

PENERAPAN ALGORITMA K-MEANS UNTUK OPTIMASI MODEL CLUSTERING DATA SUPPLIER DI APLIKASI SHOPEE

Moch Maliq Dinar Danureksa, Rudi kurnawan, Yudhist ira Arie Wijaya

STMIK IKMI Cirebon, Teknik Informatika

Jl. Perjuangan No.10B, Karyamulya, Kec. Kesambi, Kota Cirebon, Jawa Barat 45135, Indonesia

reksakusnadi6@gmail.com

ABSTRAK

Clustering adalah teknik analisis data yang digunakan untuk mengelompokkan data berdasarkan kesamaan karakteristik. Dalam konteks e-commerce, khususnya pada platform Supplier Shopee, teknik ini dapat dimanfaatkan untuk meningkatkan efisiensi pengambilan keputusan strategis. Penelitian ini berfokus pada penerapan algoritma K-Means untuk membangun model clustering yang optimal dalam mengidentifikasi pola transaksi supplier di platform tersebut. Metode analisis yang digunakan adalah Knowledge Discovery in Database (KDD), yang terdiri dari lima tahapan utama: seleksi data, pra-proses data, transformasi, penerapan algoritma K-Means, dan evaluasi hasil. Eksperimen dilakukan dengan menguji beberapa nilai K untuk menemukan hasil cluster terbaik. Evaluasi cluster dilakukan menggunakan Davies-Bouldin Index (DBI), sebuah metrik yang mengukur tingkat pemisahan antar cluster dan keseragaman dalam cluster. Hasil penelitian menunjukkan bahwa nilai K terbaik adalah 2, dengan nilai DBI sebesar 0,213. Ini menunjukkan cluster yang dihasilkan memiliki pemisahan yang baik dan struktur internal yang konsisten. Cluster 0 merepresentasikan supplier dengan jumlah transaksi sedikit, cluster 1 mencakup supplier dengan transaksi tinggi, dan cluster 2 melibatkan supplier dengan transaksi sedang. Hasil ini membuktikan bahwa algoritma K-Means dapat secara efektif mengidentifikasi pola transaksi. Informasi ini berguna untuk merumuskan strategi pemasaran, seperti insentif bagi supplier bertransaksi rendah atau penguatan hubungan dengan supplier unggul. Penelitian ini menegaskan pentingnya algoritma K-Means dan KDD dalam analisis data besar untuk mendukung pengambilan keputusan berbasis data di e-commerce.

Kata Kunci: K-Means, clustering, Shopee, penjualan produk, Davies-Bouldin Index

1. PENDAHULUAN

Dalam beberapa tahun terakhir, perkembangan teknologi informasi dan komunikasi telah membawa perubahan signifikan di berbagai sektor, termasuk bisnis dan perdagangan elektronik[1].

Dalam dunia perdagangan online, pengelolaan supplier yang efektif menjadi salah satu faktor utama dalam keberhasilan sebuah toko. Toko Rey Kuningan, yang memanfaatkan platform Shopee, menghadapi tantangan dalam mengelompokkan supplier secara tepat berdasarkan karakteristik dan kontribusinya terhadap penjualan. Permasalahan yang sering muncul adalah bagaimana mengelompokkan supplier dengan cara yang relevan agar dapat mendukung strategi bisnis, seperti optimalisasi stok, penentuan prioritas pemasok, dan peningkatan kualitas layanan. Penelitian sebelumnya menunjukkan bahwa model clustering yang tidak optimal dapat mengurangi kualitas rekomendasi dan efisiensi pengelolaan pemasok[2].

Penelitian ini bertujuan untuk meningkatkan model Algoritma K-Means Untuk Meningkatkan Model Clustering Data Pada Supplier Toko Di Aplikasi Shopee. Fokus utama penelitian adalah mengatasi permasalahan nilai k yang tidak optimal dalam algoritma K-Means, yang menjadi salah satu tantangan utama dalam pengelompokan data supplier. Dengan memberikan pendekatan baru dalam parameterisasi algoritma K-Means yang lebih adaptif terhadap dataset e-commerce, penelitian ini berkontribusi dalam mengisi kesenjangan penelitian

terkait efisiensi dan akurasi clustering. Secara praktis, hasil penelitian diharapkan dapat meningkatkan manajemen pemasok dan kualitas rekomendasi pada platform e-commerce, sehingga mendukung pengambilan keputusan strategis oleh pengelola bisnis[3].

Penelitian ini memberikan solusi teknis untuk optimasi algoritma K-Means yang dapat diaplikasikan pada berbagai aspek operasional dan strategis, khususnya untuk mendukung pengelolaan data supplier di Toko pada platform Shopee. Hasil clustering dapat digunakan untuk segmentasi supplier berdasarkan performa, seperti volume penjualan, frekuensi transaksi, atau rating pelanggan, sehingga mempermudah penerapan strategi yang spesifik, seperti pemberian insentif pada supplier terbaik atau pembinaan pada supplier berkinerja rendah.

