

Analisis Tren Pembelian Konsumen dalam E-commerce menggunakan Hierarchical Clustering

Puput Dwi Mandiri^{1*}, Joni Maulindar²,
^{1,2}Teknik Informatika

Universitas Duta Bangsa Surakarta

^{1*}202040332@mhs.udb.ac.id, ²joni_maulindar@udb.ac.id

Abstrak - E-commerce telah menjadi platform yang semakin populer untuk berbelanja secara online. Perusahaan e-commerce memiliki kebutuhan untuk memahami pola dan tren pembelian konsumen guna meningkatkan layanan, memperluas pasar, dan meningkatkan keuntungan. Dalam penelitian ini, dilakukan analisis tren pembelian konsumen dalam e-commerce menggunakan metode data mining. Metode data mining yang digunakan meliputi preprocessing data, pengelompokan, dan analisis asosiasi. Pertama, data pembelian konsumen yang tersedia diplatform e-commerce dikumpulkan dan dibersihkan untuk memastikan kualitas data. Kemudian, teknik pengelompokan digunakan untuk mengelompokkan konsumen berdasarkan pola pembelian mereka. Metode pengelompokan seperti K-means atau algoritma hierarkis dapat diterapkan dalam analisis ini. Setelah itu, dilakukan analisis asosiasi untuk mengidentifikasi pola pembelian yang sering muncul secara bersamaan. Melalui algoritma seperti Apriori atau FP-Growth, aturan asosiasi dapat digunakan untuk mengungkapkan hubungan antara produk yang dibeli bersamaan. Misalnya, dapat terungkap bahwa konsumen yang membeli produk A juga cenderung membeli produk B. Melalui analisis tren pembelian konsumen dalam e-commerce menggunakan data mining, perusahaan e-commerce dapat meningkatkan pemahaman mereka tentang preferensi konsumen dan pola pembelian yang terjadi. Hal ini membantu perusahaan mengoptimalkan strategi pemasaran, meningkatkan pengalaman konsumen, dan meningkatkan efisiensi operasional mereka.

Kata kunci : Tren pembelian konsumen, E-commerce, Data mining, Pola pembelian, Analisis asosiasi

Abstract - E-commerce has become an increasingly popular platform for shopping online. E-commerce companies have a need to understand consumer purchasing patterns and trends in order to improve services, expand markets, and increase profits. In this study, an analysis of consumer purchasing trends in e-commerce was carried out using the data mining method. Data mining methods used include data preprocessing, grouping, and association analysis. First, consumer purchase data available on e-commerce platforms is collected and cleaned to ensure data quality. Then, clustering techniques are used to group consumers based on their purchasing patterns. Clustering methods such as K-means or hierarchical algorithms can be applied in this analysis. After that, an association analysis is performed to identify purchasing patterns that often appear simultaneously. Through algorithms such as Apriori or FP-Growth, association rules can be used to reveal the relationship between products purchased together. For example, it can be revealed that consumers who buy product A also tend to buy product B. By analyzing consumer buying trends in e-commerce using data mining, e-commerce companies can improve their understanding of consumer preferences and purchasing patterns that occur. This helps companies optimize their marketing strategy, enhance the consumer experience, and increase their operational efficiency.

Keywords : Consumer buying trends, E-commerce, Data mining, Purchase patterns, Association analysis

I. PENDAHULUAN

Dalam era digital saat ini, e-commerce telah menjadi salah satu platform utama bagi konsumen untuk melakukan pembelian secara online. Dengan meningkatnya popularitas e-commerce,

perusahaan dalam industri ini berusaha untuk memanfaatkan data yang tersedia untuk memahami pola dan tren pembelian konsumen. Hal ini penting untuk meningkatkan kualitas layanan, memperluas pasar, dan meningkatkan keuntungan.

Dalam konteks ini, metode data mining memiliki peran yang krusial. Data mining adalah proses penggalian informasi berharga dan berguna dari dataset yang besar dan kompleks. Dalam kasus ini, data mining digunakan untuk menganalisis tren pembelian konsumen dalam e-commerce. Melalui teknik-teknik data mining, pola pembelian konsumen dapat diidentifikasi, hubungan antara produk dapat diungkap, dan wawasan berharga dapat diperoleh untuk mendukung pengambilan keputusan strategis perusahaan e-commerce.

