IMPLEMENTASI DATA MINING MENGGUNAKAN ALGORITMA K- MEANS CLUSTERING DALAM ANALISIS PENJUALAN PRODUK

Talitha Tiara Alifa¹, Rini Astuti², Fadhil Muhamad Basysyar³

¹ Teknik Informatika, STMIK IKMI Cirebon
² Sistem Informasi, STMIK LIKMI Bandung
³ Sistem Informasi, STMIK IKMI Cirebon
Jalan Perjuangan No.10B, Majasem, Cirebon
talitha060902@gmail.com

ABSTRAK

Fashion pria beroperasi di sebuah Department Store X dan fokus pada penjualan pakaian pria dengan berbagai macam produk seperti Kemeja, T-shirt, Celana pendek, Celana Panjang dan produk sebagainya. menghadapi tantangan memanfaatkan informasi penjualan yang melimpah untuk meningkatkan strategi penjualan. Oleh karena itu, penelitian ini menggunakan metode clustering K-Means untuk mengelompokkan data penjualan dan menentukan produk pakaian mana yang laris dan produk mana yang kurang laris di pasar mode pria. Dengan demikian, penelitian ini dapat mempermudah mengidentifikasi produk. Algoritma clustering K-Means digunakan untuk membuat kelompok produk yang saling terkait dari data penjualan pakaian pria.. Dalam konteks ini, informasi yang dimiliki, seperti Nama Barang (Artikel), Stok Awal, dan Produk yang Terjual, digunakan sebagai atribut dalam analisis K-Means Clustering. Tujuan dari studi ini adalah untuk memperdalam pemahaman terhadap pola penjualan produk Fashion Pria, sehingga hasil analisis dapat memberikan wawasan yang berharga dalam mengelola data penjualan. Hasil dari penerapan K-Means Clustering diharapkan dapat meningkatkan pemahaman tentang tren penjualan, membantu mengidentifikasi produk yang paling diminati, dan memberikan dasar untuk mengoptimalkan strategi pemasaran. Dengan demikian, penelitian ini berperan penting dalam mengubah informasi menjadi pengetahuan yang dapat meningkatkan kinerja penjualan Fashion Pria. Kesimpulan dari penelitian ini memberikan landasan untuk pengembangan strategi bisnis yang lebih cerdas dalam menghadapi dinamika pasar yang terus berubah.

Kata kunci: K Means Clustering, Data Mining, Fashion Pria

1. PENDAHULUAN

Fashion pria yang berada di sebuah Department store X, sedang berusaha memahami dan meningkatkan pola penjualan produk mereka. Meskipun mereka menghadapi peningkatan jumlah data penjualan seiring dengan kemajuan teknologi dan pertumbuhan bisnis e-commerce, memanfaatkannya secara efisien untuk mendapatkan wawasan yang berharga menjadi suatu tantangan. Oleh karena itu, dibutuhkan pendekatan yang lebih canggih, seperti menerapkan metode K-Means Clustering dalam konteks data mining.

Fashion Pria menghadapi beberapa tantangan utama, termasuk kesulitan dalam mengenali pola penjualan yang berpengaruh, keterbatasan dalam memahami preferensi konsumen secara menyeluruh, dan kendala dalam manajemen efisien stok produk. Pendekatan analisis konvensional mungkin kurang memadai mengingat kompleksitas pertumbuhan data penjualan yang terus meningkat. Oleh karena itu, dibutuhkan pendekatan lebih canggih seperti penerapan metode K-Means Clustering.

Dalam penelitian yang dilakukan oleh [1] berjudul "Penerapan Data Mining Metode K-Means Clustering untuk Analisis Penjualan pada Toko Fashion Hijab Banten," menemukan bahwa metode K-Means dapat digunakan untuk menemukan penjualan pakaian yang sangat laris, laris, dan kurang laris di toko Helai. Data stok pakaian dikelompokkan menggunakan tiga

klaster yang dipilih secara acak sebagai centroid awal selama proses implementasi. Setelah data dalam setiap klaster stabil, hasil akhir menunjukkan bahwa ada sebelas artikel yang sangat laris, lima puluh lima artikel yang laris, dan tiga puluh empat artikel yang kurang laris. Dalam studi yang dilakukan oleh [2] Data berjudul "Implementasi Mining Menentukan Tingkat Penjualan Paket Data Telkomsel Metode K-Means Menggunakan menemukan bahwa algoritma K-Means Clustering memiliki kemampuan untuk membedakan wilayah penjualan produk yang rendah, sedang, dan tinggi. Wilayah penjualan rendah dapat diatasi dengan memberikan promosi penjualan, sementara wilayah penjualan tinggi tidak perlu dipromosikan.