2. TINJAUAN PUSTAKA

Metode literature review yang digunakan dalam penelitian ini adalah dengan cara mereview pada artikel-artikel yang terkait dengan data mining. Langkah-langkah dari literature review meliputi 4 tahapan, yaitu (1) formulasi permasalahan, (2) pencarian literature, (3) evaluasi data, (4) analisis dan interpretasi.

Penelitian terdahulu membahas mengenai perbandingan algoritma K-Mean dan K-Medoids untuk mengelompokkan daerah rawan tanah longsor di Jawa Barat yang berjudul "Perbandingan Algoritma K-

Mean dan K-Medoids untuk Pengelompokan Daerah Rawan Tanah Longsor di Provinsi Jawa Barat". Penelitian ini menggunakan algoritma K-Mean dan K-Medoids untuk mengelompokkan data kejadian tanah longsor di Provinsi Jawa Barat berdasarkan wilayah yang terdampak. Dataset yang digunakan terdiri dari 24 data hasil pembersihan dari total 609 kejadian tanah longsor tahun 2019 yang diperoleh dari situs Open Data Jawa Barat [4].

Penelitian terdahulu mengusulkan modifikasi pada algoritma K-Mean dengan menambahkan metode optimasi titik awal centroid menggunakan algoritma Pillar. Dalam K-Mean, pemilihan titik awal centroid sangat mempengaruhi hasil akhir pengelompokan, dan penelitian ini berfokus pada peningkatan akurasi clustering dengan menggunakan algoritma Pillar untuk meminimalkan kesalahan dalam pemilihan centroid.[5]

Penelitian terdahulu mengkaji penerapan algoritma K-Mean untuk mengelompokkan judul skripsi dan jurnal penelitian di Fakultas Teknik Informatika dan Komputer (FTIK) Universitas Prima Indonesia (Unpri). Algoritma K-Mean digunakan untuk menemukan pola dalam pemilihan topik-topik penelitian dan memetakan kecenderungan para peneliti dalam bidang teknologi informasi.[6].

Penelitian terdahulu mengaplikasikan metode clustering K-Medoids untuk mengelompokkan toko-toko di platform e-commerce Shopee berdasarkan reputasi dan ulasan pelanggan. Dengan menggunakan data transaksi dan ulasan dari pelanggan, penelitian ini bertujuan untuk membantu pembeli dalam memilih toko dengan reputasi terbaik.[7].

Penelitian terdahulu mengkaji penerapan algoritma K-Mean untuk mengelompokkan wilayah prioritas vaksin COVID-19. Dengan menggunakan data sosial-ekonomi dan tingkat penularan COVID-19, algoritma ini digunakan untuk mengidentifikasi daerah-daerah yang membutuhkan prioritas vaksin.[8]

3. METODE PENELITIAN

Dalam penelitian ini, penekanan diberikan pada penggunaan algoritma K-Means. Penelitian ini menggunakan metode kuantitatif dengan pendekatan eksperimen, yang bertujuan untuk Meningkatkan Model Clustering Data Pada Supplier Toko Di Aplikasi Shopee menggunakan algoritma K-Means.

3.1. Selection

Tahap ini melibatkan proses pemilihan data dari sumber utama, seperti *data warehouse*, yang relevan dengan tujuan analisis. Data yang dipilih disebut sebagai *target data*, yang akan digunakan untuk tahap selanjutnya. Proses ini penting untuk memastikan bahwa data yang digunakan memiliki relevansi dan kualitas yang cukup untuk mendukung analisis.

3.2. Preprocessing

Pada tahap ini, *target data* yang telah dipilih diproses untuk menghilangkan data yang tidak lengkap, tidak konsisten, atau redundan. Proses ini bertujuan untuk meningkatkan kualitas data dan mempersiapkannya agar siap digunakan pada tahap analisis berikutnya. Data hasil proses ini disebut sebagai *pre-processed data*.

3.3. Transformation

Data yang telah dipra-pemrosesan selanjutnya ditransformasikan menjadi format yang lebih sesuai untuk analisis, seperti tabel atau matriks. Transformasi ini dapat mencakup normalisasi, agregasi, atau pembuatan atribut baru yang lebih informatif. Data hasil transformasi ini disebut sebagai *data transformation*, yang siap digunakan dalam proses *data mining*.

3.4. Data Mining

Tahap inti dari proses KDD ini melibatkan penerapan algoritma tertentu untuk menemukan pola atau informasi yang tersembunyi dalam data. Pola-pola ini bisa berupa hubungan, tren, atau klasifikasi yang relevan dengan tujuan penelitian atau bisnis. Hasil tahap ini berupa pola (*pattern*) yang menjelaskan karakteristik data.

3.5. Evaluation

Pola yang ditemukan melalui *data mining* kemudian dievaluasi untuk memastikan relevansi dan validitasnya. Evaluasi dilakukan dengan membandingkan hasil analisis dengan tujuan awal atau dengan menggunakan metrik tertentu untuk menilai kualitas pola yang ditemukan.

