

PENERAPAN DATA MINING PADA PENJUALAN PRODUK PAKAIAN DAMEYRA FASHION MENGGUNAKAN METODE K-MEANS CLUSTERING

Iwan Pii, Nana Suarna, Nining Rahaningsih

Program Studi Komputerisasi Akuntansi STMIK IKMI Cirebon

Program Studi Rekayasa Perangkat Lunak STMIK IKMI Cirebon

Jl. Perjuangan No. 10B Majasem Kec. Kesambi Kota Cirebon Telp. 0231-490480-490481

iwanpii99@gmail.com

ABSTRAK

Dameyra Fashion adalah toko online yang menjual produk Pakaian lewat E-Commerce Lazada. Dameyra Fashion berlokasi di Kabupaten Cirebon lebih tepatnya di desa arjawinangun. Metode K-Means Clustering bisa diterapkan pada Toko Dameyra Fashion untuk mengetahui Penjualan Pakaian yang laku dan tidak laku. Permasalahan yang dihadapi oleh toko dameyra fashion yaitu masih belum bisa mengelola persediaan barang. Tujuan tugas akhir ini yaitu untuk mengelompokkan data penjualan dameyra fashion menggunakan metode k-means clustering serta bagaimana Penerapan teknik K-Means Clustering oleh Dameyra Fashion Stores Pada RapidMiner, K-Means Clustering bekerja dengan cara membuat dataset dengan memasukkan data penjualan pakaian dari dameyra yang akan diubah. Setelah itu, penambangan data, pemilihan data, pembersihan data, preprocessing data, dan pengetahuan Temuan dari cluster 2. Maka dapat diketahui penjualan yang laris ada 289 item dan tidak laris ada 9 item. Dengan rincian produk laris yaitu Tunik Polos, Tunik Rempel, Tunik Dzuvia, Tunik Leopard, Kemeja Polos dan yang tidak laris Daster Kaos, Daster Kaos OZ, Daster Kimono, Daster Pretty Girl, Daster Kerah V-Neck, Daster Andin Rayon, Daster Batik, Gamis Crinkle, Gamis Benecia Allana.

Kata kunci: Data Mining, K-Means Clustering, RapidMiner, Dameyra Fashion

1. PENDAHULUAN

Dameyra Fashion Merupakan Toko Pakaian yang menjual pakaian secara online melalui aplikasi E-Commerce Lazada. Dameyra Fashion berdomisili di Desa Arjawinangun Kecamatan Arjawinangun Kabupaten Cirebon Provinsi Jawa Barat. Dameyra Fashion menjual aneka ragam jenis pakaian seperti Tunik, Kemeja Polos, Long Dress, Daster dll. Dikarenakan banyaknya ragam produk yang dijual Dameyra Fashion serta tingginya minat belanja konsumen, terkadang membuat Dameyra Fashion Kesulitan dalam pemenuhan stok produk. Penjualan bisa naik dua hingga tiga kali lipat dari biasanya, terutama pada Flash Sale Hari Belanja Nasional 11.11 dan 12.12.

Untuk menyelesaikan masalah tersebut, dibutuhkan pengendalian persediaan barang yang baik agar nantinya tidak terjadi kesulitan dalam pemenuhan stok barang. Untuk melakukan pengendalian persediaan barang, penulis menggunakan data penjualan produk pada periode tertentu yang nantinya digunakan untuk mengelompokkan persediaan barang yang akurat. Penulis menggunakan teknik data mining dan algoritma kmeans clustering untuk mengelompokkan produk yang paling laris saat mengolah data penjualan. Data mining merupakan suatu metode pengolahan data untuk menemukan suatu informasi, dan hasil dari pengolahan datanya dapat digunakan untuk pengambilan suatu keputusan dimasa yang akan datang

Permasalahan yang ditemukan pada Dameyra Fashion Adalah toko sering mengalami kesulitan dalam pengelolaan produk. Toko dameyra fashion

masih belum menemukan cara pengelolaan persediaan produk yang tepat. Pasalnya dengan banyaknya produk yang terjual serta dengan banyaknya variasi produk pada toko dameyra fashion. Dalam hal ini peneliti mencoba mengatasi permasalahan yang saat ini ada pada toko dameyra fashion. Dengan menggunakan alat RapidMiner dan metode pengelompokan k-means, peneliti akan mengklasifikasikan data penjualan fashion dameyra dari Desember 2021 hingga April 2022 untuk mengatasi masalah ini.

Tujuan dari penelitian ini adalah untuk menentukan produk mana yang banyak diminati dan mana yang tidak, yang nantinya dihasilkan oleh output k-means clustering pada tools RapidMiner.

