ANALISIS DATA TRANSAKSI PENJUALAN BARANG MENGGUNAKAN TEKNIK K-MEANS CLUSTERING

Moh Fajar Maulana Adji [1], Gifthera Dwilestari [2]
[1] Teknik Informatika, STMIK IKMI Cirebon
[2] Sistem Informasi, STMIK IKMI Cirebon
Jl. Perjuangan No.10B Majasem-Kota Cirebon, Indonesia
ggdwilestari@gmail.com

ABSTRAK

PT. Swiss Padma Jaya Cirebon menghadapi tantangan dalam memahami pola pembelian pelanggan yang semakin kompleks, terutama dengan pertumbuhan volume data transaksi yang besar dan beragam. Hal ini menyulitkan perusahaan untuk menyusun strategi pemasaran yang efektif serta mengoptimalkan manajemen stok dan alokasi sumber daya. Dalam konteks ini, analisis data menjadi sangat penting untuk memberikan wawasan yang mendalam mengenai perilaku pelanggan dan tren penjualan produk. Penelitian ini bertujuan untuk meningkatkan model clustering pada data transaksi penjualan di PT. Swiss Padma Jaya Cirebon menggunakan algoritma K-Means. Dataset yang digunakan terdiri dari berbagai atribut penting, termasuk Cust. Code, Customer Name, Shipment, Salesman, Product Code, Product Name, Sales Order No, Sales Order Date, Qty SO, Brutto SO, DPP SO, Invoice No, BRAND, dan CATEGORY2. Melalui clustering ini, diharapkan dapat diidentifikasi pola pembelian pelanggan serta produk yang sering terjual, sehingga perusahaan dapat menyusun strategi pemasaran yang lebih efektif dan efisien. Metode yang digunakan dalam penelitian ini meliputi beberapa tahapan, yaitu pengumpulan data transaksi penjualan, pembersihan data (data cleaning) untuk memastikan kualitas data, dan penerapan algoritma K-Means untuk mengelompokkan data berdasarkan pola pembelian dan kategori produk. Evaluasi model dilakukan dengan menggunakan indeks evaluasi seperti Davies-Bouldin Index untuk memastikan kualitas cluster yang terbentuk. Proses pengujian melibatkan berbagai inisialisasi dan variasi jumlah cluster untuk mendapatkan hasil optimal. Hasil penelitian menunjukkan algoritma K-Means berhasil membentuk sepuluh klaster dengan distribusi item yang beragam. Klaster terbesar mencakup 4.884 item, sementara beberapa klaster lainnya berisi lebih sedikit data. Evaluasi model menghasilkan nilai DBI sebesar 0,011, menunjukkan klaster yang sangat baik, dengan pemisahan antar-klaster yang optimal dan kesamaan dalam klaster. Penambahan atribut seperti frekuensi transaksi dan lokasi penjualan diidentifikasi dapat meningkatkan akurasi model. Hasil clustering ini mengungkap pola transaksi, seperti produk musiman dan pelanggan yang rutin membeli merek tertentu, yang dapat mendukung strategi pemasaran yang lebih terfokus dan efisien. Penelitian ini membuktikan bahwa algoritma K-Means adalah alat yang efektif untuk analisis data transaksi penjualan di perusahaan.

Kata kunci : K-Means, Clustering, Transaksi Penjualan, Evaluasi Model, Strategi Pemasaran

1. PENDAHULUAN

Perkembangan pesat di bidang informatika telah membawa dampak signifikan dalam berbagai sektor, termasuk teknologi, bisnis, dan pendidikan. Inovasi dalam teknologi analitik, khususnya dalam pengelolaan dan analisis data besar (big data), memungkinkan perusahaan untuk mengoptimalkan strategi bisnis mereka melalui pemahaman yang lebih mendalam terhadap perilaku konsumen dan tren pasar [1] Salah satu metode analisis yang efektif adalah clustering, yang digunakan untuk mengelompokkan data berdasarkan karakteristik tertentu, sehingga dapat memberikan wawasan yang berguna dalam pengambilan keputusan. PT. Swiss Padma Jaya Cirebon, sebagai perusahaan di bidang industri batik dan tekstil, memiliki data transaksi penjualan yang besar dan kompleks. Untuk meningkatkan efisiensi dan efektivitas strategi bisnis perusahaan, penerapan algoritma clustering seperti K-Means dapat menjadi solusi optimal untuk memahami pola penjualan dan preferensi pelanggan dengan lebih baik [2]

