

Contents list available at www.jurnal.unimed.ac.id

CESS
(Journal of Computing Engineering, System and Science)

journal homepage: <https://jurnal.unimed.ac.id/2012/index.php/cess>



Analisis Data Penjualan Menggunakan Algoritma K-Means Clustering Pada Toko Daun Indah Di Shopee

Analysis of Sales Data Using K-Means Clustering Algorithm at Daun Indah Shop in Shopee

Hibrizi Dzaky Dermawan^{1*}, Rudi Kurniawan², Yudhistira Arie Wijaya³

^{1,3} Program Studi Teknik Informatika, STMIK IKMI Cirebon

² Program Studi Rekayasa Perangkat Lunak, STMIK IKMI Cirebon

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

email: ¹dzakydermawan5@gmail.com, ²rudi226@gmail.com, ³yudhistira010471@gmail.com

ABSTRAK

Toko Daun Indah adalah sebuah usaha yang menjual berbagai pilihan produk kecantikan, tidak semua produk tersebut dimanati pelanggan. Namun data penjualan di Toko Daun Indah belum dikelola dengan baik untuk menentukan produk mana yang paling diminati dan mana yang kurang diminati pelanggan. Akibatnya, data tersebut berfungsi sebagai dokumen arsip dan belum dimanfaatkan untuk strategi pemasaran. Sehingga perlu diterapkannya teknik data mining dalam mengembangkan strategi pemasaran penjualan. Tujuan penelitian adalah menganalisis data penjualan untuk mengetahui *cluster* terbaik berdasarkan *Davies Bouldin Index*, *iterasi*, dan *measure type* yang menghasilkan K Optimal. Metode yang digunakan adalah *Cross-Industry Standard Process Model for Data Mining* dengan algoritma *K-Means Clustering* untuk mengelompokkan data penjualan berdasarkan karakteristiknya, karena mudah dalam penerapannya, dan relatif cepat. Berdasarkan hasil pengelompokan data penjualan produk dengan metode *K-Means* diperoleh parameter yang optimal. Dengan melakukan uji dengan jumlah *cluster* ($k= 2-25$), hasil metode *K-Means* menunjukkan nilai DBI paling optimal sebesar -0.149 dengan 2 *cluster* pada *iterasi* ke-1 sebanyak 30 *iterasi*, *Measure type Mixed Measures*.

Kata Kunci: *Data Mining, K-Means Clustering, Analisis Penjualan, Davies Bouldin Index, CRISP-DM*

ABSTRACT

Toko Daun Indah is a business that sells a wide selection of beauty products, not all of these products are enjoyed by customers. However, sales data at Toko Daun Indah has not been properly managed to determine which products are most in demand and which are less in demand by customers. As a result, the data serves as an archive document and has not been

*Penulis Korespondensi:

email: dzakydermawan5@gmail.com

utilized for marketing strategies. So it is necessary to apply data mining techniques in developing sales marketing strategies. The research objective is to analyze sales data to determine the best *cluster* based on Davies Bouldin Index, iteration, and measure type that produces K Optimal. The method used is the Cross-Industry Standard Process Model for Data Mining with *K-Means Clustering* algorithm to group sales data based on its characteristics, because it is easy to apply, and relatively fast. Based on the results of product sales data *clustering* with the *K-Means* method, the optimal parameters are obtained. By conducting tests with the number of *clusters* ($k = 2-25$), the results of the *K-Means* method show the most optimal DBI value of -0.149 with 2 *clusters* at the 1st iteration of 30 iterations, Measure type Mixed Measures.

Keywords: *Data Mining, K-Means Clustering, Sales Analysis, Davies Bouldin Index, CRISP-DM*

1. PENDAHULUAN

E-Commerce telah menjadi platform yang populer bagi konsumen untuk melakukan berbagai jenis transaksi jual beli secara online. Terutama dalam beberapa tahun terakhir, *E-Commerce* telah berkembang pesat di Indonesia [1]. Salah satu platform *E-Commerce* yang memegang peran signifikan dalam industri ini adalah *Shopee* [2]. *Shopee* adalah salah satu platform *E-Commerce* terkemuka di Indonesia, yang memungkinkan penjual seperti "Toko Daun Indah" untuk beroperasi secara online [3]. Dalam konteks persaingan yang semakin ketat, data mining dan analisis penjualan menjadi penting bagi penjual untuk memahami perilaku konsumen dan meningkatkan strategi pemasaran [4].

Permasalahan yang dihadapi oleh pemilik toko online adalah bagaimana Toko Daun Indah dapat mengoptimalkan penjualan dan memahami perilaku konsumen secara mendalam, terutama dalam menghadapi persaingan yang ketat untuk mengatur strategi penjualan [5]. Analisis data penjualan merupakan aset penting dalam upaya ini, karena data dapat menyebabkan kerugian jika tidak digunakan dengan bijak, tetapi dapat memberikan keuntungan jika dilakukan pengolahan dan analisis dengan baik [6]. Data penjualan yang dihasilkan dari transaksi pelanggan mencakup informasi berharga tentang pola pembelian dan tren yang berubah dari waktu ke waktu, yang dapat digunakan untuk meningkatkan performa penjualan [4].

Namun, metode tradisional dalam menganalisis data penjualan mungkin tidak cukup efektif mengingat volume data yang terus meningkat. Oleh karena itu, untuk mengatasi masalah tersebut diperlukan analisis data penjualan produk yang baik. Salah satu cara untuk melakukan hal tersebut adalah dengan menggunakan data mining [7]. Data mining adalah proses untuk mengekstrak informasi yang berguna dan belum diketahui sebelumnya dari kumpulan data yang besar sehingga informasi tersebut dapat digunakan untuk membuat keputusan bisnis yang penting [8]. *K-Means Clustering* merupakan salah satu metode data mining yang dapat digunakan untuk memecahkan tantangan tersebut, yaitu bagaimana mengoptimalkan penjualan dan memahami perilaku konsumen secara mendalam dalam menghadapi persaingan di bisnis *E-Commerce* [9].

Metode ini memungkinkan pemilik toko, seperti "Toko Daun Indah," untuk mengidentifikasi kelompok data penjualan produk berdasarkan kesamaan karakteristik atau perbedaannya. Hal ini dilakukan dengan cara membagi data penjualan produk ke dalam beberapa kelompok atau *cluster* menggunakan algoritma *Clustering*. [8]. Hal ini dapat

membantu pemilik toko untuk merumuskan strategi pemasaran yang lebih tepat sasaran dan mengoptimalkan data penjualan barang berdasarkan pola penjualan yang ditemukan [10]. Data mining tidak hanya sekadar alat analisis, tetapi juga merupakan Solusi efektif dalam sistem pengambilan Keputusan [11]. Hal ini terutama penting karena dapat memberikan wawasan berharga dan solusi yang dapat membantu pemilik toko online seperti Toko Daun Indah untuk menghadapi tantangan persaingan yang sengit dan meningkatkan performa pendapatan penjualan Toko Daun Indah di ekosistem *E-Commerce* [5].