2. TINJAUAN PUSTAKA.

2.1. Data Mining

Proses pengumpulan data historis untuk menemukan hubungan dan keteraturan dalam kumpulan data besar dikenal sebagai penambangan data. Penambangan data atau penambangan data adalah analisis data dengan tujuan membangun koneksi yang jelas dan sebelumnya tidak dikenal antara data dan sumber lain sehingga dapat digunakan dan dipahami.[1]

Menggunakan alat pengenalan pola seperti metode statistik dan matematika, penambangan data adalah proses menemukan hubungan, pola, dan tren yang bermakna dalam sejumlah besar data yang disimpan di memori.[2]

Penambangan data adalah proses menggunakan statistik, matematika, kecerdasan buatan, dan pembelajaran mesin untuk mengekstraksi dan

mengidentifikasi informasi yang berguna dan pengetahuan yang relevan dari berbagai basis data terkait.[3]

2.2. Penjualan

Untuk mencapai keuntungan jangka panjang dan menguntungkan baik bagi pembeli maupun penjual, penjual harus memenuhi semua persyaratan pembeli selama proses penjualan. Dalam dunia bisnis, penjualan juga mengacu pada hasil dari pemberian jasa.[4]

2.3. Pakaian

Pakaian merupakan penutup atau pelindung tubuh terbuat dari bahan tekstil dan berserat. Selain itu, pakaian telah menjadi sarana untuk mengekspresikan identitas, status, atau status pemakainya dari waktu ke waktu.[5]

2.4. Fashion

Kata Sansekerta "bhusana" adalah asal kata "fashion". Namun, konsep berpakaian dan pakaian berbeda. Segala sesuatu yang kita kenakan, dari ujung rambut sampai ujung kaki, adalah pakaian. Kostum ini meliputi pakaian dasar, asesoris (topi dan asesoris), dan tata rias. Pakaian adalah bagian dari pakaian dan termasuk pakaian utama. Pakaian juga, terutama yang menutupi bagian tubuh. padahal busana fashion lebih terkenal atau mapan. Mayoritas orang yang akan mampu memberikan kenyamanan dan membuat sesuatu menjadi lebih baik pada waktu tertentu akan memilih, menerima, menyukai, dan menggunakan fashion, yaitu perpaduan atau perpaduan gaya atau gaya dengan desain. Fashion adalah kombinasi atau perpaduan dari gaya-gaya tersebut.[5]

2.5. Algoritma K-Means

K-Means adalah metode non-hierarkis untuk mengelompokkan data. Tujuannya adalah untuk membagi data yang ada menjadi satu atau lebih cluster atau kelompok sehingga data dengan karakteristik yang mirip berada dalam cluster yang sama dan data dengan karakteristik yang berbeda berada dalam kelompok yang berbeda.[6]

Data dapat dibagi menjadi beberapa kelompok menggunakan teknik pengelompokan berbasis partisi yang dikenal dengan K-Means. Kelompok yang sama diberikan kepada kelompok dengan karakteristik yang sama, sedangkan kelompok yang berbeda diberikan kepada kelompok yang berbeda. Metode K-Means membagi data dengan mengelompokkan objek berdasarkan nilai k yang diberikan untuk memisahkan kelompok-kelompok tersebut.[7]

Algoritma K-Means adalah metode yang relatif mudah untuk mengatur sejumlah besar objek dengan atribut tertentu ke dalam kelompok (cluster) K. Jumlah k cluster ditentukan sebelumnya dalam algoritma K-Means.[3]

2.6. Clustering

Proses menempatkan data atau objek ke dalam kelas atau cluster berdasarkan seberapa mirip atributnya dikenal sebagai clustering. Salah satu metode untuk data mining adalah clustering. Pengelompokan yang baik menghasilkan objek-objek dengan kesamaan rendah dengan yang ada di cluster lain tetapi kesamaan tinggi dengan yang ada di grup atau cluster yang sama.[8]

Metode penambangan data tanpa pengawasan adalah pengelompokan. Dalam proses pengelompokan data, dua jenis pengelompokan data sering digunakan: pengelompokan data non-hierarkis (non-hierarkis) dan pengelompokan data hierarkis (hierarkis).[3]