Diharapkan bahwa aplikasi ini akan membantu mengelola data transaksi penjualan dengan lebih cepat, menghemat waktu, dan membantu dalam membangun strategi untuk meningkatkan penjualan paket data.. Penelitian oleh [3] berjudul "Penerapan Data Mining untuk Prediksi Penjualan Mobil Clustering K-Means" Menggunakan Metode menunjukkan bahwa data penjualan mobil dapat dikelompokkan berdasarkan jumlah penjualan selama lima tahun. Hasilnya adalah pembentukan tiga klaster: Sangat Laris, Kurang Laris, dan Laris. Perusahaan mobil yang berada di Klaster Kurang dapat menggunakan strategi meningkatkan penjualan mereka.

otomotif dapat melihat tingkat penjualan mobil mereka dibandingkan dengan perusahaan lain dengan menggunakan teknik clustering ini. Dengan menggunakan indeks Davies Bouldin (DBI), evaluasi menunjukkan hasil yang konsisten dan akurat dalam membagi penjualan mobil ke dalam tiga klaster.

Dengan menerapkan metode Clustering, Tujuan dari studi ini adalah untuk memperdalam pemahaman terhadap pola penjualan produk Fashion Pria, sehingga hasil analisis dapat memberikan wawasan yang berharga dalam mengelola data penjualan. Dengan pemahaman yang lebih baik terhadap perbedaan ini, diharapkan dapat ditingkatkan pengelolaan stok, pemahaman mendalam terkait preferensi pelanggan, pemantauan tren penjualan produk Fashion Pria.

Studi ini menggunakan pendekatan analisis *K-Means Clustering*. Keputusan untuk menggunakan metode ini didasarkan pada kemampuannya untuk mengelompokkan data menjadi segmen yang signifikan berdasarkan karakteristik yang serupa. Dengan menerapkan metode *K-Means Clustering* pada data penjualan *Fashion* Pria, diharapkan dapat diidentifikasi kelompok produk dengan pola penjualan serupa. Ini memberikan kesempatan untuk mempelajari preferensi pelanggan secara lebih mendalam, meningkatkan efisiensi strategi pengelolaan stok, serta mengenali produk-produk dengan tingkat penjualan yang berbeda-beda.

Harapannya, hasil dari penerapan *K-Means Clustering* akan memberikan pemahaman yang lebih mendalam terkait tren penjualan, membantu mengenali produk yang laris dan kurang laris. Dengan demikian, penelitian ini memiliki peran yang signifikan dalam mengubah informasi menjadi pengetahuan yang dapat meningkatkan kinerja penjualan *Fashion* Pria. Pemahaman yang lebih mendalam terhadap pola penjualan diharapkan dapat memberikan wawasan berharga bagi *Fashion* pria dalam membuat keputusan strategis terkait manajemen stok penjualan.

2. TINJAUAN PUSTAKA

2.1. Data Mining

Data mining adalah proses ekstraksi pola yang menarik dari sejumlah besar data. Jika suatu strategi menarik, itu harus mudah dipahami, tidak sepele, tidak diketahui sebelumnya, dan berguna. Selain itu, strategi harus berlaku untuk data yang akan diprediksi dengan tingkat kepastian tertentu, berguna, dan baru.[4]

2.2. Clustering

Menurut Metisen dan Sari (2015), klasifikasi atau clustering adalah teknik yang digunakan untuk membagi rangkaian data menjadi beberapa kelompok berdasarkan kesamaan yang telah ditentukan sebelumnya. Cluster adalah sekelompok atau sekumpulan objek data yang sangat mirip satu sama

lain dalam kumpulan yang sama dan tidak mirip dengan kumpulan yang berbeda.[5]

2.3. Penjualan

Penjualan Sistem penjualan melibatkan sumber daya organisasi, prosedur, data, dan sarana pendukung untuk menjalankan sistem penjualan, sehingga memberikan informasi yang membantu manajemen membuat keputusan.[6]