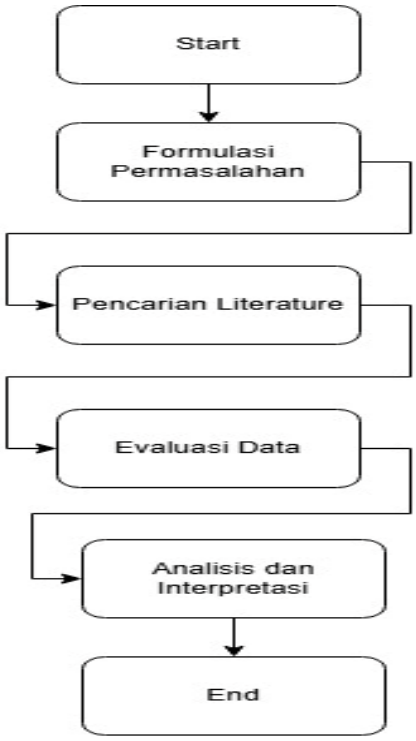
Permasalahan utama dalam penelitian ini adalah bagaimana mengoptimalkan pengelompokan data transaksi penjualan yang besar dan kompleks di PT. Swiss Padma Jaya Cirebon. Pengelolaan data yang tidak efisien dapat menyebabkan kurang akuratnya strategi pemasaran dan alokasi sumber daya perusahaan [3]. Dalam tren terbaru, algoritma K-Means telah banyak digunakan untuk analisis clustering, namun masih terdapat kesenjangan dalam literatur terkait pengoptimalan jumlah cluster dan pemanfaatan fitur tambahan seperti kategori produk dan lokasi penjualan untuk menghasilkan model clustering yang lebih presisi [4]

Penelitian sebelumnya menunjukkan bahwa algoritma *K-Means* merupakan salah satu algoritma yang paling efektif dalam pengelompokan data transaksi penjualan. [5] menunjukkan bahwa K-Means mampu mengidentifikasi pola penjualan dalam data ritel yang kompleks. [6] menyelidiki dampak dari penggunaan atribut tambahan seperti lokasi dan waktu transaksi dalam meningkatkan akurasi model clustering, namun pendekatan ini masih terbatas pada skala data yang lebih kecil.[5] membahas pentingnya inisialisasi yang tepat dalam algoritma K-Means untuk mengurangi risiko konvergensi yang lambat, namun

belum mengeksplorasi secara mendalam bagaimana fitur produk dapat dimanfaatkan untuk segmentasi yang lebih presisi. Oleh karena itu, masih terdapat peluang untuk penelitian lebih lanjut yang mengintegrasikan variabel-variabel ini dalam dataset besar seperti yang dimiliki PT. Swiss Padma Jaya Cirebon.

Penelitian ini bertujuan untuk mengembangkan model clustering yang lebih presisi dalam mengelompokkan data transaksi penjualan di PT. Swiss Padma Jaya Cirebon menggunakan algoritma *K-Means*. Penelitian ini diharapkan dapat mengisi kesenjangan dalam literatur dengan mengeksplorasi penggunaan atribut tambahan seperti kategori produk dan lokasi penjualan serta mengevaluasi model menggunakan metode seperti *Davies-Bouldin Index*. Kontribusi penelitian ini adalah memberikan pemahaman yang lebih mendalam tentang pengelompokan data penjualan, yang nantinya dapat dimanfaatkan untuk strategi pemasaran yang lebih tepat sasaran, serta pengelolaan stok yang lebih efisien di perusahaan.

Pendekatan yang digunakan dalam penelitian ini meliputi tahapan praproses data, penerapan algoritma K-Means, dan evaluasi model menggunakan indeks seperti *Silhouette Score* dan *Davies-Bouldin Index*. Algoritma *K-Means* dipilih karena kemampuannya dalam membagi data berdasarkan kesamaan karakteristik, sementara penggunaan indeks evaluasi bertujuan untuk mengukur kualitas dan efektivitas clustering [1]. Dalam pengujian, berbagai variasi jumlah cluster dan inisialisasi centroid akan diimplementasikan untuk menemukan konfigurasi terbaik yang memberikan akurasi tertinggi [5].