2.7. RapidMiner

Rapidminer adalah platform sumber terbuka untuk menganalisis data melalui visualisasi, pemodelan prediktif, dan pelaporan. Rapidminer menyediakan alat untuk mengumpulkan, mengekstrak, mengubah, dan menganalisis data, serta membangun model prediktif dan menghasilkan laporan analitik. Dalam pekerjaannya, RapidMiner memberi pengguna wawasan melalui berbagai metode prediktif dan deskriptif sehingga mereka dapat membuat keputusan terbaik. Rapidminer memiliki antarmuka pengguna grafis (GUI) untuk membuat desain pipa analisis. File XML (Extensible Markup Language) yang berisi prosedur analisis yang diinginkan pengguna dihasilkan oleh antarmuka pengguna grafis ini. Rapidminer secara otomatis melakukan analisis pada file ini.[9]

Rapidminer adalah aplikasi penambangan informasi berbasis sumber terbuka utama dan terkenal. Ini termasuk metode untuk pemuatan data, transformasi, pemodelan, dan visualisasi sebagai aplikasi mandiri untuk analisis data dan mesin penambangan data.[10]

3. METODE PENELITIAN

3.1. Teknik Pengumpulan Data

Berikut adalah beberapa metode pengumpulan data:

a. Observasi

Riset lapangan langsung untuk mempelajari situasi dan kondisi toko Dameyra Fashion untuk memudahkan penulis mengumpulkan dan menganalisis data yang dibutuhkan untuk risetnya. Dengan cara ini, penulis melihat langsung di toko dameyra fashion untuk mendapatkan informasi yang tepat. Pemilik toko memberikan referensi untuk penelitian ini dengan memaparkan proses transaksi penjualan di toko dameyra fashion.

b. Wawancara

Dengan mewawancarai pemilik toko dan operator pengepakan, data dapat dikumpulkan melalui metode wawancara. Tujuan dari wawancara ini adalah untuk mengumpulkan informasi yang akurat. Peneliti menggunakan strategi ini untuk mengumpulkan data dengan mengajukan

pertanyaan yang harus dijawab secara lisan. Penulis melakukan wawancara dengan karyawan Dameyra Fashion Store untuk mengumpulkan informasi tersebut.

c. Studi Pustaka

membaca buku dan referensi internet untuk mengumpulkan informasi yang dapat dijadikan referensi saat membahas masalah ini. Dalam strategi ini analis mencari informasi atau data penelitian melalui catatan harian logis, buku referensi dan bahan distribusi lain yang dapat diakses di perpustakaan. Untuk membantu sistem pengelompokan penjualan yang dikembangkan pada toko dameyra fashion, penulis mencari referensi dan penunjang penelitian berupa buku atau contoh sistem pengelompokan penjualan yang sejenis atau yang sudah ada.

d. Studi Dokumentasi

Untuk memperoleh data yang lengkap, seperti dokumen latar belakang dan kegiatan yang telah diselesaikan, maka dilakukan studi dokumentasi dalam penelitian ini. Informasi yang akan dijadikan sebagai sumber penunjang penelitian dapat diperoleh dengan menggunakan pendekatan ini. Penulis menggunakan strategi ini untuk melakukan pencatatan mengenai proses penjualan pada toko Dameyra Fashion. Metode dokumentasi ini menambah teknik observasi dan wawancara.

3.2. Sumber Data

a. Sumber Data Penelitian

Seorang peneliti akan dihadapkan pada metodologi penelitian untuk penelitian ini yaitu mengelompokkan produk yang penjualannya paling banyak (laku) dengan produk yang penjualannya sedikit (tidak laku). Peneliti mendapatkan sumber data penjualan periode Desember 2021 – April 2022 langsung dari pemilik Dameyra Fashion melalui wawancara. Dameyra Fashion yang berdomisili di Arjawinangun, Kecamatan Arjawinangun, Kabupaten Cirebon Minggu, 22 Januari 2023 Pukul 13.00 WIB.

b. Sumber Data Primer

Data Primer Peneliti mengumpulkan data primer langsung dari lapangan. Data yang berasal langsung dari sumber aslinya disebut data primer. Pemilik toko fashion di Dameyra menjadi subjek penyelidikan ini.

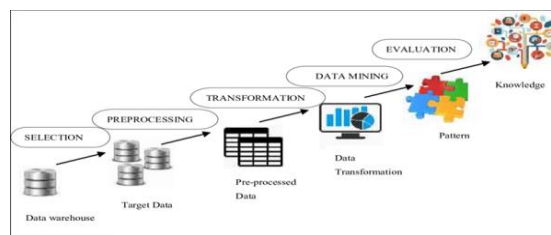
c. Sumber Data Sekunder

Data Sekunder Sumber data sekunder digunakan untuk mengumpulkan data penelitian. Berbagai informasi yang sudah ada sebelumnya yang sengaja dikumpulkan oleh peneliti untuk memenuhi kebutuhan data penelitian disebut sebagai data sekunder.