2.4. Produk

Suatu produk adalah kumpulan atribut yang nyata dan tidak nyata, seperti kemasan, warna, harga, kualitas, dan merk, ditambah dengan jasa dan reputasi penjual, menurut Stanton (1996:222).[7]

2.5. Davies bouldin index (DBI)

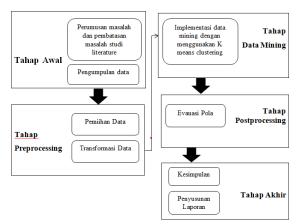
Salah satu metode evaluasi yang dapat digunakan untuk mengukur keakuratan atau akurasi cluster dalam penelitian adalah *Davis Bouldin Index (DBI)*. Pengujian DBI dilakukan dengan mengukur kekuatan atau keakuratan jumlah K (cluster) yang dibentuk menggunakan algoritma K-Means dan K-Medoids.[8]

2.6. K- Means

K-Means membagi data ke dalam banyak cluster dengan pendekatan berbasis jarak. Khususnya, algoritma ini efektif pada data numerik karena hanya bekerja pada atribut numerik atau angka. Salah satu fitur algoritma K-Means adalah partitioning clustering, yang memisahkan data menjadi beberapa bagian terpisah. Algoritma K-Means terkenal karena kemudahan dan kecepatan penggabungan data yang besar dan outlier dengan cepat. Algoritma K-Means memastikan bahwa setiap data termasuk dalam cluster tertentu. Selain itu, adalah mungkin bagi setiap data yang termasuk dalam cluster tertentu pada satu tahap proses untuk berpindah ke cluster lain pada tahap berikutnya.[9]

3. METODE PENELITIAN

3.1. Tahapan penelitian



Gambar 1. Tahap Metode Penelitian

KDD menghasilkan pengetahuan dari database yang ada. Pengetahuan ini akan digunakan oleh basis pengetahuan atau pengambilan keputusan. [10] Namun, data mining adalah komponen dari tahapan proses KDD, seperti yang ditunjukkan pada gambar 1.

3.1.1. Tahap Awal

a. Perumusan masalah dan pembatasan masalah studi literature

Berdasarkan konteks latar belakang penelitian yang telah dijelaskan, perumusan masalah menjadi langkah selanjutnya. Pembatasan masalah diperlukan untuk memfokuskan pembahasan penelitian. Setelah itu, tahap berikutnya adalah meninjau literatur dari berbagai sumber yang terkait dengan masalah penelitian untuk mendapatkan basis teori yang kuat..

b. Pengumpulan Data

Data akan dikumpulkan sebagai langkah awal, dan selanjutnya akan diolah pada tahapan berikutnya.

3.1.2. Tahap Preprocessing

a. Pemilihan Data

Atribut-atribut yang memiliki relevansi untuk proses data mining akan dipilih, sementara atribut-atribut yang tidak relevan akan dihapus dari perhitungan.

b. Transformasi Data

Data akan mengalami transformasi atau mengubahnya menjadi format yang sesuai untuk diproses selama proses data mining.

3.1.3. Tahap Data Mining

K means clustering digunakan untuk menerapkan data mining: Dalam tahap data mining, proses pengolahan data dilakukan dengan menerapkan metode *clustering* menggunakan perangkat lunak pengolah data, yaitu *RapidMiner*. Perangkat lunak tersebut menggunakan algoritma kmeans *clustering*.

3.1.4. Tahap Postprocessing

Evaluasi Pola: Proses identifikasi dan evaluasi dilakukan terhadap informasi, pola yang signifikan, atau temuan menarik yang ditemukan melalui hasil data mining.

3.1.5. Tahap Akhir

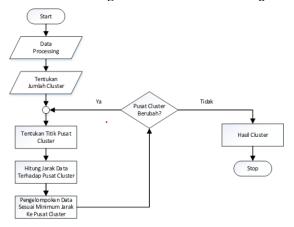
a. Kesimpulan

Pada langkah ini, dibuat kesimpulan berdasarkan hasil penelitian.

b. Penyusunan Laporan

Dokumentasi hasil penelitian dinyatakan dalam bentuk laporan skripsi sebagai wujud laporan akhir dari penelitian tersebut.