4. HASIL DAN PEMBAHASAN

4.1. Hasil Penelitian

Dalam penelitian ini, telah dilakukan serangkaian analisis untuk mengkaji Algoritma K-Mean Untuk Meningkatkan Model Clustering Data Pada Supplier Toko di Aplikasi Shoppe dengan menggunakan metode yang telah dirancang secara sistematis. Berdasarkan data yang diperoleh, beberapa temuan utama berhasil diidentifikasi dan dianalisis guna menjawab rumusan masalah yang telah dirumuskan sebelumnya

4.2. Selection

Proses seleksi yang ditunjukkan dalam gambar "Read Excel" adalah bagian integral dari sistem pengolahan data. Dalam konteks ini, seleksi dilakukan pada file Excel yang dimasukkan sebagai input untuk memastikan bahwa hanya data yang relevan diproses dan diteruskan ke tahap berikutnya. Narasi seleksi berdasarkan informasi gambar dijelaskan sebagai berikut:



Gambar 1. Operator Membaca Excel

Operator *Read Excel* dalam penelitian digunakan untuk baca data dari *file Excel* (.xlsx atau .xls) ke dalam perangkat lunak analisis data atau data mining. Fungsi utamanya adalah membaca dan mengekstrak informasi dari tabel Excel, baik seluruh data maupun kolom dan baris tertentu yang relevan. Operator ini memastikan data dapat diakses dalam format yang sesuai untuk proses analisis lebih lanjut, seperti eksplorasi, pembersihan, atau transformasi data

Tabel 1 menjelaskan parameter yang digunakan dalam proses membaca file Excel untuk analisis data. File yang digunakan berjudul *6000+ List Supplier Terlaris Shopee dataset maliq.xlsx* dan dipilih sheet pertama berdasarkan nomor sheet (Sheet Number 1). Sel yang diimpor dari file ini ditentukan pada rentang sel A1, dengan metode pengkodean karakter

(encoding) yang disesuaikan secara otomatis oleh sistem. Parameter ini dirancang untuk memastikan data yang relevan dapat diakses secara akurat sesuai kebutuhan analisis.

Tabel 1. Parameter Membaca Excel

No	Parameter	Isi
1	Excel file	6000+ List Supplier Terlaris Shopee dataset maliq.xlsx
2	Sheet Selection	Sheet Number
3	Sheet Number	1
4	Imported cell Range	A 1
5	Encoding	System

Data yang digunakan dalam penelitian ini adalah data Primer yang di peroleh langsung dari sumber pertama. yang menyediakan dataset Clustering data terlaris pada Supplier toko rey kuningan di aplikasi shopee. Dataset ini terdiri dari 6.000 entri yang mencakup berbagai atribut Data ini dipilih karena kelengkapannya dan relevansinya dengan tujuan penelitian seperti pada Tabel 4.1.1.

Tabel 2. Hasil Preview Membaca Excel

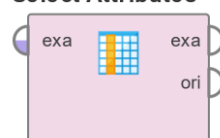
No	Link Product	Hyperlink	Nama produk	Link Toko	Hyperlink 2	% kenaikan penjualan	Review Bulan sebelum	Kenaikan 3	% kenaikan review
1	https://shopee.co.id/Promosi-BATERAI-JAM-TANGAN-SONY-626-LR-626-BATTERY-BATRAI-i.180705320.4879128419	Link Produk	Promosi BATERAI JAM TANGAN SONY 626 / LR 626 BATTERY BATRAI	https://shopee.co.id/cille520?categoryId=37&itemId=4879128419	Link Toko	376,19 %	103	766	743,69 %
2	https://shopee.co.id/Tas-Selempang-Pushop-Shoulder-Bag-Korea-CROSS-i.4818709.9912793796	Link Produk	Tas Selempang Pushop Shoulder Bag Korea CROSS	https://shopee.co.id/hotshop?categoryId=10217&itemId=9912793796	Link Toko	273,08 %	1.100	3.400	309,09 %
3	https://shopee.co.id/Milky-House-Kantong-ASI-120ml-dengan-Thermal-Sensor-i.374860314.6876983183	Link Produk	Milky House Kantong ASI 120ml dengan Thermal Sensor	https://shopee.co.id/partymurahonline.id?categoryId=45&itemId=6876983183	Link Toko	246,15 %	501	1.399	279,24 %
4	https://shopee.co.id/PROMO-Star-Grip-Kain-Raket-Badminton-i.186650205.5162274104	Link Produk	PROMO Star Grip Kain Raket Badminton	https://shopee.co.id/badmintonyuk?categoryId=41&itemId=5162274104	Link Toko	203,23 %	97	206	212,37 %
5	https://shopee.co.id/SETELAN-CHUPA-CHUP-TIE-DYE-TIE-DYE-SET-KAOS-VIRAL-	Link Produk	SETELAN CHUPA CHUP TIE DYE / TIE DYE SET /KAOS	https://shopee.co.id/supplierhijabfashion?categoryId=14535&itemId=8214010087	Link Toko	203,03 %	520	1.480	284,62 %