3.3. Tahapan Perancangan

Tahap Knowledge Discovery in Databases (KDD) digunakan oleh peneliti selama tahap desain penelitian ini. Menemukan informasi yang bermakna

dari sejumlah besar data dikenal sebagai KDD. Langkah-langkah KDD yaitu sebagai berikut:



Gambar 1. Tahapan Perancangan

1. Data Selection (Seleksi Data)

Sebelum diintegrasikan ke dalam satu database dan gudang, prosedur ini digunakan untuk menghapus data yang berisik dan tidak konsisten dari data yang terdapat di banyak database. Basis data ini mungkin memiliki format atau platform yang berbeda.

2. Data Transformation (Transformasi Data)

Data database kemudian diubah menggunakan berbagai metode. Khususnya ketika berhadapan dengan masalah skala besar, prosedur pemilihan sangat penting untuk mendapatkan hasil yang lebih tepat dan mempercepat perhitungan. Agar data yang telah diproses siap untuk ditambang, diperlukan transformasi data sebagai langkah preprocessing. Di antara metode transformasi adalah:

- A. Centering, Dimana setiap data direduksi dengan rata-rata dari berbagai atributnya.
- B. Normalisasi, Dimana setiap data desentralisasi dibagi dengan standar deviasi atribut.
- C. Scaling, Mengubah data ke ukuran tertentu melalui penskalaan

3. Data Mining (Penambangan Data)

Proses penambangan data atau data mining dapat dilakukan setelah data yang telah diseleksi dan ditransformasikan ditambang dengan menggunakan berbagai macam metode. Selama proses penambangan data, fungsi-fungsi khusus harus digunakan untuk mencari pola atau informasi menarik dalam data yang dipilih. Tujuan dan keseluruhan proses pencarian pengetahuan sangat mempengaruhi pemilihan fungsi atau algoritma yang sesuai.

4. Evaluation (Evaluasi)

Tahap ini melibatkan penggambaran visualisasi, penyajian pengetahuan dalam bentuk yang mudah dipahami oleh pengguna, dan menentukan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesis yang ada sebelumnya.

5. Knowledge (Pengetahuan)

Memvisualisasikan data yang diproses untuk memudahkan pengguna memahami dan mengambil tindakan berdasarkan analisis adalah langkah terakhir dalam proses KDD. Akibatnya, salah satu langkah penting dalam proses penambangan data adalah menyajikan hasil

dengan cara yang dapat dipahami semua orang. Hasil data mining dengan menggunakan algoritma K-Means juga dapat dikomunikasikan melalui visualisasi dalam presentasi ini.

4. HASIL DAN PEMBAHASAN

4.1. Hasil Penelitian

Proses penggunaan Algoritma K-Means untuk mengklasifikasikan atau mengelompokkan dataset penjualan pakaian akan dijelaskan dalam temuan pembahasan ini. Pengelompokkan ini dilakukan dengan proses pengujian menggunakan machine learning yaitu *Software RapidMiner Studio*.

4.2. a. Data

Dataset yang digunakan dalam penelitian ini memiliki 6 atribut dan 398 record. diantaranya adalah: no, nama produk, keterangan, tanggal, pesanan, unit terjual. Dataset tersebut bersumber dari toko dameyra *fashion*. Data yang diterima dalam bentuk dokumen *soft file* yang disajikan dalam bentuk table 1 dibawah ini:

Tabel 1. Data Penjualan Produk Pakaian

No	Nama Produk	Kategori	Tanggal	Pesanan	Unit Terjual
1	Tunik Rempel	Pakaian Wanita	01-Des	21	21
2	Tunik Polos	Pakaian Wanita	01-Des	57	65
3	Tunik Polos	Pakaian Wanita	01-Des	65	80
4	Tunik Polos	Pakaian Wanita	01-Des	80	80
5	Tunik Dzuvia	Pakaian Wanita	01-Des	79	95
6	Tunik Rempel	Pakaian Wanita	01-Des	95	95
7	Tunik Rempel	Pakaian Wanita	01-Des	101	110
8	Tunik Leopard	Pakaian Wanita	01-Des	97	109
9	Tunik Leopard	Pakaian Wanita	01-Des	91	91
10	Kemeja Crinkle	Pakaian Wanita	01-Des	60	75
11	Daster Ruffle	Pakaian Wanita	01-Des	76	76
12	Tunik Polos	Pakaian Wanita	01-Des	80	80
13	Janda Bolong	Pakaian Wanita	01-Des	56	67
14	Tunik Polos	Pakaian Wanita	02-Des	145	155
15	Kemeja Polos	Pakaian Wanita	02-Des	19	28
...
398	Tunik Leopard	Pakaian Wanita	30-Apr	50	59

4.3. b. Memilih Data

Data penjualan pakaian untuk bulan Desember 2021 hingga April 2022 dipilih untuk penelitian ini. Informasi ini memiliki 398 record dan 6 atribut. Dua atribut merupakan hasil seleksi yang tidak dimanfaatkan. karena masing-masing dari keempat atribut memiliki nilai yang sama.