Dalam penelitian ini, akan dilakukan analisis tren pembelian konsumen menggunakan metode data mining. Preprocessing data, pengelompokan, dan analisis asosiasi akan menjadi fokus utama untuk menggali wawasan dari data pembelian konsumen yang ada. Hasil dari analisis ini akan memberikan informasi yang berharga bagi perusahaan e-commerce dalam mengembangkan strategi pemasaran yang lebih efektif, meningkatkan pengalaman konsumen, dan mengoptimalkan operasional mereka.

II. METODOLOGI PENELITIAN

Analisis tren pembelian konsumen dalam e-commerce menggunakan metode clustering dalam data mining dapat membantu mengidentifikasi pola-pola pembelian yang berbeda dalam data pelanggan. Clustering adalah salah satu metode yang berguna untuk mengelompokkan data berdasarkan kesamaan fitur atau atribut tertentu. Berikut adalah langkah-langkah umum yang dapat Anda ikuti dalam melakukan analisis tren pembelian konsumen menggunakan clustering:

1. **Persiapan Data:** Pertama, pastikan Anda memiliki dataset yang mencakup informasi pembelian konsumen, seperti ID pembelian, produk yang dibeli, jumlah pembelian, waktu pembelian, dan atribut lain yang relevan. Bersihkan dan format data sesuai kebutuhan.

2. **Pemilihan Fitur:** Pilih fitur atau atribut yang akan digunakan dalam analisis clustering. Misalnya, Anda dapat mempertimbangkan produk yang dibeli, jumlah pembelian, atau kategori produk sebagai fitur yang relevan.

3. **Pemilihan Metode Clustering:** Ada beberapa

metode clustering yang dapat Anda gunakan, seperti K-means, Hierarchical Clustering, atau DBSCAN. Pilih metode yang paling sesuai dengan data dan tujuan analisis Anda. Misalnya, jika Anda ingin mengelompokkan pembelian berdasarkan kategori produk, K-means dapat menjadi pilihan yang baik.

4. **Penerapan Clustering:** Terapkan metode clustering yang dipilih pada dataset Anda. Metode clustering akan mengelompokkan data ke dalam beberapa kelompok berdasarkan kesamaan fitur. Setiap kelompok akan mewakili pola pembelian yang berbeda.

5. **Evaluasi Hasil Clustering:** Evaluasi hasil clustering untuk memastikan kelompok-kelompok yang dihasilkan memiliki kesamaan internal yang tinggi dan perbedaan antar kelompok yang signifikan. Anda dapat menggunakan metrik evaluasi seperti Coherence Index, Silhouette Score, atau Elbow Method untuk mengevaluasi kualitas clustering.

6. **Interpretasi dan Analisis:** Setelah mendapatkan kelompok-kelompok pembelian, analisis pola-pola pembelian dalam setiap kelompok. Identifikasi karakteristik atau preferensi pembelian yang unik dalam setiap kelompok. Misalnya, apakah ada kelompok yang cenderung membeli produk-produk diskon, atau kelompok yang lebih memilih produk-produk premium.

7. **Implementasi Hasil:** Gunakan hasil analisis clustering untuk meningkatkan strategi pemasaran e-commerce Anda. Anda dapat mengirimkan tawaran khusus atau rekomendasi produk yang relevan kepada setiap kelompok pembeli. Hal ini dapat membantu meningkatkan pengalaman konsumen dan mengoptimalkan penjualan.

Penting untuk diingat bahwa analisis clustering adalah proses iteratif dan dapat melibatkan beberapa percobaan dan penyesuaian. Selain itu, kualitas hasil clustering sangat bergantung pada kualitas data yang digunakan dan pemilihan atribut yang tepat.

III. HASIL DAN PEMBAHASAN

Berdasarkan olah data dengan metode Hierarchical Clustering, bahwa data telah dikelompokkan menjadi 3 kluster (Cluster 1, Cluster 2, dan Cluster 3). Setiap kluster memiliki karakteristik yang berbeda dalam hal harga dan

rating produk. Berikut adalah beberapa informasi yang dapat disimpulkan dari ringkasan statistik tersebut:

D. Dataset

Dataset yang digunakan ada 9 kategori untuk data yang digunakan dapat dilihat pada gambar dibawah ini.