Gambar 1. Metode Literatur riview

Jika penelitian ini berhasil, model clustering yang dikembangkan dapat meningkatkan pemahaman terhadap pola penjualan dan preferensi pelanggan di PT. Swiss Padma Jaya Cirebon. Temuan ini diharapkan dapat memberikan kontribusi signifikan pada bidang informatika, khususnya dalam penerapan algoritma clustering pada data transaksi penjualan skala besar. Praktisi bisnis dapat menggunakan hasil penelitian ini untuk memperbaiki strategi pemasaran, sementara peneliti dapat mengembangkan lebih lanjut metode optimalisasi clustering dengan algoritma lain yang lebih kompleks [7].

2. TINJAUAN PUSTAKA

Metode literature review yang digunakan dalam penelitian ini adalah dengan cara mereview pada artikel-artikel yang terkait dengan Algoritma *K-Means* Untuk Meningkatkan Model *Clustering* Data Transaksi Penjualan Barang Pt. Swiss Padma Jaya Cirebon. Langkah-langkah dari literature review meliputi 4 tahapan , yaitu (1) formulasi permasalahan, (2) pencarian literature, (3) evaluasi data, (4) analisis dan interpretasi. Formulasi permasalahan dilakukan memilih topik. Pada penelitian ini topik yang dipilih adalah mengenai Algoritma *K-Means* Untuk Meningkatkan Model *Clustering* Data Transaksi Penjualan Barang Pt. Swiss Padma Jaya Cirebon. Langkah selanjutnya adalah pencarian literatur yang relevan dengan topik penelitian. Langkah ini dapat memberikan gambaran mengenai Cluster Data Transaksi Penjualan Barang Pt. Swiss Padma Jaya Cirebon, Proses pencarian dan pengumpulan artikel dengan menggunakan tools publish or perish, yang diambil dari database akademik yaitu *Google Scholar, Shinta dan Scopus* dengan kata kunci " K-Means", "Clustering", dan "Transakdi Penjualan, Evaluasi model, strategi pemasaran Dari hasil pencarian ditemukan sebanyak 200, artikel jurnal. Langkah ketiga adalah evaluasi data yaitu dengan melakukan penyaringan dan pemilihan serta pemilahan artikel jurnal yang benar benar relevan dengna topik penelitian serta mempertimbangkan keterbaruannya. Keterbaruan dibatasi dengan memilih artikel 5 tahun terakhir yaitu mulai dari tahun 2023, 2022, 2021, 2020 dan 2019. Sedangkan kesesuaian dilihat dari kesesuain judul paper, sinta index, atau hal lain yang dianggap memiliki reputasi. Berdasarkan hasil evaluasi data, kemudian dipilih 15 artikel jurnal untuk di review. Setelah keempat tahapan tersebut dilakukan, proses selanjutnya adalah pelaksanaan literature review. Adapun teknik yang dilakukan dalam literature review yaitu mencai kesamaan (compare), mencari ketidaksamaan (contrast), memberikan pandangan (criticize), membandingkan (synthesize), dan meringkas (summary)

2.1. Penjualan

Penjualan merupakan bentuk interaksi antara individu yang dilakukan secara tatap muka dengan tujuan untuk menciptakan, memperbaiki, menguasai,

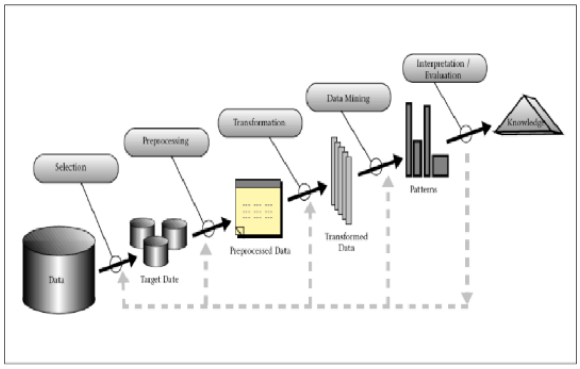
atau mempertahankan hubungan pertukaran yang saling menguntungkan. Selain itu, penjualan juga dapat diartikan sebagai suatu usaha yang dilakukan manusia untuk menyediakan barang atau jasa kepada pihak yang membutuhkan dengan imbalan berupa uang. Transaksi ini dilakukan berdasarkan harga yang telah disepakati oleh kedua belah pihak melalui kerjasama yang saling menguntungkan [8].