No	Link Product	Hyperlink	Nama produk	Link Toko	Hyperlink 2	% kenaikan penjualan	Review Bulan sebelumnya	Kenaikan 3	% kenaikan review
	KAOS-ANAK-TIE-DYE-SHIRT-i.67693195.8214010087		VIRAL/KAO S ANAK/TIE DYE SHIRT						
...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...
6403	https://shopee.co.id/Voucher-Kuota-Data-Smartfren-Unlimited-7GB-7-Hari-i.106958964.3250356685	Link Produk	Voucher Kuota Data Smartfren Unlimited 7GB, 7 Hari	https://shopee.co.id/3storesbyjatim?categoryId=2160&itemId=3250356685	Link Toko	-	-	-	-
6404	https://shopee.co.id/SPOTIFY-1-2-3BULAN-PROMO-TERMURAH!!-DIJAMIN-GARANSI!!-i.151414835.3863242686	Link Produk	SPOTIFY 1-2-3BULAN PROMO TERMURAH !! DIJAMIN GARANSI!!	https://shopee.co.id/spotifydhika1?categoryId=2160&itemId=3863242686	Link Toko	-	-	-	-
6406	https://shopee.co.id/pulsa-5000-10000-15000-20000-25000-30000-bebas-semua-operator-i.8007186.73398793	Link Produk	pulsa 5000 10000 15000 20000 25000 30000 bebas semua operator	https://shopee.co.id/stivee88?categoryId=2160&itemId=733987943	Link Toko	-	-	-	-

Dataset yang diunggah berisi informasi rinci mengenai produk yang dijual di Shopee, termasuk URL produk (LinkProduk), nama produk (NamaProduk), dan informasi toko (NamaToko, LinkToko). Dataset ini mencakup detail harga seperti Harga, HargaAtas, dan estimasi Omset Minimum, serta data penjualan seperti JumlahTerjual, Penjualan Bulan Sebelumnya, dan kenaikan dalam penjualan (Kenaikan dan % Kenaikan Penjualan). Selain itu, terdapat informasi kategori produk (Kategori1, Kategori2, Kategori3), lokasi pengiriman (KotaPengiriman), serta data ulasan seperti Review Bulan Sebelumnya, Kenaikan3, dan % Kenaikan Review.

Melalui analisis data ini, diharapkan dapat memberikan kontribusi dalam pengambilan keputusan bagi penjual dan meningkatkan pengalaman berbelanja bagi konsumen di Shopee.

Select Attributes



Gambar 2. operator Select Attributes

Operator Select Attributes dalam penelitian ini berfungsi untuk menyaring atribut-atribut yang relevan dari dataset yang akan digunakan dalam proses klusterisasi dengan algoritma K-Mean. Tahap ini memastikan bahwa hanya atribut yang signifikan terhadap tujuan penelitian, seperti % Kenaikan Review, Harga, Jumlah Terjual, Kategori Produk (Kategori1), Kota Pengiriman, Omset Minimum, Penjualan Bulan Sebelumnya, dan Review Bulan Sebelumnya.

Tabel 3. Parameter Select Attributes

No	Parameter	Isi
1	Type	Include attributes
2	Atribut filter type	A subset
326	Attribute :	% Kenaikan Review
25		Harga
		JumlahTerjual
		Kategori1
		KotaPengiriman
		Omset Minimum
		Pejualan Bulan Sebelumnya
		Review Bulan Sebelumnya
		Kenaikan penjualan

Tabel parameter di atas menggambarkan proses seleksi atribut data yang digunakan dalam penelitian untuk analisis clustering menggunakan algoritma K-Mean. Type disetel pada "Include attributes," yang berarti hanya atribut yang relevan dan signifikan yang akan dimasukkan dalam analisis. Pemilihan atribut dilakukan secara spesifik melalui Atribut filter type dengan pendekatan "A subset," di mana hanya

beberapa atribut utama yang dianggap relevan dipilih dari dataset awal. Atribut-atribut ini mencakup persentase kenaikan review, harga produk, jumlah terjual, kategori produk (*Kategori1*), kota pengiriman, omset minimum, penjualan bulan sebelumnya, review bulan sebelum dan kenaikan penjualan