3.2. Flowchart Algoritma K Means Clustering



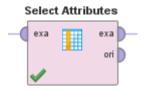
Gambar 2. Flowchart Algoritma K Means Clustering

Gambar 2 di bawah ini menunjukkan flowchart yang menunjukkan langkah-langkah metode K-means dengan ketentuan berikut.

- 1. Tentukan jumlah cluster yang akan digunakan,
- Tentukan titik pusat cluster dari data yang sudah ditambahkan,
- 3. Setelah itu, hitung jarak terdekat tiap cluster dengan menggunakan proses jarak geometris
- 4. Kemudian mengelompokkan data sesuai dengan jarak terdekat ini ,
- 5. Jika nilai jarak terdekat atau nilai hasil cluster berubah maka perhitungan akan diulang lagi jika tidak berubah maka perhitungan akan diulang lagi jika tidak berubah maka perhitungan akan diulang lagi jika.

4. HASIL DAN PEMBAHASAN

Penulis menggunakan metode clustering K Means dalam penelitian ini. melalui aplikasi *RapidMiner*. Data yang menjadi fokus penelitian adalah data stok fashion pria dari bulan Januari hingga April 2022. Proses data mining menggunakan metode K Means Clustering ini mengikuti tahapan Knowledge Discovery Database (KDD), yang mencakup pengumpulan data, pemilihan data, transformasi data, dan data mining.



Gambar 3. Operator Select Attributes

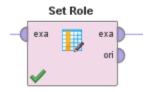
Pada bagian Select Attribute untuk menentukan kolom atau atribut khusus dalam dataset yang akan digunakan dalam proses analisis atau pemodelan data. Dengan menggunakan fitur ini, dapat menyaring dan memilih variabel yang dianggap relevan atau kritikal untuk mencapai tujuan analisis atau pembuatan

model. Ini membantu menyederhanakan dataset dengan fokus pada informasi yang memiliki signifikansi tertinggi, mengurangi kompleksitas data.



Gambar 4. Menyeleksi Atribut

Kemudian jika sudah berada dalam box select atribut, dilakukan pemilihan data atribut mana saja yang akan dipakai. Adapun atribut yang akan dipakai meliputi nama barang, bulan penjualan yaitu bulan januari, febuari, maret, dan April. dan untuk data atribut yang tidak terpakai yaitu Artikel (kode barang), Nama item, dan Stok awal.



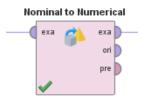
Gambar 5. Operator Set role

menyeleksi Setelah atribut yang akan digunakan, langkah berikutnya adalah menetapkan atribut tersebut sebagai id dalam rangka mempermudah analisis visualisasi statistik. Dalam konteks ini, atribut yang akan dijadikan id adalah nama barang. RapidMiner menyediakan operator Set Role yang dapat digunakan untuk mengubah peran atribut tersebut menjadi id. Dalam situasi ini, atribut yang akan difungsikan sebagai id adalah nama barang. Pada RapidMiner, untuk mengonversi atribut tersebut menjadi id, dapat dilakukan dengan memanfaatkan operator Set Role.



Gambar 6. Memilih target pada Set role

Setelah menentukan atribut name dan target role, kemudian apply dengan manual atribut nama barang akan diubah menjadi id. Setelah mengubah atribut nama barang menjadi id, langkah berikutnya adalah mengubah atribut Artikel menjadi tipe numerik. Hal ini diperlukan agar ketika melakukan uji performa nilai cluster di *RapidMiner*, tidak terjadi kesalahan, karena untuk melaksanakan uji performa, atribut harus memiliki tipe numerik.



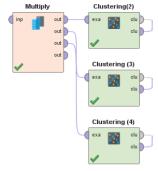
Gambar 7. Operator Nominal to Numerical

Dalam filter tipe atribut, pilih subset, lalu pilih atribut yang akan diubah menjadi tipe numerik, yakni Artikel dan untuk opsi coding type dengan memilih unique integers.



