terhadap produk fashion kasual, dengan perbedaan pada jenis produk dan variasi warna. Klaster 1 mencerminkan preferensi konsumen terhadap produk fashion formal dan semi-formal, sedangkan Klaster 3 merepresentasikan preferensi konsumen terhadap produk sportswear dan aktivitas santai.

Hasil ini menunjukkan bahwa metode *K-Means Clustering* mampu mengelompokkan produk *fashion* berdasarkan pola gaya secara efektif, sehingga preferensi konsumen dapat diidentifikasi melalui hasil pengelompokan tersebut. Dengan demikian, hipotesis penelitian yang menyatakan bahwa metode *K-Means Clustering* dapat digunakan untuk mengidentifikasi preferensi konsumen dapat diterima.

V. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan mengenai penerapan *data mining* dengan metode *K-Means Clustering* untuk analisis pola gaya pada produk *fashion* dalam mengidentifikasi preferensi konsumen, dapat disimpulkan:

1. *fashion* yang telah melalui tahap pra-pemrosesan, meliputi pembersihan data, transformasi data kategorikal menggunakan label encoding, serta normalisasi data. Proses ini menghasilkan data yang siap dianalisis dan sesuai dengan kebutuhan algoritma *K-Means*.
2. Penentuan jumlah klaster optimal menggunakan metode *Elbow* menunjukkan bahwa jumlah klaster terbaik adalah empat klaster. Hal ini ditunjukkan oleh titik siku *(elbow point)* pada grafik *Within-Cluster Sum of Squares (WCSS),* di mana penurunan nilai WCSS mulai melandai setelah nilai K = 4.
3. Hasil penerapan *K-Means Clustering* mampu mengelompokkan produk *fashion* ke dalam empat klaster dengan karakteristik yang berbeda, yang ditentukan berdasarkan atribut dominan seperti *usage, articleType*, dan *baseColour*. Setiap klaster merepresentasikan pola gaya produk fashion yang berbeda, yaitu gaya kasual sehari-hari, gaya formal dan semi-formal, gaya kasual modern dengan variasi warna, serta gaya *sportswear* dan aktivitas santai.
4. Preferensi konsumen dapat diidentifikasi secara tidak langsung melalui hasil *clustering*, di mana setiap klaster mencerminkan kecenderungan konsumen terhadap jenis gaya

produk fashion tertentu. Konsumen dengan preferensi kasual, formal, maupun *sportswear* dapat dikenali berdasarkan karakteristik produk yang tergabung dalam masing-masing klaster.

Hasil penelitian ini membuktikan bahwa metode *K-Means Clustering* efektif digunakan untuk menganalisis pola gaya dan mengidentifikasi preferensi konsumen pada produk fashion digital. Informasi yang dihasilkan dari proses clustering dapat dimanfaatkan sebagai dasar dalam pengambilan keputusan strategis, khususnya dalam pengembangan produk, segmentasi pasar, dan penyusunan strategi pemasaran pada industri *fashion* digital.

DAFTAR PUSTAKA

Babu, M. M., & Arunraj, P. (2019). Fashion marketing management. New Delhi: New Century Publications.

Bahar, A., Pramono, B., & Sagala, L. H. S. (2016). Penentuan strategi penjualan menggunakan metode K-Means. SemanTIK, 2(2), 75–86.

Byjuith, A. (2023, October). Data preprocessing for machine learning. ResearchGate. Retrieved November 13, 2025, from https://www.researchgate.net/publication/375081512

Dalimunthe, R. F., & Putri, R. A. (2024). Data mining on women's clothing sales in marketplaces with the K-Means clustering algorithm. Indonesian Journal of Artificial Intelligence and Data Mining, 7(2), 476–484.

Kaggle. (2024). Fashion product images (small) dataset. Retrieved from https://www.kaggle.com/datasets/paramaggarwal/fashion-product-images-small

Kemenparekraf. (2022). Laporan tahunan ekonomi kreatif Indonesia.

Mardalius. (2018). Pemanfaatan Rapid Miner Studio 8.2 untuk pengelompokan data penjualan aksesoris menggunakan algoritma K-Means. Jurnal Teknologi dan Sistem Informasi, 4(2), 123–132.

Nazir, M. (2017). Metode penelitian. Bogor: Ghalia Indonesia.

Normah, N., Nurajizah, S., & Salbinda, A. (2021). Penerapan data mining metode K-Means clustering untuk analisa penjualan pada toko fashion hijab Banten. Jurnal Teknik Komputer AMIK BSI, 7(2).

Nurajizah, S. (2019). Analisa transaksi penjualan obat menggunakan algoritma Apriori. INOVTEK, 4(1), 35–44.

Pradnyana, G. A., & Agustini, K. (Edisi 1). Konsep dasar data mining (Modul 01, MSIM4403).

PublishJurnal. (2024, Februari 15). Penelitian kuantitatif eksploratori. https://publishjurnal.com/2024/02/15/penelitian-kuantitatif-eksploratori/ (Diakses 12 November 2025)

Siregar, M. H. (2018). Data mining klasterisasi penjualan alat bangunan menggunakan metode K-Means. Jurnal Teknologi dan Open Source, 1(2).

Statista. (2024). Global apparel market report. Retrieved from https://www.statista.com/topics/5091/apparel-market-worldwide/

Sugiyono. (2018). Metode penelitian kuantitatif, kualitatif, dan R&D. Alfabeta.

Sulianta, F., Ulfah, K., & Amalia, E. (2024). Revealing consumer preferences in the fashion industry using K-Means clustering. International Journal of Engineering Continuity, 3(2).

Yulianti, Y., Utami, D. Y., Hikmah, N., & Hasan, F. N. (2019). Penerapan data mining menggunakan algoritma K-Means untuk mengetahui minat customer di toko hijab. Jurnal Pilar Nusa Mandiri, 15(2), 241–246.