ID	Produk	Harga	Kategori	Deskripsi	Stok	Toko	Rating	Variasi
1	Sepatu	\$50	Aksesori	Sepatu dengan kenyamanan maksimal	8	Store B	4.2	Hitam, Cokelat
2	Kamera	\$500	Elektronik	Kamera dengan resolusi tinggi	5	Store C	4.8	DSLR, Mirrorless
3	Lipstik	\$15	Kosmetik	Lipstik dengan warna-warna memamk	12	Store D	4.6	Merah, Pink, Nude
4	Tas	\$30	Aksesori	Tas dengan kapasitas besar	10	Store E	4.3	Hitam, Cokelat
5	Smartphone	\$800	Elektronik	Smartphone dengan kinerja yang cepa	3	Store F	4.1	Hitam, Silver
61	Foundation	\$25	Kosmetik	Foundation dengan formula tahan lama	8	Store BG	4.3	Light Beige, Natural Beige
62	Handphone	\$600	Elektronik	Handphone dengan kamera berku	6	Store BH	4.5	Hitam, Biru
63	Shampoo	\$10	Kosmetik	Shampoo dengan aroma segar	12	Store BI	4.6	Jeruk, Apel
64	Hoodie	\$50	Pakaian	Hoodie dengan desain modern	10	Store BJ	4.3	Abu-abu, Biru
65	Wireless Mouse	\$20	Aksesori	Wireless mouse dengan konektivitas 2.4GHz	8	Store BK	4.2	Hitam, Putih
66	Smart TV	\$800	Elektronik	Smart TV dengan resolusi 4K	5	Store BL	4.7	50 inch, 55 inch

Gambar 14 Dataset Ecommerce

Setelah data siap digunakan, pilih metode Clustering yang kana digunakan ada beberpad metode yaitu K-means, Hierarchical Clustering, atau DBSCAN. Tapi untuk dataset konsumen E-commerce ini kita menggunakan metode Hierarchical Clustering.

E. Metode Hierarchical Clustering

Metode Hierarchical Clustering adalah salah satu teknik clustering yang digunakan untuk mengelompokkan data ke dalam hierarki kluster. Metode ini membangun hierarki kluster dengan mempertimbangkan jarak antara data yang ada. Terdapat dua pendekatan utama dalam Hierarchical Clustering: Agglomerative (penggabungan) dan Divisive (pemisahan).

1. Agglomerative Hierarchical Clustering: Pendekatan ini dimulai dengan setiap data sebagai kluster individu, kemudian secara iteratif menggabungkan kluster berdasarkan jarak antara mereka.

2. Divisive Hierarchical Clustering: Pendekatan ini dimulai dengan semua data dalam satu kluster dan secara iteratif memisahkan kluster menjadi sub-kluster yang lebih kecil berdasarkan jarak antara data

Dalam Hierarchical Clustering, hasilnya dapat direpresentasikan dalam bentuk dendrogram, yang menunjukkan hierarki kluster dan jarak antara kluster. Pemotongan dendrogram dapat digunakan untuk menentukan jumlah kluster yang optimal. Keuntungan dari Hierarchical Clustering adalah kemampuannya untuk menghasilkan struktur hierarkis kluster yang intuitif dan visual. Selain itu,

tidak perlu menentukan jumlah kluster sebelumnya dan dapat digunakan untuk dataset berukuran besar. Namun, Hierarchical Clustering cenderung lebih lambat dibandingkan dengan metode clustering lainnya dan dapat terpengaruh oleh noise atau pencilan dalam data. Metode Hierarchical Clustering dapat diimplementasikan dengan menggunakan berbagai bahasa pemrograman, termasuk Python dengan menggunakan library seperti SciPy, scikit-learn, atau hcluster.

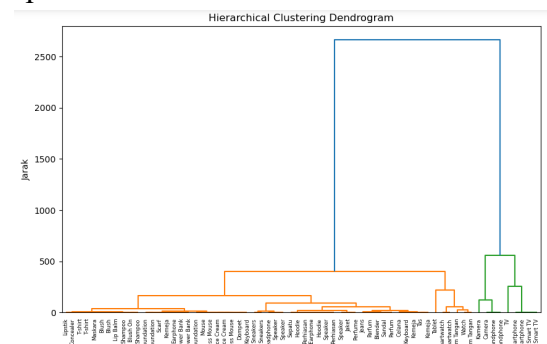
F. Pseudocode Program

Berikut adalah pseudocode dari proses pengolahan dataset konsumen E-commerce :