2.2. Data Mining

Data mining adalah metode analisis data yang menggabungkan berbagai teknik untuk mengidentifikasi pola atau informasi tersembunyi dari data yang telah terkumpul. Proses ini memungkinkan pengguna untuk mengakses, memahami, dan mengolah sejumlah besar data dengan cepat dan efisien. Dengan pendekatan yang menyatukan teknologi pembelajaran mesin, pengenalan pola, statistik, pengelolaan basis data, dan visualisasi, data mining dapat menangani permasalahan kompleks dalam pengambilan informasi dari basis data yang besar. Hal ini mendukung pengambilan keputusan berbasis data di berbagai bidang, termasuk bisnis, kesehatan, dan sains [9].

2.3. Proses Data Mining

Dalam penelitian ini dilakukan beberapa tahapan yang dilakukan secara sistematis, bertujuan untuk memermudah peneliti dalam mengangkat prosedur dan rangkaian dalam sebuah penelitian dan lebih terarah. Knowladge Discovery in Database (KDD) adalah metode yang digunakan untuk dapat memperoleh pengetahuan yang berasal dari database yang ada. Hasil pengetahun yang diperoleh dapat dimanfaat kan untuk basis pengetahuan (knowledge base) yang digunakan dalam keperluan mengambil keputusan [10].



Gambar 2. Proses KDD

Setiap proses dalam suatu kegiatan melibatkan beberapa tahap yang harus dilalui untuk mencapai hasil yang diinginkan. Hal ini juga berlaku dalam proses Data Mining, yang sering disebut sebagai Knowledge Discovery Data (KDD)[4]. terdapat beberapa proses:

a. Pembersihan Data (Data Cleaning) melibatkan penghapusan data duplikat, memeriksa data yang tidak konsisten, dan memperbaiki kesalahan data seperti kesalahan ketik.
b. Integrasi Data (Data Intergration) proses penggabungan data dari database yang berbeda ke dalam database baru yang memerlukan KDD
c. Seleksi data (Data Selection) Memilih data yang relevan dan dapat dianalisis dari data operasional. Data hasil pemilu disimpan dalam database tersendiri.
d. Transformasi Data (Data Transformation) Proses mengubah data ke dalam format tertentu sehingga cocok untuk proses data mining. Misalnya, beberapa metode standar seperti analisis asosiasi dan pengelompokan hanya dapat menerima data kategorikal sebagai masukan.
e. Penambangan Data (Data Mining) Proses pencarian pola atau informasi menarik dengan menggunakan teknik, metode, atau algoritma tertentu.
f. Evaluasi Pola (Pettern Evaluation) Mengidentifikasi pola yang benar-benar menarik dari hasil data mining. Pada tahap ini, hasil teknik data mining dievaluasi dalam bentuk pola tipikal dan model prediksi untuk menilai apakah hipotesis yang ada terpenuhi atau tidak.
g. Presentasi Pengetahuan (Knowladge Presentation) Menampilkan pola informasi hasil proses data mining ini membantu mengkomunikasikan hasil data mining dengan cara yang mudah dipahami.