Penggunaan aplikasi RapidMiner versi 10.1 sebagai alat pengelolaan data dalam penelitian ini didasarkan pada hasil studi sebelumnya [12]. Berdasarkan latar belakang yang telah diuraikan, penelitian ini bertujuan menerapkan metode Data Mining *Clustering* menggunakan algoritma *K-Means* untuk menganalisis data penjualan produk. Tujuan penelitian ini menentukan nilai K optimal dari beberapa atribut yang digunakan dalam pengelompokan data. Selain itu, penelitian ini juga akan menentukan parameter *Measure Type* dan menganalisis berdasarkan *Davies Bouldin Index* (DBI) optimal pada setiap *Cluster* yang dihasilkan oleh algoritma *K-Means*. Selanjutnya, penelitian ini menganalisis pengaruh nilai *iterasi* terhadap nilai K optimal pada dataset penjualan produk dari Toko Daun Indah dengan menggunakan algoritma *K-Means* dan parameter *Cluster Distance Performance*.

Penelitian ini menggunakan metode kuantitatif dengan pendekatan eksperimental. Data dalam penelitian ini berasal dari Toko Daun Indah di platform *E-Commerce Shopee*. Data yang digunakan data transaksi bulan November 2022 hingga Mei 2023 sebanyak 18.119 baris data. data diolah menggunakan metode analisis data *Clustering* algoritma *K-Means* dengan metode *Data Mining Cross-Industry Standard Process Model For Data Mining* (CRISP-DM). Penelitian ini dapat memberikan pemahaman terhadap perilaku penjualan pada platform *E-Commerce Shopee*, terutama dalam hal pengukuran dengan Parameter *Measure Type* terbaik menghasilkan *Davies Bouldin Index* (DBI) terbaik dari sebuah data. Sehingga dapat memberikan wawasan mendalam mengenai produk-produk yang paling diminati, serta kurang diminati di dalam platform tersebut. Informasi yang diperoleh dari penelitian ini dapat menjadi landasan yang kuat dalam mengembangkan strategi pemasaran yang lebih efektif bagi Toko Daun Indah.

2. TINJAUAN TEORI

2.1. Algoritma K-Means

Algoritma *K-Means* merupakan salah satu metode partisi yang membagi data menjadi kelompok dengan menentukan nilai awal centroid. Proses iteratif digunakan dalam algoritma ini untuk mendapatkan *cluster* data. Pengguna menginput jumlah *cluster* awal yang diinginkan, dan algoritma menghasilkan jumlah *cluster* akhir. Jika algoritma harus menghasilkan K *cluster*, maka akan ada K awal dan K akhir. Metode *K-Means* secara acak memilih K pola sebagai titik awal centroid. Jumlah *iterasi* yang diperlukan untuk mencapai centroid *cluster* dipengaruhi oleh kandidat centroid *cluster* awal, dan *iterasi* berhenti jika posisi centroid tidak berubah. Nilai K dihitung menggunakan rumus Jarak *Euclidean*, yakni mencari jarak terpendek antara titik centroid dan data atau objek [2]. Data yang memiliki jarak terpendek ke centroid membentuk sebuah *cluster*. Algoritma *K-Means* mempartisi data ke dalam *cluster* atau segmen sehingga data dengan karakteristik serupa dikumpulkan dalam satu *cluster* [13].

Berikut adalah langkah-langkah dalam melakukan *K-Means Clustering*.

1. Menentukan nilai *k-cluster* secara acak.
2. Memulai nilai acak ke pusat *cluster* awal (centroid) dari sejumlah *k-cluster* yang telah ditentukan.
3. Menghitung jarak setiap data input ke centroid berdasarkan rumus jarak (Euclidian Distance) untuk mencapai jarak terpendek maksimum dari setiap data ke centroid. Rumus Persamaan Euclidian Distance dapat dilihat dibawah ini.

$$d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \quad (2)$$
4. Mengelompokkan data berdasarkan jarak ke centroid.
5. Memperbarui nilai centroid dengan menghitung rata-rata dari *cluster* yang bersangkutan menggunakan rumus:
 - a. d : jarak antar dua objek
 - b. x_1 : objek pertama
 - c. x_2 : objek kedua
 - d. d_i = jumlah dari nilai jarak yang masuk dalam masing-masing *Cluster* Langkah 2-5 diulangi sampai anggota setiap *cluster* tidak berubah.
6. Setelah langkah 6 selesai dilakukan, nilai rata-rata pusat *cluster* (μ_j) pada percobaan terakhir akan menjadi dasar untuk menentukan klasifikasi data.

2.2. Clustering

Clustering adalah teknik untuk mengelompokkan sejumlah data menjadi kelompok-kelompok berdasarkan kemiripan data. Kemiripan data yang maksimum terdapat pada satu kelompok, sedangkan kemiripan data yang minimum terdapat pada kelompok lainnya [14].

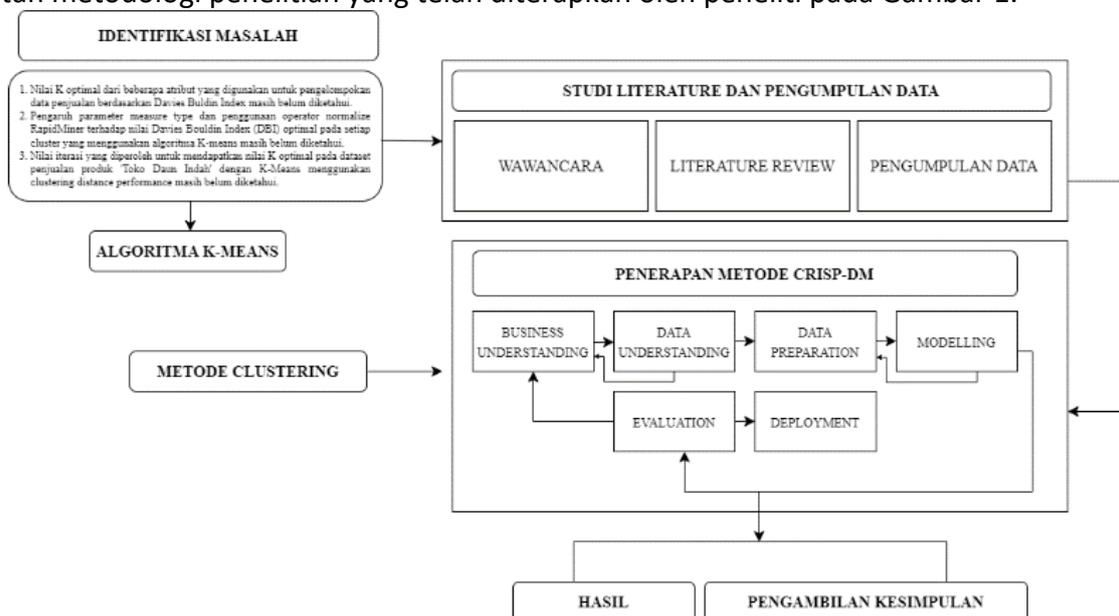
2.3. Davies Bouldin Index (DBI)

Davies Bouldin Index (DBI) merupakan sebuah pengukuran yang diperkenalkan pada tahun 1979 oleh David L. Davies dan Donald W. Bouldin untuk mengevaluasi jumlah *cluster* optimal atau validitas pada suatu metode pengelompokan. DBI mengukur kohesi dan jarak antar *cluster*. Kohesi didefinisikan sebagai jumlah kedekatan data terhadap titik pusat *cluster* dari *cluster* yang diikuti [15].