Tabel 2. Seleksi Atribut

No	Nama Atribut	Type
1	No	Integer
2	Tanggal	Integer

4.4. c. Data Preprocessing

Data Penjualan Pakaian Dameyra Fashion sebelum Preprocessing.

Data Penjualan Pakaian ini termasuk data primer yang artinya data langsung diambil dari pemilik toko dameyra *fashion*. Tabel 3 memuat informasi tersebut.

Tabel 3. Data Sebelum Preprocessing

No	Nama Produk	Kategori	Tanggal	Pesanan	Unit Terjual
1	Tunik Rempel	Pakaian Wanita	01-Des	21	21
2	Tunik Polos	Pakaian Wanita	01-Des	57	65
3	Tunik Polos	Pakaian Wanita	01-Des	65	80
4	Tunik Polos	Pakaian Wanita	01-Des	80	80
5	Tunik Dzuvia	Pakaian Wanita	01-Des	79	95
6	Tunik Rempel	Pakaian Wanita	01-Des	95	95
7	Tunik Rempel	Pakaian Wanita	01-Des	101	110
8	Tunik Leopard	Pakaian Wanita	01-Des	97	109
9	Tunik Leopard	Pakaian Wanita	01-Des	91	91
10	Kemeja Crinkle	Pakaian Wanita	01-Des	60	75
11	Daster Ruffle	Pakaian Wanita	01-Des	76	76
12	Tunik Polos	Pakaian Wanita	01-Des	80	80
13	Janda Bolong	Pakaian Wanita	01-Des	56	67
14	Tunik Polos	Pakaian Wanita	02-Des	145	155
15	Kemeja Polos	Pakaian Wanita	02-Des	19	28
...
398	Tunik Leopard	Pakaian Wanita	30-Apr	50	59

4.5. d. Data Penjualan Pakaian Setelah Preprocessing

Menghapus atau menghilangkan data yang tidak penting dalam melakukan data mining ini

Tabel 4. Data Sesudah Preprocessing

Nama Produk	Kategori	Pesanan	Unit Terjual
Tunik Rempel	Pakaian untuk wanita	21	21
Tunik Polos	Pakaian untuk wanita	57	65
Tunik Polos	Pakaian untuk wanita	65	80
Tunik Polos	Pakaian untuk wanita	80	80
Tunik Dzuvia	Pakaian untuk wanita	79	95
Tunik Rempel	Pakaian untuk wanita	95	95
Tunik Rempel	Pakaian untuk wanita	101	110
Tunik Leopard	Pakaian untuk wanita	97	109
Tunik Leopard	Pakaian untuk wanita	91	91
Kemeja Crinkle	Pakaian untuk wanita	60	75
Daster Ruffle	Pakaian untuk wanita	76	76
Tunik Polos	Pakaian untuk wanita	80	80
Janda Bolong	Pakaian untuk wanita	56	67
Tunik Polos	Pakaian untuk wanita	145	155
Kemeja Polos	Pakaian untuk wanita	19	28
Tunik Polos	Pakaian untuk wanita	91	91
...
Tunik Leopard	Pakaian untuk wanita	50	59

4.6. d. Data Transformation

Transformasi data dilakukan untuk mengubah data bertipe nominal menjadi numerik dengan menggunakan operator *nominal to numerical*.

Row No.	Nama Produk	Kategori = P...	Pesanan	Unit Terjual
1	Tunik Rempel	1	21	21
2	Tunik Polos	1	57	65
3	Tunik Polos	1	65	80
4	Tunik Polos	1	80	80
5	Tunik Dzuvia	1	79	95
6	Tunik Rempel	1	95	95
7	Tunik Rempel	1	101	110
8	Tunik Leopard	1	97	109
9	Tunik Leopard	1	91	91
10	Kemeja Crink...	1	60	75
11	Daster Ruffle	1	76	76
12	Tunik Polos	1	80	80
13	Janda Bolong	1	56	67
14	Tunik Polos	1	145	155
15	Kemeja Polos	1	19	28

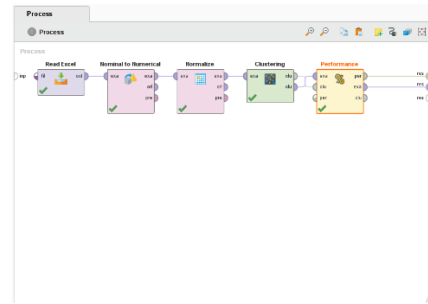
Gambar 2. Transformasi Data

Penjelasan:

Hasil proses dari rapidminer untuk operator terdiri dari 398 example, 1 special attribute, 3 regular attributes diantaranya kategori produk, pesanan, unit terjual.