Tabel 4. Hasil dari oprator Select Atribut

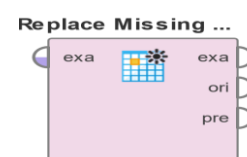
NO	Harga	Omset Maksimum	Jumlah Terjual	Kategori 1	Kota Pengiriman	Penjualan Bulan Sebelumnya	%Kenaikan Penjualan	Review Bulan Sebelumnya	%Kenaikan Review
1	99	990000	10000	Jam Tangan	KOTA JAKARTA UTARA - TANJUNG PRIOK, DKI JAKARTA, ID	2100	3.762	103	7.437
2	25000	242500000	9700	Tas Pria	KAB. BANDUNG - DAYEUKHOLOT, JAWA BARAT, ID	2600	2.731	1100	3.091
3	21500	193500000	9000	Ibu & Bayi	KOTA JAKARTA UTARA - KELAPA GADING, DKI JAKARTA, ID	2600	2.462	501	2.792
4	2750	25850000	9400	Olahraga & Outdoor	KOTA SURAKARTA (SOLO) - JEBRES, JAWA TENGAH, ID	3100	2.032	97	2.124
5	11640	116400000	10000	Fashion Bayi & Anak	KOTA JAKARTA BARAT - TAMBORA, DKI JAKARTA, ID	3300	2.030	520	2.846
...	...	...	...	...	...	...	...	...	...
6402	18796	187960000	10000	Voucher	KOTA SURABAYA - MULYOOREJO, JAWA TIMUR, ID	—	—	—	—
6403	6000	15600000	2600	Voucher	KOTA JAKARTA BARAT - CENGKARENG, DKI JAKARTA, ID	—	—	—	—
6404	6000	166200000	27700	Voucher	KAB. JEMBER - UMBULSARI, JAWA TIMUR, ID	—	—	—	—

Tabel 4 menyajikan hasil seleksi data awal yang relevan untuk analisis clustering menggunakan algoritma K-Means, dengan kolom-kolom utama mencakup nomor urut data, harga maksimum produk, omset maksimum dari penjualan, jumlah produk terjual, kategori produk (seperti "Jam Tangan" dan "Tas Pria"), serta kota asal pengiriman produk (contoh: "KOTA JAKARTA UTARA - TANJUNG PRIOK").

4.3. Preprocessing

Pada proses pengolahan data, langkah penting yang dilakukan adalah menangani nilai-nilai yang hilang (missing values) dalam dataset. Penanganan ini

bertujuan untuk memastikan bahwa analisis data dapat dilakukan secara akurat tanpa gangguan akibat data yang tidak lengkap. Gambar berikut menunjukkan metode yang digunakan untuk menggantikan nilai-nilai yang hilang dalam dataset



Gambar 3. operator Replace Missing Value

Dalam penelitian ini, operator *Replace Missing Values* digunakan untuk menangani data yang hilang (missing values) dalam dataset, sehingga kualitas data tetap terjaga untuk analisis clustering menggunakan algoritma K-Mean.

Tabel 5. Parameter Replace Missing Values

No	Parameter	Isi
1	Atribut filter type	All
2	Default	Average
3	Columns	Edit list

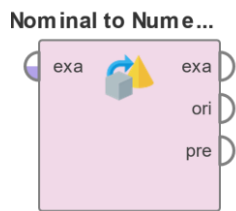
Tabel ini menggambarkan parameter yang digunakan dalam tahap pra-pemrosesan data pada eksperimen clustering menggunakan algoritma K-Mean. Parameter Atribut filter type: All menunjukkan bahwa pada tahap ini, semua atribut yang telah dipilih sebelumnya akan dipertimbangkan tanpa ada pengecualian. Artinya, tidak ada pembatasan atau filter tambahan yang diterapkan pada data yang akan diproses.

Tabel 6. Hasil dari oprator Parameter Replace Missing

Row n0	Harga	Omset minimum	Jumlah terjual	kategori 1	Kota Pengiriman	Bulan penjualan Sebelumnya	%Kenaikan penjualan	Review Bulan Sebelumnya	%Kenaikan Review
1	99	990000	10000	Jam Tangan	KOTA JAKARTA UTARA - TANJUNG PRIOK, DKI JAKARTA, ID	2100	3.7619	103	7.4369
2	25000	242500000	9700	Tas Pria	KAB. BANDUNG - DAYEUEHKOLOT, JAWA BARAT, ID	2600	2.7308	1100	3.0909
3	21500	193500000	9000	Ibu & Bayi	KOTA JAKARTA UTARA - KELAPA GADING, DKI JAKARTA, ID	2600	2.462	501	2.7924
4	2750	25850000	9400	Olahraga & Outdoor	KOTA SURAKARTA (SOLO) - JEBRES, JAWA TENGAH, ID	3100	2.032	97	2.124
5	11640	116400000	10000	Fashion Bayi & Anak	KOTA JAKARTA BARAT - TAMBORA, DKI JAKARTA, ID	3300	2.030	520	2.8466
...	...	...	...		...	...	...	...	...
6494	64943	5.302705476791068	12725	Perlengkapan Rumah	KOTA JAKARTA UTARA - PENJARINGAN, DKI JAKARTA, ID	14016	0.087	9700	0.296
6495	64943	5.302705476791068	12725	Perlengkapan Rumah	KOTA JAKARTA UTARA - PENJARINGAN, DKI JAKARTA, ID	14016	0.087	9700	0.296
6496	64943	5.302705476791068	12725	Perlengkapan Rumah	KOTA JAKARTA UTARA - PENJARINGAN, DKI JAKARTA, ID	14016	0.087	9700	0.296

Dalam proses identifikasi data hilang (missing value), langkah awal yang dilakukan adalah memeriksa keberadaan sel kosong atau nilai yang tidak valid di dalam dataset, seperti NULL, NaN, atau angka nol jika dianggap sebagai data hilang.