Gambar 8. Menambahkan operator *Nominal to Numerical*

Untuk menguji keakuratan clustering, RapidMiner akan melakukan perbandingan dalam tiga kelompok. Kelompok pertama menggunakan jumlah k=2, kelompok kedua dengan jumlah k=3, dan kelompok ketiga dengan jumlah k=4. Untuk menambahkan lebih dari satu operator clustering dalam RapidMiner, dapat menggunakan operator multiply yang berfungsi untuk menggandakan operator pada platform tersebut.



Gambar 9. Memasukan operator Clustering

Hasil dari ketiga clustering tersebut yang masing-masing memiliki nilai k sebanyak 2, 3, dan 4, hasilnya dapat dilihat seperti yang tergambar pada gambar berikut.

Cluster Model	Cluster Model	Cluster Model
Cluster 0: 257 items Cluster 1: 243 items	Cluster 0: 207 items Cluster 1: 210 items Cluster 2: 83 items	Cluster 0: 160 items Cluster 1: 155 items Cluster 2: 73 items Cluster 3: 112 items
Total number of items: 500	Total number of items: 5	00

Gambar 10. Hasil perbandingan nilai cluster

Tahapan ini bertujuan untuk menilai tingkat akurasi yang diperoleh dari proses tersebut, dengan memeriksa sejauh mana hasilnya sesuai dengan nilai kinerja pada data mining. Semakin kecil nilai kinerja pada data mining, maka hasilnya dianggap semakin baik dan akurat.

Davies Bouldin	Davies Bouldin	Davies Bouldin
Davies Bouldin: 0.546	Davies Bouldin: 0.462	Davies Bouldin: 0.514

Gambar 11. Hasil perbandingan nilai *Davies Bouldin Index*

Hasil dari masing-masing klaster dapat diamati. Pada klaster dengan jumlah k=2, hasilnya terlihat pada gambar, dengan nilai *Davies-Bouldin Index* (DBI) sebesar 0.546, k=3 dengan nilai indeks *Davies-Bouldin Index* (DBI) sebesar 0.462, dan k=4 dengan nilai *Davies Bouldin Index* (DBI) sebesar 0.514.

Davies Bouldin

Davies Bouldin: 0.462

Gambar 12 Nilai Davies Bouldin Index

Dengan merujuk pada hasil *Davies-Bouldin Index* (DBI), diputuskan untuk memilih jumlah klaster k=3 karena menghasilkan nilai *Davies Bouldin Index* yang lebih rendah dibandingkan dengan klaster lainnya. Hasil dari klaster dengan jumlah k=3 dapat dilihat pada tabel berikut:

Table 1. Hasil model cluster

Cluster	Jumlah
Cluster 0	207 Items
Cluster 1	210 Items
Cluster 2	83 Items

Berdasarkan hasil table model cluster diatas nilai *Davies Bouldin Index* Untuk klaster dengan jumlah k=3, hasilnya pada gambar, dengan nilai *Davies Bouldin Index* (DBI) sebesar 0.462. Dalam uji coba menggunakan *RapidMiner*, hasil yang diperoleh juga serupa dengan perhitungan manual. Dalam *RapidMiner*, terdapat penentuan sebanyak k=3 untuk membentuk kelompok klaster produk dari dataset yang terdiri dari 500 data. *cluster* dalam *RapidMiner* dinamai *Cluster*_0, *Cluster*_1, *Cluster* memiliki kesamaan yang sejalan dengan perhitungan manual. Berbeda dengan perhitungan manual, *RapidMiner* tidak menentukan nilai klaster awal, tetapi hasil yang diperoleh tetap tidak jauh berbeda. *Cluster*_0 terdiri

dari 207 produk fashion pria yang laris, sesuai dengan kelompok tinggi pada perhitungan manual. Sementara *Cluster_1* terdiri dari 210 produk *fashion* pria yang biasa, dan *Cluster_2* terdiri dari 83 Produk *Fashion* pria yang kurang laris, sesuai dengan kelompok rendah pada perhitungan manual.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil diskusi, kesimpulan penelitian adalah bahwa metode k means clustering dapat menghasilkan produk yang laris dan kurang laris. Selain itu, pengujian dengan alat RapidMiner juga dapat menghasilkan hasil yang serupa, yaitu masingmasing klaster memiliki anggota kelompok klaster sesuai dengan perhitungan manual seperti Cluster_0 pada RapidMiner memiliki 207 Produk fashion pria yang laris, Cluster_1 pada RapidMiner memiliki 210 Produk fashion pria yang biasa, dan Cluster_3 pada RapidMiner memiliki 83 Poduk Fashion pria yang kurang laris. Dengan bantuan K-Means, metode data mining menawarkan banyak keuntungan dan mempermudah proses pembuatan rencana manajemen stok produk fashion pria.