4.7. Data Mining

Dengan menggunakan algoritma k-means, data mining adalah proses pencarian pola atau informasi yang menarik pada data tertentu.



Gambar 3. Proses Clustering

Pada gambar 3 peneliti menggunakan *algoritma k-means*. Untuk mendapatkan operator yang diinginkan, pada menu operator ketik k-means jika sudah menemukan maka drag operator k-means tersebut. Parameter yang digunakan adalah k=2 dan max run=10. K adalah cluster, jadi pada parameter tersebut peneliti menggunakan 2 cluster yaitu penjualan laris dan penjualan tidak laris dan max run = 10 cluster.

4.8. f. Parameter K-Means

Tabel 5. Parameter K-Means

Parameter	Keterangan
k min	2
k max	10
measure type	Numerical Measurement
numerical measure	Euclidean Distance
algorithm clustering	K-Means

Penjelasan:

Hasil Parameter K-means, k min =2, k max = 10 adalah untuk menguji k-means pada langkah ke 2 untuk menghasilkan nilai terbaik *Davies Bouldin Index* yang mendekati angka 0.

4.9. Evaluation

Untuk mengetahui apakah hasil dari teknik data mining telah terealisasi maka pada tahap ini dievaluasi dalam bentuk skema dan model tipikal. Jika hasil yang diperoleh tidak sesuai, ada beberapa opsi, termasuk menerima hasil sebagai hasil yang tidak terduga dan berpotensi bermanfaat atau menggunakannya sebagai umpan balik untuk meningkatkan ekstraksi data lain yang lebih sesuai.

Alur informasi proses, atau aliran yang dihasilkan oleh penambangan data, harus disajikan dengan cara yang mudah dipahami oleh pihak yang berkepentingan. Ini adalah proses KDD yang mencakup penentuan apakah aliran atau informasi yang ditemukan bertentangan dengan fakta atau hipotesis evaluasi sebelumnya.

Dari 398 data penjualan pakaian, empat kluster diidentifikasi menggunakan temuan Software RapidMiner. 4 cluster diperoleh setelah dilakukannya

percobaan dari cluster 2 sampai cluster 10 untuk mencari nilai *Davies Bouldin Indeks*. Setelah melakukan percobaan menghasilkan nilai k optimum k=2 dengan nilai DBi = -0,414. Dapat diartikan bahwa tingkat kemiripan data dalam cluster 2 yaitu -0,414. Berikut rekapitulasi 9 cluster yang diperoleh:

Tabel 6. Cluster dan DBi

No	Cluster	Dbi
1	2	-0,414
2	3	-0,537
3	4	-0,499
4	5	-0,48
5	6	-0,426
6	7	-0,426
7	8	-0,422
8	9	-0,416
9	10	-0,445

4.10. Pembahasan

Algoritma pengelompokan k-means dan tampilan proses digunakan untuk mengelompokkan dataset penjualan pakaian dengan RapidMiner 9.8 dalam penelitian ini.

a. Pengolahan Data

Langkah clustering dalam proses pengolahan data harus dilakukan dua kali yaitu langkah pertama membagi data menjadi dua cluster berdasarkan jenis atribut, pesanan dan unit terjual, dan langkah kedua membagi data menjadi dua cluster berdasarkan jenis atribut, nama produk dan deskripsi. Algoritma k-means clustering dapat digunakan untuk mengelompokkan data setelah semuanya diubah menjadi angka.

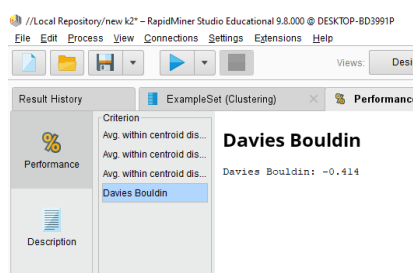
Secara khusus, prosedur untuk mengevaluasi pengujian, yang menghasilkan daftar nilai kriteria berdasarkan centroid cluster. yang membuatnya berguna untuk memahami seberapa baik prosedur pengelompokan bekerja. Davies Bouldin Index (dbi) dari peringkat RapidMiner berfungsi sebagai dasar evaluasi klaster dalam penyelidikan ini. Di sisi lain, jika Davies Bouldin Index (dbi) menghasilkan nilai yang lebih tinggi, proses evaluasi cluster semakin buruk atau tidak sebaik yang seharusnya.