3. METODE

Penelitian dilakukan dengan menggunakan metode penelitian eksperimen dengan pendekatan kuantitatif. Pendekatan kuantitatif ini digunakan untuk mengolah dan menganalisis data dengan sistematis dan objektif guna memperoleh kesimpulan yang akurat dan dapat dipercaya. Dalam penelitian ini, diperlukan interaksi langsung seperti wawancara, dan observasi lapangan. Metode penelitian eksperimental memungkinkan pengumpulan data melalui proses pengujian dan eksperimen yang dilakukan. Data yang terkumpul selanjutnya akan dianalisis menggunakan algoritma *K-Means Clustering*. Dalam penelitian ini, metode yang tepat untuk membangun model adalah CRISP-DM (*Cross-Industry Standard Process for Data Mining*), yaitu metode data mining yang terstandar di berbagai industri. Tujuan CRISP-DM untuk menemukan pola yang menarik dan bermakna dalam data, yang dapat digunakan untuk menganalisis industri dan mengembangkan strategi pemecahan masalah bisnis. Metodologi penelitian mengacu pada tahapan yang disusun secara sistematis dalam

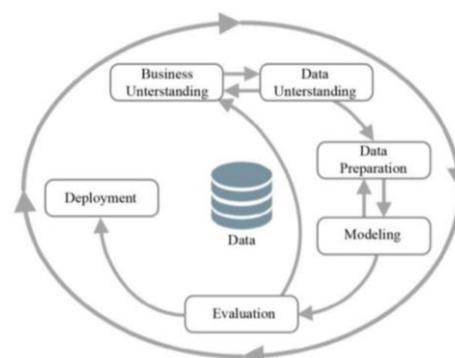
penelitian untuk memastikan bahwa penelitian berjalan dengan terstruktur. Berikut ini adalah urutan metodologi penelitian yang telah diterapkan oleh peneliti pada Gambar 1.



Gambar 1. Tahapan metode penelitian [16]

4. HASIL DAN PEMBAHASAN

Pengelompokan data dilakukan menggunakan algoritma *K-Means*. Dalam penelitian ini, metode yang tepat untuk membangun model adalah CRISP-DM (*Cross-Industry Standard Process for Data Mining*), yaitu metode data mining yang terstandar di berbagai industri. CRISP-DM terdiri dari enam tahapan, yaitu: 1. *Business Understanding*, 2. *Data Understanding*, 3. *Data Preparation*, 4. *Modelling*, 5. *Evaluation*, 6. *Deployment*, Tahapan-tahapan tersebut dapat dilihat pada Gambar 2.

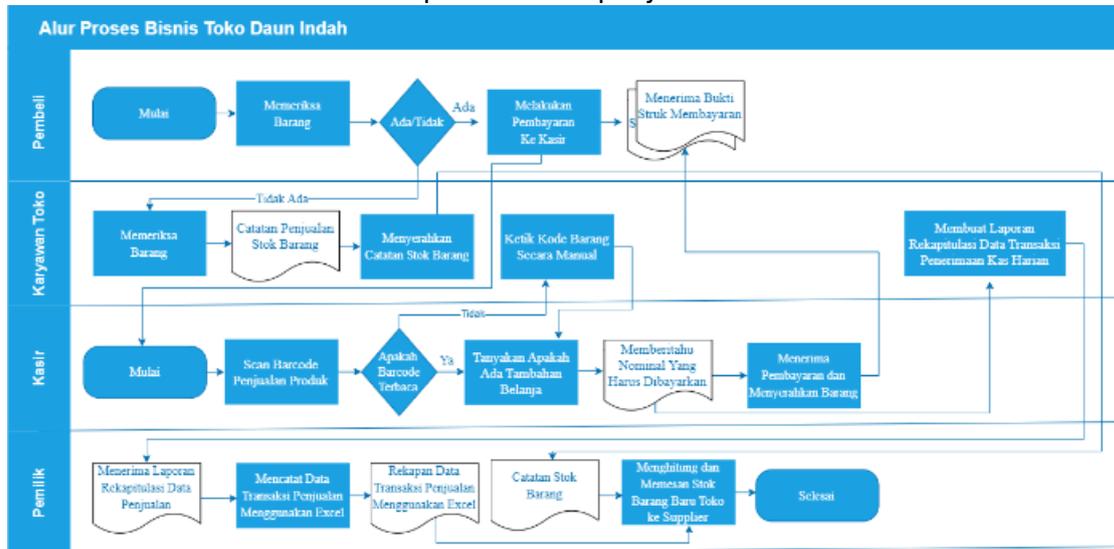


Gambar 2. Metode *Cross Industry Standard Process for Data Mining* [17]

4.1. *Business Understanding*

Pada tahap *Business Understanding*, analisis data perlu memahami kebutuhan bisnis dari analisis data penjualan yang akan dilakukan. Dalam hal ini, Toko Daun Indah memiliki tujuan untuk meningkatkan penjualan produk kecantikannya. Banyaknya produk yang dihasilkan tersebut terkadang tidak merata terjual, sehingga ada saja produk yang kurang diminati. Hal ini menyebabkan penumpukan stok barang, yang pada akhirnya menyebabkan

toko merugi. Alur Pemahaman bisnis tersebut dapat dilihat pada Gambar 3. Oleh karena itu, analisis data harus memahami alur proses bisnis penjualan Toko Daun Indah berikut:



Gambar 3. Alur Proses Tahapan Toko Daun Indah

Dari Gambar 3 terlihat bahwa pembeli yang datang ke Toko Daun Indah akan memilih produk kecantikan yang tersedia di toko tersebut. Jika barang yang dicari tidak tersedia, pembeli akan bertanya kepada staf toko. Staf toko akan mencatat barang yang tidak tersedia dan menyerahkannya kepada pemilik toko. Setelah pembeli memilih barang, konsumen akan membayar langsung ke kasir. Kasir menerima uang dan menyerahkan barang yang telah ditandai lunas kepada pembeli. Kasir kemudian menyerahkan hasil data transaksi kepada pemilik. Pemilik mencatat transaksi penjualan di Excel berdasarkan data transaksi penjualan. Pemilik juga melakukan proses pembelian barang dari supplier produk kecantikan atau stok barang baru berdasarkan catatan dari staf toko apabila tidak ada jenis varian dan produk di toko. Namun, kondisi ini sangat tidak baik untuk kelangsungan Toko Daun Indah karena pemilik hanya mencatat data transaksi penjualan menggunakan Excel sebagai dokumen arsip saja dan belum dimanfaatkan sepenuhnya untuk mengembangkan strategi pemasaran. Oleh karena itu, diperlukan metode yang tepat dalam melakukan evaluasi produk dan strategi penjualannya. Sebagai solusinya, Toko Daun Indah perlu menerapkan teknik data mining dalam menyusun strategi pemasaran penjualan produk kecantikan.