4.4. Transformation



Gambar 4. operator Nominal to Numerical

Fungsi dari operator Nominal to Numerical dalam penelitian ini adalah untuk mengonversi data kategori (nominal) menjadi format numerik agar dapat diproses oleh algoritma K-Mean yang hanya bekerja dengan data numerik.

Tabel 8. Parameer Nominal to Numerical

No	Parameter	Isi
1	Atribut filter type	All
2	Coding type	Unique integers

Tabel ini menjelaskan dua parameter penting dalam tahap transformasi data dalam eksperimen clustering menggunakan algoritma K-Mean. Parameter pertama, Atribut filter type: All, menunjukkan bahwa semua atribut yang telah dipilih sebelumnya akan dipertimbangkan tanpa ada pengecualian atau pembatasan tambahan. Dalam konteks penelitian ini, hal tersebut berarti semua variabel yang terkait dengan performa supplier, seperti % Kenaikan Review, Harga, Jumlah Terjual, dan lainnya, akan digunakan tanpa filter lebih lanjut.

Tabel 8. Hasil dari oprator Nominal To Nomorical

NO	Kategori 1	Kota Pengiriman	Harga	Omset Maksimum	Jumlah Terjual	Penjualan Bulan Sebelumnya	%Kenaikan Penjualan	Review Bulsn Sebelumnya	%Kenaikan Review
1	0	0	99	990000	10000	2100	3.762	103	7.437
2	1	1	25000	242500000	9700	2600	2.731	1100	3.091
3	2	2	21500	193500000	9000	2600	2.462	501	2.792
4	3	3	2750	25850000	9400	3100	2.032	97	2.124
5	4	4	11640	116400000	10000	3300	2.030	520	2.846
...	...	...	...	...	...	...	...	...	...
2186	12	344	990	9900000	10000	100	0	282	0.415
2187	12	286	7500	75000000	10000	100	0	2200	0.182
2188	22	80	170	1700000	10000	100	0	979	0.839

Berdasarkan Tabel 8 yang menampilkan hasil dari operator *Nominal To Numerical*, dapat disimpulkan bahwa proses ini bertujuan untuk mengonversi data kategorikal, seperti *Kategori 1* dan *Kota Pengiriman*, menjadi bentuk numerik.



Gambar 5. operator Clustering

Fungsi dari clustering dalam penelitian ini adalah untuk mengelompokkan data supplier pada platform e-

commerce Shopee berdasarkan atribut-atribut yang relevan,

Tabel 9. Parameter Clustering

No	Parameter	Isi
1	K	2,3,4,5
2	Max runs	10

Tabel yang diberikan menggambarkan parameter yang digunakan dalam eksperimen untuk klastering data dengan metode *K-Mean*. Di kolom pertama, terdapat nama parameter yang digunakan dalam eksperimen, sedangkan kolom kedua menjelaskan konfigurasi atau nilai yang diatur untuk masing-masing parameter.

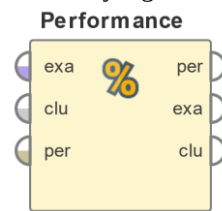
Tabel 10. Hasil dari operator Data Mining

No	Id	Clustering	Kategori 1	Kota Pengiriman	Harga	Omset Maksimum	Jumlah Terjual	Penjualan Bulan Sebelumnya	%Kenaikan Penjualan	Review Bulsn Sebelumnya	%Kenaikan Review
1	1	cluster_0	0	0	99	990000	10000	2100	3.762	103	7.437
2	2	cluster_0	1	1	25000	242500000	9700	2600	2.731	1100	3.091
3	3	cluster_0	2	2	21500	193500000	9000	2600	2.462	501	2.792
4	4	cluster_0	3	3	2750	25850000	9400	3100	2.032	97	2.124
5	5	cluster_0	4	4	11640	116400000	10000	3300	2.030	520	2.846
...	...	...	...	...	...	...	...	...	...	...	...
6494	6v494	cluster_0	19	10	64943	5.30270547679	12725	14016	0.087	9700	0.296
6495	6495	cluster_0	19	10	64943	5.30270547679	12725	14016	0.087	9700	0.296
6496	6496	cluster_0	19	10	64943	5.30270547679	12725	14016	0.87	9700	0.296

Tabel 10 menunjukkan hasil proses data mining menggunakan algoritma K-Means yang merepresentasikan hasil pengelompokan data ke dalam cluster tertentu, dalam hal ini semua data berada dalam "cluster_0." Tabel ini memuat sejumlah atribut, seperti nomor urut data (N O),

4.5. Evaluation

Kinerja algoritma sangat bergantung pada kualitas data dan parameter yang digunakan selama proses analisis. Evaluasi performa dilakukan untuk memastikan bahwa model yang dihasilkan mampu memberikan hasil yang optimal dan sesuai dengan tujuan penelitian. Gambar berikut menunjukkan hasil evaluasi performa model yang telah diterapkan.