Beberapa metode *clustering* lainnya perlu dianalisis sebagai referensi untuk pengembangan metode *k-means clustering*, Peneliti diharapkan akan mempertimbangkan penggunaan berbagai analisis statistik saat memproses data.

DAFTAR PUSTAKA

- [1] N. Normah, S. Nurajizah, and ..., "Penerapan data mining metode k-means clustering untuk analisa penjualan pada toko fashion hijab Banten," Jurnal download.garuda.kemdikbud.go.id, 2021. [Online]. Available: http://download.garuda.kemdikbud.go.id/article. php?article=2192225&val=10507&title=Penera pan Data Mining Metode K-Means Clustering Untuk Analisa Penjualan Pada Toko Fashion Hijab Banten
- [2] S. Handoko, F. Fauziah, and ..., "Implementasi Data Mining Untuk Menentukan Tingkat Penjualan Paket Data Telkomsel Menggunakan Metode K-Means Clustering," ... Ilm. Teknol. dan ..., 2020, [Online]. Available: https://ejournal.gunadarma.ac.id/index.php/tekn o/article/view/2677
- [3] S. Butsianto and N. T. Mayangwulan, "Penerapan Data Mining Untuk Prediksi Penjualan Mobil Menggunakan Metode K-Means Clustering," *J. Nas. Komputasi dan ...*. ojs.serambimekkah.ac.id, 2020. [Online]. Available:
 - http://ojs.serambimekkah.ac.id/jnkti/article/dow nloadSuppFile/2428/169
- [4] M. Marsono, "Analisis Data Mining Pada Strategi Penjualan Produk PT Aquasolve Sanaria Dengan Menggunakan Metode K-Means Clustering," J. Teknol. Sist. Inf. dan Sist.

- ..., 2019, [Online]. Available: http://ojs.trigunadharma.ac.id/index.php/jsk/article/view/87
- [5] S. Aulia, "Klasterisasi Pola Penjualan Pestisida Menggunakan Metode K-Means Clustering (Studi Kasus Di Toko Juanda Tani Kecamatan Hutabayu Raja)," *Djtechno J. Teknol. Inf.*, 2020, [Online]. Available: http://jurnal.dharmawangsa.ac.id/index.php/djte chno/article/view/964
- [6] Y. Anggraini, D. Pasha, and D. Damayanti, "Sistem Informasi Penjualan Sepeda Berbasis Web Menggunakan Framework Codeigniter," *J. Teknol. Dan ...*, 2020, [Online]. Available: http://jim.teknokrat.ac.id/index.php/sisteminformasi/article/view/236
- [7] G. Triyandana, L. A. Putri, and ..., "Penerapan Data Mining Pengelompokan Menu Makanan dan Minuman Berdasarkan Tingkat Penjualan Menggunakan Metode K-Means," *J. Appl. ...*, 2022, [Online]. Available: https://jurnal.polibatam.ac.id/index.php/JAIC/art icle/view/3824

- [8] S. Ramadhani, D. Azzahra, and Z. Tomi, "Comparison of K-Means and K-Medoids Algorithms in Text Mining based on Davies Bouldin Index Testing for Classification of Student's Thesis," *Digit. Zo. J. ...*, 2022, [Online]. Available: http://journal.unilak.ac.id/index.php/dz/article/view/9292
- [9] B. M. Metisen and H. L. Sari, "Analisis clustering menggunakan metode K-Means dalam pengelompokkan penjualan produk pada Swalayan Fadhila," *J. media infotama*, 2015, [Online]. Available: https://jurnal.unived.ac.id/index.php/jmi/article/view/258
- [10] I. S. Hidayat, S. Defit, and G. W. Nurcahyo, "Simulasi dalam Optimalisasi Pengadaan Barang menggunakan Metode K-Mean Clustering," *J. Sistim Inf. Dan* ..., 2021, [Online]. Available: http://www.jsisfotek.org/index.php/JSisfotek/art icle/view/79