4.2. Data Understanding

Setelah mendapatkan data transaksi penjualan, langkah selanjutnya dalam metode CRISP-DM adalah memahami kebutuhan data yang terkait dengan pencapaian tujuan dalam menentukan strategi penjualan. Data transaksi penjualan tersebut merupakan kumpulan data yang didapat dari bagian admin Toko Daun Indah, berupa data transaksi penjualan dari November 2022 hingga Mei 2023. Data tersebut kemudian diubah ke dalam bentuk dokumen dengan format Excel, dengan jumlah 18.119 *record* dan 46 atribut. Untuk data pengumpulan data, digunakan data transaksi penjualan yang terdiri dari atribut-atribut yang terdapat pada Tabel 1 berikut:

Tabel 1. Karakteristik Dataset Penjualan Toko Daun Indah

No.	Atribut	Type Data	Keterangan
1	No. Pesanan	<i>Polynomial</i>	Nomor unik yang diberikan oleh <i>Shopee</i> untuk setiap pesanan yang dibuat oleh pembeli.
2	Status Pesanan	<i>Polynomial</i>	Status terkini dari pesanan yang dibuat oleh pembeli
3	Alasan Pembatalan	<i>Polynomial</i>	Alasan mengapa pembeli membatalkan pesanan.
4	Status Pembatalan/ Pengembalian	<i>Polynomial</i>	Status terkini dari pembatalan atau pengembalian pesanan.
5	No. Resi	<i>Polynomial</i>	Nomor unik yang diberikan oleh kurir untuk setiap pengiriman.
...
40	Catatan dari Pembeli	<i>Polynomial</i>	Catatan dari pembeli untuk penjual.
41	Catatan	<i>Polynomial</i>	Catatan dari penjual untuk pembeli.
42	Nama Penerima	<i>Polynomial</i>	Nama penerima pesanan.
43	No. Telepon	<i>Polynomial</i>	Nomor telepon penerima pesanan.
44	Alamat Pengiriman	<i>Polynomial</i>	Alamat pengiriman pesanan.
45	Kota/Kabupaten	<i>Polynomial</i>	Kota atau kabupaten tempat pengiriman pesanan.
46	Provinsi	<i>Polynomial</i>	Provinsi tempat pengiriman pesanan.

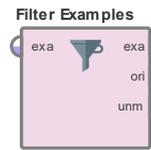
4.3. Data Preparation

Data Preparation dilakukan untuk mempersiapkan data yang telah didapatkan sebelumnya. Persiapan yang dilakukan meliputi pemilihan data, *preprocessing* data, dan *transformation*. Untuk membaca dataset penjualan toko Daun Indah dari file excel, digunakan operator *Read Excel*. Operator digunakan untuk memasukkan data excel dari komputer pengguna ke dalam proses RapidMiner. Tahap ini bertujuan membangun data mentah awal menjadi dataset akhir yang selanjutnya akan digunakan pada tahap pemodelan *cluster* analisis data penjualan Toko Daun Indah. Dalam tahap *Data Selection*, dilakukan pemilihan data yang relevan dan sesuai dengan tujuan bisnis. Dalam kasus ini, atribut-atribut yang dipilih Jumlah, Nama Produk, Nama Variasi Produk, dan Nomor Pesanan. Data tersebut akan disaring dengan hanya memilih 4 atribut yang akan dijadikan analisis *cluster* dari 46 atribut. Hasil seleksi data dapat dilakukan dengan menggunakan operator *Select Attributes*, sebagaimana terlihat pada Gambar 4.

Name	Type	Missing	Statistics	Filter (4 / 4 attributes):
✓ No. Pesanan	Nominal	0	Least 230505TPF8WV47 (1) Most 230410NYUF0PAJ (46) Values 230410NYUF0F...	Search for Attributes
✓ Nama Produk	Nominal	0	Least maybelli [...] nude (1) Most Miranda [...] um (3335) Values Miranda [...] g F...	
✓ Nama Variasi	Nominal	4523	Least Tancho Pomade 40g (1) Most Resistance (763) Values Resistance (763...	
✓ Jumlah	Integer	0	Min 1 Max 221 Average 2.591	

Gambar 4. Tampilan Statistik Operator *Filter Example Attribute*

Sedangkan pada tahap *Data Preprocessing*, dilakukan pembersihan data dari *noise*, *Missing value*, dan data yang tidak konsisten bahwa terdapat 4.523 data yang memiliki nilai hilang pada atribut Nama Variasi untuk mengatasi nilai hilang pada atribut tersebut. Operator *filter Examples* digunakan untuk menghapus data yang memiliki nilai Missing, sebagaimana terlihat pada Gambar 5.

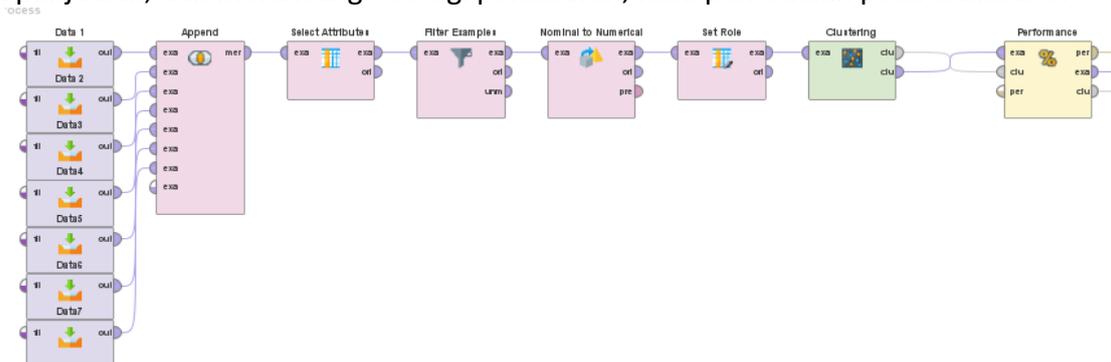


Gambar 5. Operator *Filter Examples*

Pada tahap *Transformation*, data yang telah dipilih Dalam dataset Penjualan Toko Daun Indah, atribut Nama Produk, Nama Variasi, dan No. Pesanan akan diubah menjadi *numeric* Tahap ini diperlukan karena *K-Means* adalah algoritma yang hanya bisa memproses data dengan variabel *numeric* dan dibersihkan akan ditransformasikan ke dalam bentuk yang sesuai untuk data mining.