Gambar 6 Performance

Fungsi dari *Performance* dalam penelitian ini adalah untuk mengevaluasi kualitas hasil klusterisasi yang dihasilkan oleh algoritma K-Means.

Tabel 11 Parameter Performance

No	Parameter	Isi
1	Main criterion	Avg within centroid dista

Tabel yang diberikan menunjukkan salah satu parameter dalam eksperimen yang terkait dengan evaluasi hasil klastering menggunakan algoritma K-Mean.

4.6. Data Mining

Dalam analisis data, langkah awal yang penting adalah memastikan data berada dalam format yang sesuai dan siap untuk diproses. Hal ini meliputi penanganan data yang hilang, konversi jenis data, dan pengelompokan awal sebagai bagian dari eksplorasi data. Berikut adalah ilustrasi langkah-langkah tersebut dalam konteks analisis data clustering

Tabel 12. Parameter Performance

No	Parameter	Isi
1	Main criterion	Avg within centroid dista

Tabel yang diberikan menunjukkan salah satu parameter dalam eksperimen yang terkait dengan evaluasi hasil klastering menggunakan algoritma K-Mean.

Tabel 13. Hasil dari oprator Evaluation

No	Id	Cluster ing	Katego ri 1	Kota Pengiri man	Har ga	Omset Maksim um	Juml ah Terj ual	Penjual an Bulan Sebelum nya	%Kenai kan Penjual an	Review Bulsn Sebelum nya	%Kenai kan Review
1	1	cluster_0	0	0	99	990000	10000	2100	3.762	103	7.437
2	2	cluster_0	1	1	25000	242500000	9700	2600	2.731	1100	3.091
3	3	cluster_0	2	2	21500	193500000	9000	2600	2.462	501	2.792
4	4	cluster_0	3	3	2750	25850000	9400	3100	2.032	97	2.124
5	5	cluster_0	4	4	11640	116400000	10000	3300	2.030	520	2.846
...	...	...	...	...	...	...	...	...	...	...	...
6494	6494	cluster_0	19	10	64943	5.30270547679	12725	14016	0.087	9700	0.296
6495	6495	cluster_0	19	10	64943	5.30270547679	12725	14016	0.087	9700	0.296
6496	6496	cluster_0	19	10	64943	5.30270547679	12725	14016	0,087	9700	0.296

Tabel 1 menyajikan hasil evaluasi dari operator clustering yang digunakan pada data penelitian, yang meliputi beberapa metrik evaluasi untuk setiap cluster.

4.7. Berapa nilai k terbaik berdasarkan Davies Bouldin indens

Nilai k terbaik berdasarkan Davies-Bouldin Index (DBI) dapat ditentukan dengan mengevaluasi hasil clustering untuk berbagai nilai k (jumlah cluster), di mana nilai DBI dihitung untuk setiap nilai k, dan semakin rendah nilai DBI, semakin baik kualitas clustering tersebut. Langkah-langkah untuk

menentukan nilai k terbaik dimulai dengan menginisialisasi rentang k yang akan diuji (misalnya, 2 hingga 10), kemudian menjalankan algoritma K-Means untuk setiap nilai k dalam rentang tersebut. Selanjutnya, hasil clustering dievaluasi dengan menghitung nilai DBI, yang mempertimbangkan variabel dispersion (variasi dalam cluster yang menunjukkan seberapa kompak cluster) dan separation (jarak antar cluster yang menunjukkan seberapa terpisah cluster satu sama lain). Nilai k terbaik dipilih berdasarkan nilai DBI terkecil, karena nilai tersebut menunjukkan pembentukan cluster yang lebih

optimal. Jika data hasil clustering, seperti pusat cluster, jarak antar cluster, atau data penjualan tersedia, nilai DBI dapat dihitung untuk mendukung analisis lebih lanjut.

4.8. Atribut apa aja yang mempengaruhi nilai k terbaik berdasarkan Davies Boulding Index

Hasil analisis data menunjukkan bahwa faktor harga, jumlah ulasan, dan kenaikan penjualan memiliki pengaruh signifikan terhadap performa produk di platform Shopee. Produk-produk dengan harga yang kompetitif, ulasan positif, dan tingkat kenaikan penjualan yang stabil cenderung memiliki jumlah penjualan yang lebih tinggi. Temuan ini konsisten dengan penelitian yang menyatakan bahwa ulasan konsumen dapat memengaruhi keputusan pembelian secara signifikan. Selain itu, analisis nilai k terbaik berdasarkan Davies-Bouldin Index menunjukkan bahwa atribut-atribut seperti % Kenaikan Review, Harga, Jumlah Terjual, Kategori, Kota Pengiriman, Omset Minimum, Penjualan Bulan Sebelumnya, Review Bulan Sebelumnya, dan Kenaikan penjualan memiliki pengaruh besar terhadap pemilihan jumlah cluster optimal. Atribut-atribut ini membantu menciptakan cluster yang lebih terpisah secara signifikan dan memiliki tingkat kekompakan yang tinggi di dalam masing-masing cluster, sehingga menghasilkan nilai Davies-Bouldin Index yang lebih rendah, yang menunjukkan performa clustering yang lebih baik.