b. Proses Davies Bouldin

Dari percobaan menggunakan nilai parameter diatas, berikut hasil percobaan yang dilakukan:

1. Data Cluster 2

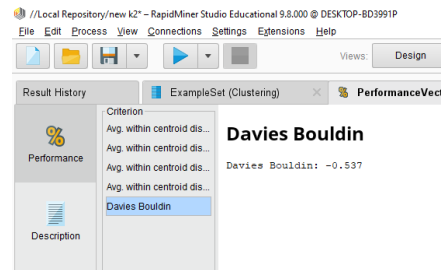
Pada k=2 diperoleh nilai *Davies Bouldin* : -0,414



Gambar 4. Meta Data Cluster 2

2. Data Cluster 3

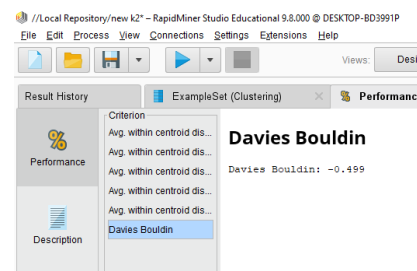
Pada k=3 diperoleh nilai *Davies Bouldin*: - 0.537



Gambar 5. Meta Data Cluster

3. Data Cluster 4

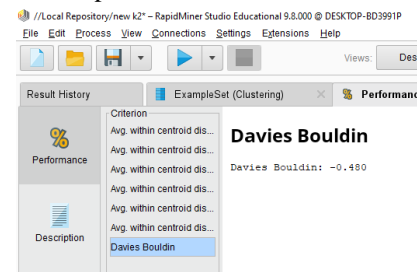
Pada k=4 diperoleh nilai *Davies Bouldin*: -0.499



Gambar 6. Meta Data Cluster 4

4. Data Cluster 5

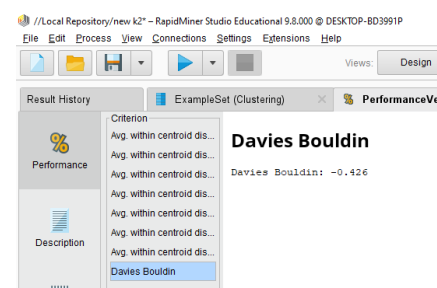
Pada k=5 diperoleh nilai *Davies Bouldin*: -0.480



Gambar 7. Meta Data Cluster 5

5. Data Cluster 6

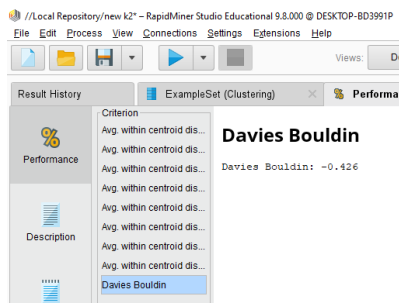
Pada k=6 diperoleh nilai *Davies Bouldin*: -0.424



Gambar 8. Meta Data Cluster 6

6. Data Cluster 7

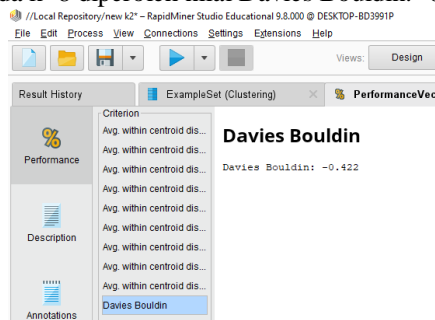
Pada k=7 diperoleh nilai *Davies Bouldin*: -0.424



Gambar 9. Meta Data Cluster 7

7. Data Cluster 8

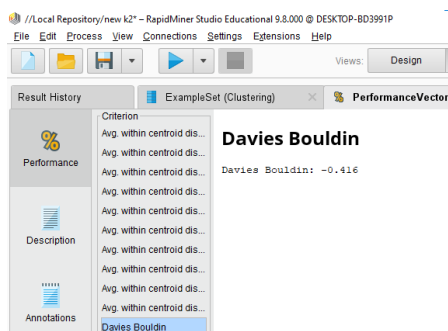
Pada $k=8$ diperoleh nilai Davies Bouldin: -0.422