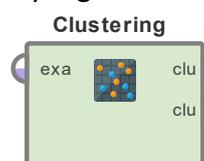
4.4. Modelling

Pada tahap pemodelan, kegiatan awal yang dilakukan adalah persiapan dataset penjualan Toko Daun Indah untuk pengelompokan data penjualan produk. Dalam penelitian ini, metode pemodelan yang dipilih adalah data mining dengan algoritma *K-Means Clustering*. Algoritma ini berfungsi untuk mengelompokkan data dengan cara mengukur jarak nilai antar data. Pendekatan ini bertujuan untuk memberikan pemahaman mendalam terkait pola penjualan produk di Toko Daun Indah. Proses pengelompokan ini dapat memberikan wawasan terkait karakteristik penjualan produk, membantu pemilik Toko Daun Indah untuk memahami tren penjualan, dan merancang strategi pemasaran, hasil pemodelan pada Gambar 6.



Gambar 6. Pemodelan *K-Means Clustering*

Dalam penelitian ini, diperlukan parameter operator *K-Means Clustering* menggunakan parameter default. Namun, untuk mengetahui pengaruh nilai k terhadap hasil *Clustering*, nilai K diubah mulai dari 2 hingga 25, seperti yang terlihat pada Gambar 7.



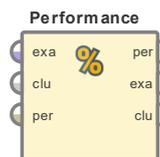
Gambar 7. Operator *K-Means Clustering*

Untuk mengetahui informasi parameter pada Operator *K-Means Clustering* yang digunakan, dapat dilihat pada Tabel 2 berikut.

Tabel 2. Keterangan Operator *K-Means Clustering*

Parameter	Keterangan
<i>K min</i>	2
<i>K max</i>	25
<i>Max optimization steps</i>	30
<i>Measure type</i>	<i>NumericalMeasures</i> <i>BregmanDivergences</i> <i>MixedMeasures</i>
<i>Clustering algorithm</i>	<i>K-Means</i>

Untuk mengetahui nilai DBI, operator *Performance (Cluster Distance Performance)* dengan *Davies Bouldin Index (DBI)* ditambahkan pada tahap ini, seperti yang dilihat pada Gambar 8.



Gambar 8. Operator *Performance (Cluster Distance Performance)*

Untuk memudahkan pemahaman tentang parameter pada *Operator Performance (Cluster Distance Performance)* yang digunakan, berikut ini disajikan informasinya dalam bentuk Tabel 3.

Tabel 3. Parameter Operator *Performance*

Parameter	Keterangan
<i>Main criterion</i>	<i>Davies Bouldin</i>

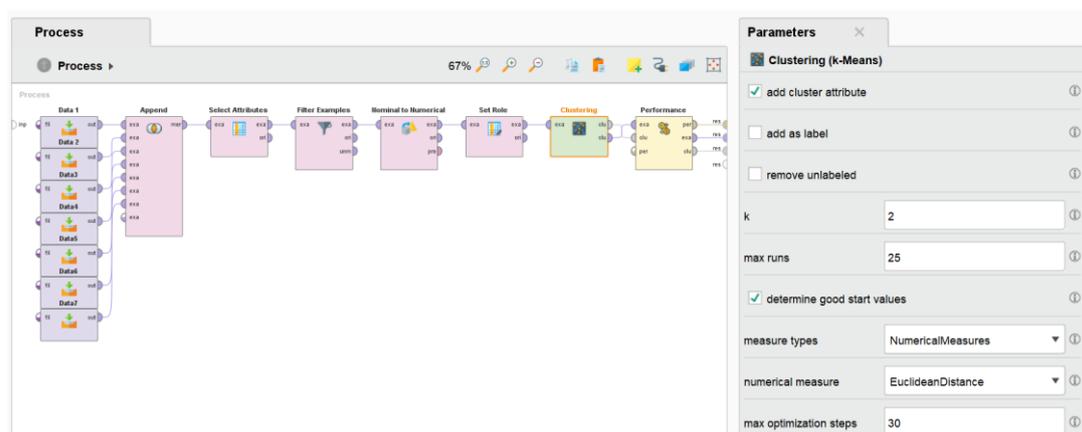
Berdasarkan hasil pengujian Parameter *Performance Vector* pengelompokan menggunakan algoritma *K-Means* dengan Software RapidMiner versi 10.1, nilai K yang paling baik performanya adalah nilai K yang mendekati angka 0. K=2 memiliki performa terbaik dengan nilai -0.149. menggunakan *Measure Types, MixedMeasures* dan pada *iterasi ke-1* dari 30 *iterasi*. Performa algoritma *K-Means* dapat diketahui dari nilai *Davies Bouldin Index*. DBI yang mendekati angka 0 menunjukkan bahwa performa algoritma semakin baik. Hal ini dapat dilihat pada Gambar 9.

```
Performance:
PerformanceVector [
----Avg. within centroid distance: -604535.967
----Avg. within centroid distance_cluster_0: -562186.275
----Avg. within centroid distance_cluster_1: -640100.538
*****Davies Bouldin: -0.149
```

Gambar 9. Parameter *Performance Vektor*

4.5. Evaluation

Pada tahap *Evaluation*, struktur pemodelan yang telah disusun akan dievaluasi menggunakan metode *Davies Bouldin Index* (DBI). Metode ini digunakan untuk menguji kualitas pemodelan berdasarkan karakteristik *Cluster* yang dihasilkan. Evaluasi ini bertujuan untuk mengidentifikasi apakah pemodelan yang telah dikembangkan tepat dan relevan ketika diterapkan pada konteks penelitian ini, serta sejalan dengan rencana awal penelitian. Hasil evaluasi akan menentukan langkah-langkah berikutnya, yaitu apakah proses ini dapat diteruskan atau perlu dilakukan perbaikan atau pengulangan dari awal jika terdapat ketidaksesuaian dengan rencana awal penelitian, Pemodelan terdapat pada Gambar 10.



Gambar 10. Pemodelan Evaluasi *Davies Bouldin Index*

Berdasarkan Gambar 11, hasil perhitungan *Davies Bouldin Index* menggunakan software RapidMiner pada *iterasi* ke-1 sebanyak 30 *iterasi* disajikan dalam bentuk tabel. Dalam 25 percobaan pengelompokan dengan jumlah *cluster* K yang bervariasi dari K=2 hingga K=25, dilakukan uji parameter jumlah *max run* sebanyak 25 kali, dengan *iterasi* sebanyak 30. Uji ini menggunakan tiga jenis *Measure*, yaitu *Numerical Measures*, *Bregman Divergences*, dan *Mixed Measures*. Berdasarkan hasil uji sebelumnya, peneliti mendapatkan informasi *Davies Bouldin Index*. prinsip DBI, nilai yang baik dalam data mining adalah semakin kecil atau mendekati nol, Menurut hasil riset yang telah dilakukan oleh Shanaz Khairunnisa, Muhammad Ihsan Jambak [18]. Dari data pada tabel tersebut, terdapat jumlah *cluster* yang memiliki nilai DBI dan jumlah *cluster* dengan nilai DBI terbaik adalah *Cluster 2*, dengan nilai DBI mendekati nol yaitu -0.149 jenis *measure type MixedMeasures*, Hasil *cluster* model terdapat pada Tabel 4.