5. KESIMPULAN DAN SARAN

Penelitian ini berhasil mengelompokkan supplier terlaris di Shopee menggunakan algoritma K-Means dengan kualitas clustering yang baik, ditunjukkan oleh nilai Davies-Bouldin Index (DBI) sebesar 0,213. Atribut yang paling berpengaruh dalam pembentukan cluster meliputi jumlah penjualan, harga produk, dan rating produk, yang mencerminkan performa supplier, daya tarik pasar, dan kepercayaan konsumen. Temuan ini menunjukkan bahwa kombinasi strategi harga kompetitif dan kualitas produk yang tinggi merupakan kunci keberhasilan supplier di Shopee.

Penjual disarankan untuk mengoptimalkan strategi penjualan dengan menawarkan harga yang kompetitif dan menjaga kualitas produk agar dapat memperoleh ulasan serta rating yang baik, yang secara signifikan memengaruhi performa penjualan. Selain itu, hasil clustering dapat dimanfaatkan untuk menyusun strategi pemasaran yang lebih spesifik, misalnya dengan memberikan promosi pada kelompok supplier dengan performa rendah. Penelitian juga dapat dikembangkan lebih lanjut melalui pembangunan model prediksi berbasis machine learning, serta penambahan atribut data seperti biaya pengiriman dan lokasi penjual untuk menghasilkan analisis yang lebih komprehensif. Untuk memastikan kualitas clustering, disarankan menggunakan metrik evaluasi tambahan seperti Silhouette Score yang dapat

memberikan validasi hasil clustering secara lebih menyeluruh.

DAFTAR PUSTAKA

- [1] 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.
- [2] H. Al Rasyid, B. F. K. Soebari, and D. S. Y. Kartika, "Implementasi Algoritma K-Means Clustering Untuk Pengelompokan Penjualan Produk Pada Online Shop Toko Gizi," *Pros. Semin. Nas. Teknol. dan Sist. Inf.*, vol. 2, no. 1, pp. 242–248, 2022, doi: 10.33005/sitasi.v2i1.304.
- [3] B. Susilo, N. A. Ramdhan, and O. S. Bachri, "Application of the K-Nearest Neighbor Algorithm for Predicting Digital Product Sales Penerapan Algoritma K-Nearest Neighbor untuk Prediksi Penjualan Produk Digital," vol. 4, no. October, pp. 1466–1476, 2024.
- [4] M. Herviany, S. Putri Delima, T. Nurhidayah, and K. Kasini, "Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Daerah Rawan Tanah Longsor Pada Provinsi Jawa Barat," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 1, no. 1, pp. 34–40, 2021, doi: 10.57152/malcom.v1i1.60.
- [5] D. Lestari, A. Charis Fauzan, F. Ilmu Eksakta, P. Studi Ilmu Komputer, U. Nahdlatul Ulama Blitar, and J. Masjid No, "Penerapan Algoritma Pillar Untuk Optimasi Penentuan Titik Awal Centroid Pada Algoritma K-Means Clustering," *JOISIE J. Inf. Syst. Informatics Eng.*, vol. 6, no. 1, pp. 15–24, 2022.
- [6] C. S. D. B. Sembiring, L. Hanum, and S. P. Tamba, "Penerapan Data Mining Menggunakan Algoritma K-Means Untuk Menentukan Judul Skripsi Dan Jurnal Penelitian (Studi Kasus Ftik Unpri)," *J. Sist. Inf. dan Ilmu Komput. Prima (JUSIKOM PRIMA)*, vol. 5, no. 2, pp. 80–85, 2022, doi: 10.34012/jurnalsisteminformasidanilmukomputer.v5i2.2393.
- [7] F. M. Evanita, I. Cholissodin, and ..., "Pengelompokan Toko E-commerce Shopee berdasarkan Reputasi Toko menggunakan Metode Clustering K-Medoids," ... *dan Ilmu Komput. e-ISSN*, vol. 5, no. 3, pp. 1230–1236, 2021, [Online]. Available: <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/8784/4029>
- [8] M. Y. A. Thoriq, I. A. Siradjuddin, and K. E. Permana, "Deteksi Wajah Manusia Berbasis One Stage Detector Menggunakan Metode You Only Look Once (Yolo)," *J. Teknoinfo*, vol. 17, no. 1, p. 66, 2023, doi: 10.33365/jti.v17i1.1884.