- 1.Import library yang diperlukan (seperti pandas, matplotlib)
- 2.Baca dataset dari file CSV
- 3.Lakukan pembersihan dan preprocessing data (jika diperlukan)
- 4.Lakukan Hierarchical Clustering
- 5.Tambahkan kolom kluster ke DataFrame
- 6.Tampilkan ringkasan statistik untuk setiap kluster:
 - 6.1.Kelompokkan data berdasarkan kolom kluster menggunakan groupby
 - 6.2.Hitung ringkasan statistik menggunakan describe()
 - 6.3.Tampilkan ringkasan statistik

G. Diagram Dendrogram

Diagram dendrogram adalah alat visual yang digunakan dalam analisis Hierarchical Clustering untuk menggambarkan hubungan hierarkis antara kelompok atau objek data. Dendrogram memberikan representasi visual yang intuitif tentang bagaimana objek-objek dalam dataset dikelompokkan dan bagaimana kelompok-kelompok tersebut terbentuk.



Gambar 15 Digram Dendrogram

Dalam dendrogram, setiap objek data direpresentasikan sebagai titik atau simpul, dan jarak antara titik-titik tersebut menggambarkan tingkat kesamaan atau jarak antara objek-objek tersebut. Titik-titik yang lebih dekat secara vertikal menunjukkan objek-objek yang lebih mirip satu sama lain, sedangkan titik-titik yang lebih jauh secara vertikal menunjukkan objek-objek yang lebih berbeda satu sama lain.

H. Hasil Pembagian Cluster

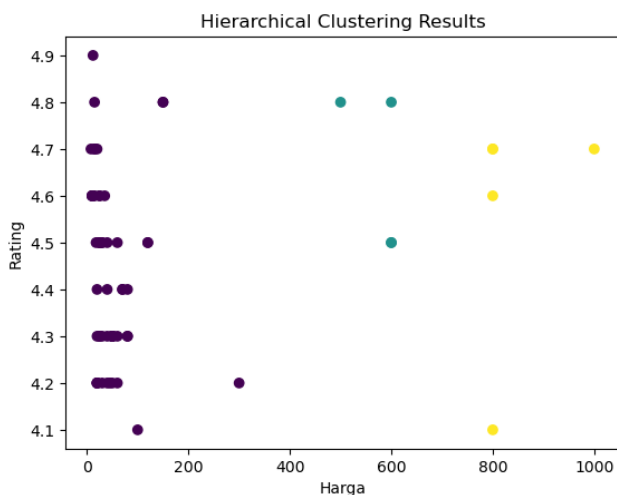
Cluster	ID							
	count	mean	std	min	25%	50%	75%	max
1	57.0	33.45614	18.655782	1.0	18.00	33.0	50.00	65.0
2	4.0	30.000000	26.191602	2.0	13.25	28.0	44.75	62.0
3	5.0	36.800000	23.983328	5.0	21.00	43.0	49.00	66.0

Cluster	Stok		Rating	
	mean	std	min	25%
1	46.964912	10.0	15.0	57.0
2	575.000000	6.0	6.0	4.0
3	840.000000	5.0	5.0	5.0

Cluster	Rating		
	50%	75%	max
1	4.40	4.6	4.9
2	4.65	4.8	4.8
3	4.70	4.7	4.7

[3 rows x 32 columns]

Gambar 16 Pembagian Cluster



Gambar 17 Hasil Clustering Results

Berdasarkan ringkasan statistik yang Anda berikan, terlihat bahwa data telah dikelompokkan menjadi 3 kluster (Cluster 1, Cluster 2, dan Cluster 3) berdasarkan hasil Hierarchical Clustering. Setiap kluster memiliki karakteristik yang berbeda dalam hal

harga dan rating produk.

Semua tautan *hypertext* dan bagian *bookmark* akan dihapus. Jika paper perlu merujuk ke alamat email atau URL di artikel, maka alamat atau URL lengkap harus ditulis dengan font biasa.

IV. KESIMPULAN

Berdasarkan pembagian kluster yang telah dilakukan dengan menggunakan Hierarchical Clustering pada dataset yang Anda berikan, berikut ini adalah beberapa kesimpulan yang dapat diambil:

1. Terdapat 3 kluster yang dihasilkan: Cluster 1, Cluster 2, dan Cluster 3.
2. Cluster 1 memiliki jumlah data yang paling banyak (57 data) dibandingkan dengan kluster lainnya. Cluster ini memiliki rentang harga yang luas (dari 1 hingga 65) dengan rata-rata harga sekitar 33.45614. Stok produk dalam kluster ini berkisar antara 3 hingga 15, dengan rata-rata sekitar 8.192982. Rating produk pada kluster ini memiliki rentang 4.1 hingga 4.9, dengan rata-rata rating sekitar 4.435088.