Tabel 4. *Cluster Model*

<i>Cluster Model</i>	
<i>Cluster 0</i>	6.125
<i>Cluster 1</i>	7.471
<i>Total Number of item</i>	13.596

Hasil pengelompokan dan pengujian menggunakan RapidMiner menghasilkan 2 *cluster*, yaitu *Cluster 0* (Kurang Diminati) dengan 6.125 item dan *Cluster 1* (Diminati) dengan 7.471 item. Kedua *Cluster* tersebut terbentuk dari 13.596 dataset, Hasil terdapat pada Tabel 5.

Tabel 5. Hasil Analisis Data Penjualan *Davies Bouldin Index Optimal*

<i>Cluster</i>	<i>Measure Type</i>	<i>Iterasi</i>	<i>Davies Bouldin Index</i>
K=2	<i>NumericalMeasures</i>	1	-0.150
	<i>BregmanDivergences</i>		-0.150
	<i>MixedMeasures</i>		-0.149
K=3	<i>NumericalMeasures</i>	1	-0.159
	<i>BregmanDivergences</i>		0.160
	<i>MixedMeasures</i>		0.160
K=4	<i>NumericalMeasures</i>	1	-0.170
	<i>BregmanDivergences</i>		-0.169
	<i>MixedMeasures</i>		-0.169
K=5	<i>NumericalMeasures</i>	1	-0.169
	<i>BregmanDivergences</i>		-0.176
	<i>MixedMeasures</i>		-0.172
K=6	<i>NumericalMeasures</i>	1	-0.173
	<i>BregmanDivergences</i>		-0.176
	<i>MixedMeasures</i>		-0.173
K=7	<i>NumericalMeasures</i>	1	-0.173
	<i>BregmanDivergences</i>		-0.173
	<i>MixedMeasures</i>		-0.171
K=8	<i>NumericalMeasures</i>	1	-0.167
	<i>BregmanDivergences</i>		-0.187
	<i>MixedMeasures</i>		-0.175
K=9	<i>NumericalMeasures</i>	1	-0.177
	<i>BregmanDivergences</i>		-0.171
	<i>MixedMeasures</i>		-0.172
K=10	<i>NumericalMeasures</i>	1	-0.178
	<i>BregmanDivergences</i>		-0.176
	<i>MixedMeasures</i>		-0.181
K=11	<i>NumericalMeasures</i>	1	-0.184
	<i>BregmanDivergences</i>		-0.184
	<i>MixedMeasures</i>		-0.186
K=12	<i>NumericalMeasures</i>	1	-0.180
	<i>BregmanDivergences</i>		-0.183
	<i>MixedMeasures</i>		-0.183
K=13	<i>NumericalMeasures</i>	1	-0.192
	<i>BregmanDivergences</i>		-0.186
	<i>MixedMeasures</i>		-0.183
K=14	<i>NumericalMeasures</i>	1	-0.195
	<i>BregmanDivergences</i>		-0.185
	<i>MixedMeasures</i>		-0.189

K=15	<i>NumericalMeasures</i>		-0.194
	<i>BregmanDivergences</i>	1	-0.202
	<i>MixedMeasures</i>		-0.203
K=16	<i>NumericalMeasures</i>		-0.196
	<i>BregmanDivergences</i>	1	-0.193
	<i>MixedMeasures</i>		-0.187
K=17	<i>NumericalMeasures</i>		-0.199
	<i>BregmanDivergences</i>	1	-0.192
	<i>MixedMeasures</i>		-0.206
K=18	<i>NumericalMeasures</i>		-0.201
	<i>BregmanDivergences</i>	1	-0.196
	<i>MixedMeasures</i>		-0.206
K=19	<i>NumericalMeasures</i>		-0.209
	<i>BregmanDivergences</i>	1	-0.199
	<i>MixedMeasures</i>		-0.201
K=20	<i>NumericalMeasures</i>		-0.207
	<i>BregmanDivergences</i>	1	-0.204
	<i>MixedMeasures</i>		-0.203
K=21	<i>NumericalMeasures</i>		-0.203
	<i>BregmanDivergences</i>	1	-0.211
	<i>MixedMeasures</i>		-0.207
K=22	<i>NumericalMeasures</i>		-0.222
	<i>BregmanDivergences</i>	1	-0.204
	<i>MixedMeasures</i>		-0.225
K=23	<i>NumericalMeasures</i>		-0.208
	<i>BregmanDivergences</i>	1	-0.210
	<i>MixedMeasures</i>		-0.214
K=24	<i>NumericalMeasures</i>		-0.231
	<i>BregmanDivergences</i>	1	-0.221
	<i>MixedMeasures</i>		-0.224
K=25	<i>NumericalMeasures</i>		-0.213
	<i>BregmanDivergences</i>	1	-0.212
	<i>MixedMeasures</i>		-0.231

Untuk anggota data penjualan produk tiap *cluster* ditampilkan pada Tabel 6.

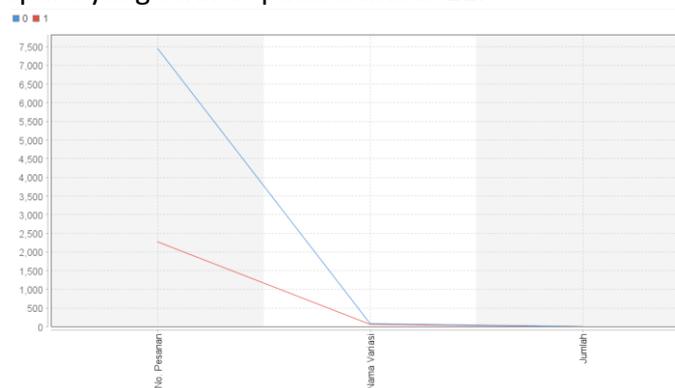
Tabel 6. Jumlah Data Penjualan Produk Tiap *Cluster*

<i>Cluster</i>	Jumlah Anggota	Nomor	Nama Produk	Nama Variasi Produk
0	6125	1	<i>BIZARRE 50ml EAU DE TOILETTE100% ORIGINAL PRODUCT -50ml</i>	<i>Sexy Killer</i>
		2	<i>Parfum Morris Travel 100ml</i>	<i>HOKKAIDO</i>
		3	<i>BIZARRE 50ml EAU DE TOILETTE100% ORIGINAL PRODUCT -50ml</i>	<i>Chic Haute</i>