Gambar 10. Meta Data Cluster 8

8. Data Cluster 9

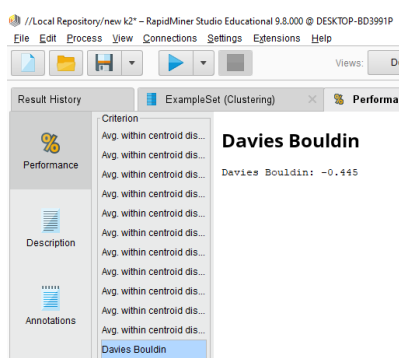
Pada $k=9$ diperoleh nilai Davies Bouldin: -0.416



Gambar 11. Meta Data Cluster 9

9. Data Cluster 10

Pada $k=9$ diperoleh nilai Davies Bouldin: -0.445



Gambar 12. Meta Data Cluster 10

5. KESIMPULAN DAN SARAN

Berikut adalah kesimpulan penulis terkait hasil penelitian setelah tahapan k-means clustering: Setelah melaksanakan proses *Cluster distance performance* maka dihasilkan *K-means clustering* $k=10$, dapat menghasilkan *davies bouldin* dengan urutan $k_2=0,414$, urutan $k_3=0,537$, urutan $k_4=0,499$, urutan $k_5=0,480$, urutan $k_6=0,426$, urutan $k_7=0,426$, urutan $k_8=0,422$, urutan $k_9=0,416$, urutan $k_{10}=0,445$, maka kesimpulan dari proses *Cluster distance performance* menghasilkan dbi yang terbaik yaitu 0,414 pada urutan ke 2.

Saran untuk peneliti lain: Alangkah baiknya pada penelitian berikutnya menggunakan data berdasarkan data penjualan toko lain dan sebaiknya menggunakan atribut yang lebih banyak lagi agar penelitian menjadi lebih baik. Dalam penentuan pengelompokan data penjualan selanjutnya sebaiknya dilakukan dengan metode yang lain selain pengujian menggunakan *algoritma k-means clustering*.

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] A. P. Natasuwarna, "Tantangan menghadapi era revolusi 4.0 - Big data and data mining," *Semin. Nas. Has. Pengabd. Kpd. Masy.*, pp. 23–27, 2019.
- [3] N. K. Surbakti, "Data Mining Pengelompokan Pasien Rawat Inap Peserta BPJS Menggunakan Metode Clustering (Studi Kasus : RSUD.Bangkalan)," *J. Inf. Technol.*, vol. 1, no. 2, pp. 47–53, 2021, doi: 10.32938/jitu.v1i2.1470.
- [4] R. F. Ahmad and N. Hasti, "Sistem Informasi Penjualan Sandal Berbasis Web," *J. Teknol. dan Inf.*, vol. 8, no. 1, pp. 67–72, 2018, doi: 10.34010/jati.v8i1.911.
- [5] A. Munif, "Borongan Menurut Prespektif Hukum Islam," *J. ALSYIRKAH (Jurnal Ekon. Syariah)*, vol. 2, no. 2, pp. 46–51, 2021.
- [6] Z. Nabila, A. Rahman Isnain, and Z. Abidin, "Analisis Data Mining Untuk Clustering Kasus Covid-19 Di Provinsi Lampung Dengan Algoritma K-Means," *J. Teknol. dan Sist. Inf.*, vol. 2, no. 2, p. 100, 2021, [Online]. Available: <http://jim.teknokrat.ac.id/index.php/JTSI>
- [7] F. Nurdiyansyah and I. Akbar, "Implementasi Algoritma K-Means untuk Menentukan Persediaan Barang pada Poultry Shop," *J. Teknol. dan Manaj. Inform.*, vol. 7, no. 2, pp. 86–94, 2021, doi: 10.26905/jtmi.v7i2.6377.
- [8] D. A. I. C. Dewi and D. A. K. Pramita, "Analisis Perbandingan Metode Elbow dan Silhouette pada Algoritma Clustering K-Medoids dalam Pengelompokan Produksi Kerajinan Bali," *Matrix J. Manaj. Teknol. dan Inform.*, vol. 9, no.

- 3, pp. 102–109, 2019, doi: 10.31940/matrix.v9i3.1662.
- [9] N. Anggraini, J. Jasmir, and P. A. Jusia, “Penerapan Metode K-Means Clustering untuk Menentukan Persediaan Stok Barang Pada Toko Pensmart Jambi,” *J. Ilm. Mhs. ...*, vol. 1, pp. 63–77, 2019.
- [10] R. Nofitri and N. Irawati, “Integrasi Metode Neive Bayes Dan Software Rapidminer Dalam Analisis Hasil Usaha Perusahaan Dagang,” *JURTEKSI (Jurnal Teknol. dan Sist. Informasi)*, vol. 6, no. 1, pp. 35–42, 2019, doi: 10.33330/jurteks.v6i1.393.