3. Cluster 2 memiliki jumlah data yang lebih sedikit (4 data) dibandingkan dengan kluster lainnya. Cluster ini memiliki rentang harga yang lebih rendah (dari 2 hingga 62) dengan rata-rata harga sekitar 30.00000. Stok produk dalam kluster ini hanya memiliki nilai 6. Rating produk pada kluster ini memiliki rentang 4.5 hingga 4.8, dengan rata-rata rating sekitar 4.650000.

4. Cluster 3 memiliki jumlah data yang sedikit (5 data). Cluster ini memiliki rentang harga yang lebih tinggi (dari 5 hingga 66) dengan rata-rata harga sekitar 36.80000. Stok produk dalam kluster ini hanya memiliki nilai 5. Rating produk pada kluster ini memiliki rentang 4.1 hingga 4.7, dengan rata-rata rating sekitar 4.560000.

Kesimpulan yang dapat diambil dari hasil pembagian kluster ini adalah terdapat perbedaan karakteristik antara kluster-kluster tersebut. Cluster 1 memiliki jumlah data yang paling banyak dengan rentang harga yang lebih luas, sedangkan Cluster 2 memiliki harga yang lebih tinggi namun jumlah data yang lebih sedikit.

Cluster 3 memiliki harga yang sedikit lebih tinggi dibandingkan dengan Cluster 1 dan rentang rating yang lebih kecil. Kesimpulan ini dapat membantu pemahaman lebih lanjut tentang kelompok-kelompok produk yang ada dalam dataset dan memungkinkan pengambilan keputusan yang lebih terarah dalam hal strategi pemasaran, pengelolaan stok, atau peningkatan kualitas produk berdasarkan karakteristik kluster yang berbeda.

REFERENSI

- [1] Wulandari, N. P., & Erti Pratiwi, E. (2017). Penerapan Metode Hierarchical Clustering pada Data Mining untuk Analisis Klasifikasi Tingkat Kesuksesan Mahasiswa. *Jurnal Informatika Mulawarman*, 12(1), 48-54.
- [2] Parman, S., & Kurniawan, R. (2019). Aplikasi Metode Hierarchical Clustering dalam Data Mining untuk Penentuan Kategori Kualitas Pendidikan Tinggi. *Jurnal Ilmiah Teknologi Informasi Asia*, 13(1), 17-23.
- [3] Fajriah, A., & Hidayatullah, N. (2018). Pengklasteran Menggunakan Metode Hierarchical Clustering pada Data Penjualan Produk UKM. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 2(11), 5459-5466.
- [4] Agustina, L., & Fitriyani, N. L. (2017). Analisis Klasterisasi Data Kinerja Dosen Menggunakan Algoritma Hierarchical Clustering. *Jurnal Informatika UPGRIS*, 3(1), 45-54.
- [5] Rahayu, E., & Hidayatullah, N. (2020). Aplikasi Metode Hierarchical Clustering dalam Pengelompokan Produk di Toko Online. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 4(1), 27-34.
- [6] Adiwijaya, & Atmadja, A. S. (2014). *Data Mining: Pemahaman, Implementasi, dan Aplikasi Menggunakan Python*. Penerbit Informatika.
- [7] Fadlisyah, A. (2016). *Data Mining: Algoritma, Implementasi, dan Aplikasi Menggunakan R*. Penerbit Informatika.
- [8] Rahmadani, D., & Wulandari, D. A. (2018). *Data Mining: Konsep dan Aplikasi dengan Software R*. Penerbit Andi.
- [9] Abdullah, N. A., & Mursanto, P. (2014). *Data Mining: Metode dan Implementasi dengan RapidMiner*. Penerbit Informatika.
- [10] Rostianingsih, S., & Sitompul, O. S. (2018). *Data Mining: Konsep, Algoritma, dan Implementasi dalam R*. Penerbit PT Gramedia Pustaka Utama.
- [11] Wijaya, A. F. (2017). *Data Mining untuk Keunggulan Kompetitif Perusahaan*. Penerbit Prenadamedia Group.