Cluster	Jumlah Anggota	Nomor	Nama Produk	Nama Variasi Produk
		4	<i>Azzura Compact Powder</i>	<i>01 Light Biege</i>
		5	<i>BIZARRE 50ml EAU DE TOILETTE100% ORIGINAL PRODUCT -50ml</i>	<i>Sexy Killer</i>
		6	<i>BIZARRE 50ml EAU DE TOILETTE100% ORIGINAL PRODUCT -50ml</i>	<i>Chic Haute</i>
		7	<i>Implora lip cream matte</i>	<i>02-Terra Cotta</i>
		8	<i>Implora lip cream matte</i>	<i>01-Dusky Nude</i>
		9	<i>BIZARRE 50ml EAU DE TOILETTE100% ORIGINAL PRODUCT -50ml</i>	<i>Sexy Killer</i>
	
		6116	<i>BIZARRE 50ml EAU DE TOILETTE100% ORIGINAL PRODUCT -50ml</i>	<i>Dream Angel</i>
		6117	<i>BIZARRE 50ml EAU DE TOILETTE100% ORIGINAL PRODUCT -50ml</i>	<i>Sexy Killer</i>
		6118	<i>MATRIX OPTISTRAIGHT MINI RESISTANT & NORMAL 125</i>	<i>Resistance</i>
		6119	<i>MATRIX OPTISTRAIGHT MINI RESISTANT & NORMAL 125</i>	<i>Resistance</i>
		6120	<i>Hanasui Tintdorable Lip Stain-Lip Tint</i>	<i>01.cherry</i>
		6121	<i>BEDAK MARCKS 40g</i>	<i>Rose</i>
		6122	<i>BEDAK MARCKS 40g</i>	<i>Putih</i>
		6123	<i>BEDAK MARCKS 40g</i>	<i>Invisible</i>
		6124	<i>Miranda Hair Colour Cat Semir Rambut / Bleaching Premium</i>	<i>06 Bleaching</i>
		6125	<i>Miranda Hair Colour Cat Semir Rambut / Bleaching Premium</i>	<i>16 Ash Blonde</i>
1	7471	1	<i>CASABLANCA BODY MIST 100</i>	<i>Royal (Ungu)</i>
		2	<i>Implora lip cream matte</i>	<i>01-Dusky Nude</i>
		3	<i>CASABLANCA BODY MIST 100</i>	<i>Fantasy (Light Green)</i>
		4	<i>Parfum 110ml Tropical EDT Parfum Edition Morris</i>	<i>Romance</i>
		5	<i>Parfum 110ml Tropical EDT Parfum Edition Morris</i>	<i>Romance</i>
		6	<i>BEDAK MARCKS 40g</i>	<i>Rose</i>

Cluster	Jumlah Anggota	Nomor	Nama Produk	Nama Variasi Produk
		7	Miranda Hair Colour Cat Semir Rambut / Bleaching Premium	05 Pink
		8	Miranda Hair Colour Cat Semir Rambut / Bleaching Premium	14 Golden Brown
		9	Miranda Hair Colour Cat Semir Rambut / Bleaching Premium	11 Green
	
		7462	BIZARRE 50ml EAU DE TOILETTE100% ORIGINAL PRODUCT -50ml	Chic Haute
		7463	Implora lip cream matte	01-Dusky Nude
		7464	Implora lip cream matte	03-Dark Berry
		7465	BIZARRE 50ml EAU DE TOILETTE100% ORIGINAL PRODUCT -50ml	Chic Haute
		7466	Parfum 110ml Tropical EDT Parfum Edition Morris	Paradise
		7467	Parfum 110ml Tropical EDT Parfum Edition Morris	Marine
		7468	Parfum 110ml Tropical EDT Parfum Edition Morris	Nature
		7469	BIZARRE 50ml EAU DE TOILETTE100% ORIGINAL PRODUCT -50ml	Chic Haute
		7470	BIZARRE 50ml EAU DE TOILETTE100% ORIGINAL PRODUCT -50ml	Sexy Killer
		7471	BIZARRE 50ml EAU DE TOILETTE100% ORIGINAL PRODUCT -50ml	Chic Haute

Dari grafik plot hasil *cluster*, dapat disimpulkan bahwa atribut Nama Variasi dan No. Pesanan menjadi faktor yang menyebabkan terbentuknya *cluster* pola karakteristik data penjualan produk, seperti yang terlihat pada Gambar 11.



Gambar 11. Grafik Plot Hasil Cluster Attribute

Plot cluster di atas tidak dapat menunjukkan karakteristik setiap *Cluster* dengan jelas. Oleh karena itu, perlu melihat Gambar hasil *Centroid* yang dihasilkan pada Gambar 12. untuk melihat karakteristik tersebut secara lebih jelas.

Attribute	cluster_0	cluster_1
No. Pesanan	7454.241	2271.870
Nama Variasi	83.842	61.367
Jumlah	1.849	2.215

Gambar 12. Hasil *Centroid* Setiap *Cluster*

Pada Gambar 13, terdapat nilai *Centroid* untuk setiap *Attribute*. *Attribute* No. Pesanan untuk *Cluster* 0 memiliki nilai 7454.2421, *Cluster* 1 memiliki nilai 2271.870. *Attribute* Nama Variasi untuk *Cluster* 0 memiliki nilai 83.842, *Cluster* 1 memiliki nilai 61.367. *Attribute* Jumlah untuk *Cluster* 0 memiliki nilai 1.849, *Cluster* 1 memiliki nilai 2.215.

4.6. Deployment

Dalam proses pembuatan laporan hasil kegiatan data mining, tahap akhir disebut *Deployment*. Laporan akhir tersebut berisi pengetahuan yang diperoleh atau pengenalan pola dalam proses data mining. Berikut adalah beberapa analisis *Cluster* pada penelitian ini yang dapat dijelaskan sebagai berikut. Terdapat empat atribut yang digunakan dalam proses *Clustering*, yaitu Jumlah, Nama Produk, Nama Variasi Produk, dan Nomor Pesanan. Dalam penelitian ini, terdapat dua *cluster* yang digunakan, yaitu *Cluster* 0 yang diambil dari produk kategori kurang diminati dan *cluster* 1 yang diambil dari produk yang dijual dalam kategori diminati. Berikut adalah hasil *Clustering*:

1. *Cluster* 0, Produk Yang Dijual Pada Kategori Kurang Diminati Berjumlah 6125 Item Produk Terjual.
2. *Cluster* 1, Produk Yang Dijual Pada Kategori Diminati Berjumlah 7471 Item Produk Terjual.

5. KESIMPULAN

Penerapan metode data mining menggunakan Algoritma *K-Means Clustering* pada data penjualan Toko Daun Indah dapat memberikan manfaat yang signifikan. Dengan metode ini, dapat mengelompokkan konsumen berdasarkan peminatan produk. Hal ini dapat membantu dalam mengambil keputusan strategi pemasaran yang tepat, sehingga dapat meningkatkan penjualan produk. Berdasarkan hasil penelitian penerapan metode *K-Means* dalam analisis data penjualan produk selama 6 bulan, yaitu dari bulan November 2022 hingga Mei 2023, diperoleh jumlah *cluster* terbaik $K=2$ menunjukkan nilai DBI paling optimal sebesar -0.149 pada *iterasi* ke-1 sebanyak 30 *iterasi*, *type measure Mixed Measures*, setelah dilakukan pengujian dengan $K=2$ sampai $K=25$ menggunakan Software RapidMiner. Dari 18.119 baris data yang telah diolah Data *Preprocessing* menjadi 13.596 baris, dapat disimpulkan bahwa produk yang termasuk kategori kurang diminati terdapat pada *cluster* 0 dengan jumlah 6.125 item, sedangkan produk yang termasuk kategori diminati terdapat pada *cluster* 1 dengan jumlah 7.471 item. Hasil ini dapat menjadi dasar bagi Toko Daun Indah untuk menyusun strategi promosi penjualan produk kecantikan. Toko Daun Indah dapat meningkatkan promosi

untuk produk-produk yang termasuk kategori diminati, sedangkan untuk produk-produk yang termasuk kategori kurang diminati, Toko Daun Indah dapat melakukan evaluasi produk dan strategi penjualannya.

REFERENSI

- [1] Decky Hendarsyah, "E-Commerce Di Era Industri 4.0 Dan Society 5.0," Dec. 2019. doi: <https://doi.org/10.46367/iqtishaduna.v8i2.170>.
- [2] I. Ramadhaniati, "Product Clustering Using K-Means Method in CV. Jaya Abadi," P ISSN, 2023.
- [3] Sharen Kangean and Farid Rusdi, "Analisis Strategi Komunikasi Pemasaran Shopee dalam Persaingan E-Commerce di Indonesia," Oct. 2020. doi: 10.24912/pr.v4i2.6504.
- [4] E. Febrianty, L. Awalina, and W. I. Rahayu, "Optimalisasi Strategi Pemasaran dengan Segmentasi Pelanggan Menggunakan Penerapan K-Means Clustering pada Transaksi Online Retail Optimizing Marketing Strategies with Customer Segmentation Using K-Means Clustering on Online Retail Transactions," *Jurnal Teknologi dan Informasi (JATI)*, vol. 13, 2023, doi: 10.34010/jati.v13i2.
- [5] I. P. Mulyadi, "Clusterisasi Menggunakan Metode Algoritma K-Means dalam Meningkatkan Penjualan Tupperware," *Jurnal Informatika Ekonomi Bisnis*, pp. 172–179, Sep. 2022, doi: 10.37034/infv4i4.164.
- [6] S. Ika Murpratiwi, I. Gusti Agung Indrawan, and A. Aranta, "Analisis Pemilihan Cluster Optimal Dalam Segmentasi Pelanggan Toko Retail," *Jurnal Pendidikan Teknologi dan Kejuruan*, vol. 18, no. 2, 2021.
- [7] Chandra Purnama, Wina Witanti, and Puspita Nurul Sabrina, "Clusterisasi Penjualan Pakaian Untuk Meningkatkan Strategi Penjualan Barang Menggunakan K-Means," 2022. doi: <https://doi.org/10.47292/joint.v4i1.79>.
- [8] Fintri Indriyani and Ani Irfiani, "Clustering Data Penjualan pada Toko Perlengkapan Outdoor Menggunakan Metode K-Means," 2019. [Online]. Available: <https://jurnalnasional.ump.ac.id/index.php/JUITA/article/view/5529>
- [9] Iis Setiawan Mangku Negara, Purwono, and Imam Ahmad Ashari, "Analisa Cluster Data Transaksi Penjualan Minimarket Selama Pandemic Covid-19 dengan Algoritma K-Means," 2021. [Online]. Available: <https://publishing-widyagama.ac.id/ejournal-v2/index.php/jointecs/article/view/2693/1589>
- [10] S. Handoko, F. Fauziah, and E. T. E. Handayani, "Implementasi Data Mining Untuk Menentukan Tingkat Penjualan Paket Data Telkomsel Menggunakan Metode K-Means Clustering," *Jurnal Ilmiah Teknologi dan Rekayasa*, vol. 25, no. 1, pp. 76–88, 2020, doi: 10.35760/tr.2020.v25i1.2677.
- [11] Y. Yun, D. Ma, and M. Yang, "Human–computer interaction-based Decision Support System with Applications in Data Mining," *Future Generation Computer Systems*, vol. 114, pp. 285–289, Jan. 2021, doi: 10.1016/j.future.2020.07.048.
- [12] S. Nurajizah and A. Salbinda, "Penerapan Data Mining Metode K-Means Clustering Untuk Analisa Penjualan Pada Toko Fashion Hijab Banten," *Jurnal Teknik Komputer AMIK BSI*, vol. 7, no. 2, 2021, doi: 10.31294/jtk.v4i2.
- [13] K. Anam, D. Sudrajat, D. A. Kurnia, and N. Masuk, "Analisis Segmentasi Pelanggan Menggunakan Metode K-Means Clustering," *Jurnal ICT: Information Communication & Technology*, vol. 21, pp. 273–278, 2022.

- [14] M. Siahaan, "Data Mining Strategi Pembangunan Infrastruktur Menggunakan Algoritma *K-Means*," *Jurnal Sisfokom (Sistem Informasi dan Komputer)*, vol. 11, no. 3, pp. 316–324, Dec. 2022, doi: 10.32736/sisfokom.v11i3.1453.
- [15] E. Muningsih, I. Maryani, and V. R. Handayani, "Penerapan Metode *K-Means* dan Optimasi Jumlah *Cluster* dengan Index Davies Bouldin untuk *Clustering* Propinsi Berdasarkan Potensi Desa," *Jurnal Sains dan Manajemen*, vol. 9, no. 1, 2021, [Online]. Available: www.bps.go.id
- [16] Tri Wahyudi and Titi Silfia, "Implementation of Data Mining Using *K-Means Clustering* Method to Determine Sales Strategy in S&R Baby Store," 2022. doi: <https://doi.org/10.37385/jaets.v4i1.913>.
- [17] M. Rafi Muttaqin, T. Iman Hermanto, M. Agus Sunandar, P. Studi Teknik Informatika, and S. Tinggi Teknologi Wastukencana, "Penerapan *K-Means Clustering* Dan Cross-Industry Standard Process for Data Mining (CRISP-DM) Untuk Mengelompokkan Penjualan Kue," vol. 19, no. 1, pp. 38–53, 2022, [Online]. Available: <https://journal.unpak.ac.id/index.php/komputasi>
- [18] S. Khairunnisa and M. I. Jambak, "Pengelompokan Cuaca Kota Palembang Menggunakan Algoritma *K-Means Clustering* Untuk Mengetahui Pola Karakteristik Cuaca," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 6, no. 4, p. 2352, Oct. 2022, doi: 10.30865/mib.v6i4.4810.