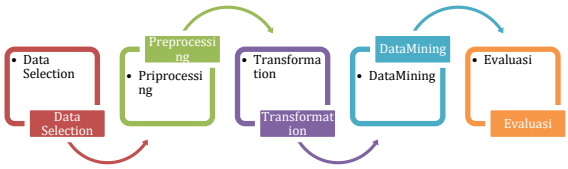
2.4. Analisis Cluster

Menurut Tan, analisis kelompok (cluster analysis) adalah metode pengelompokan data (objek) yang didasarkan hanya pada informasi yang ditemukan dalam data yang menggambarkan objek tersebut dan hubungan di antaranya. Tujuannya adalah agar objek-objek yang bergabung dalam sebuah cluster merupakan objek-objek yang mirip (atau berhubungan) satu sama lain dan berbeda (atau tidak berhubungan) dengan objek dalam cluster yang lain. Lebih besar kemiripannya (homogenitas) dalam cluster dan lebih besar perbedaanya diantara cluster yang lain. Analisis cluster membuat pengelompokan objek berdasarkan jarak antara pasangan [11] objek. Jarak merupakan ukuran yang digunakan untuk mengukur kemiripan dari suatu objek. Pada proses pemilihan ini, dapat dipilih salah satu dari beberapa jarak yang biasa digunakan salah satunya adalah jarak Euclidean. Jarak Euclidean adalah akar dari jumlah kuadrat perbedaan/deviasi di dalam nilai untuk setiap variabel (Supranto, 2004). Jarak euclidean antara cluster objek ke-i dan cluster objek ke-g dari p variabel didefinisikan sebagai berikut:

$$d(i,g)\sqrt{\sum_{j=1}^{p}(x_{ij}-x_{gj})}\,2 \qquad (1)$$

3. METODE PENELITIAN

Dalam penelitian ini dilakukan beberapa tahapan yang dilakukan secara sistematis, bertujuan untuk memermudah peneliti dalam mengangkat prosedur dan rangkaian dalam sebuah penelitian dan lebih terarah. *Knowladge Discovery in Database* (KDD) adalah metode yang digunakan untuk dapat memperoleh pengetahuan yang berasal dari database yang ada. Hasil pengetahun yang diperoleh dapat dimanfaat kan untuk basis pengetahuan (knowledge base) yang digunakan dalam keperluan mengambil keputusan. Dalam menganalisis data pada penelitian ini menggunakan Teknik analisis data mining menggunakan proses tahapan knowledge discovery in databases (KDD) yang terdiri dari Data, Data Cleaning, Data transformation, Data mining, Pattern evolution, knowledge [12].



Gambar 3. Alur Penelitian

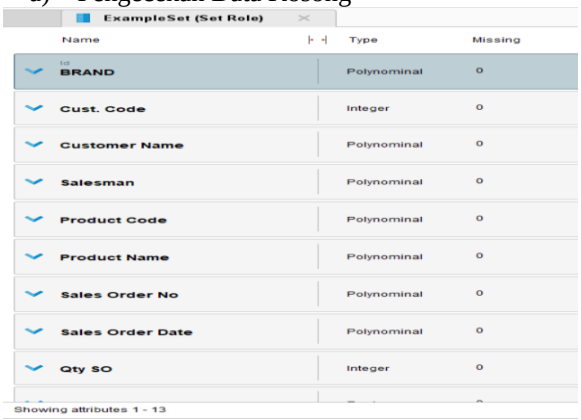
Tabel 1. Data Seletion.

No	Product Name	Qty SO	Brutto SO	DPP SO	……	CATEGORY2
1	NC TLT CORE NON EMB 1 ROLL 238S X 120 GT	12	32618,16	32618,16	……	TOILET
2	PASEO SMART FACIAL TRAVEL PACK 50'S GT	12	20160	20160	……	FACIAL
3	PASEO SMART FACIAL TRAVEL PACK 50'S GT	40	67200	67200	……	FACIAL
4	PS SMRT FCL SOFT PACK 48X540PLY GT	1	8818,19	8818,19	……	FACIAL
5	PS SMRT FCL SOFT PACK 48X540PLY GT	1	8818,19	8818,19	……	FACIAL
6	PS SMRT FCL SOFT PACK 48X540PLY GT	1	8818,19	8818,19	……	FACIAL
7	NICE FACIAL SOFT PACK 60 X 360 HELAI- GT	1	5150	5150	……	FACIAL
8	PS SMRT HKY PERF 6 PACK 12S X 40	1	4160	4160	……	FACIAL
9	PASEO SMART FACIAL TRAVEL PACK 50'S GT	1	1680	1680	……	FACIAL
10	PASEO ANTI BACTERIAL WIPES 18X25S BOGOF	3	21000	21000	……	WET WIPES
11	PASEO SMART FACIAL TRAVEL PACK 50'S GT	30	50400	50400	……	FACIAL
12	PASEO BABY WIPES GAZETTE 50'S X 18	1	9009	9009	……	WET WIPES
…..	……	……..	……..	……..		……
6429	PS AROMA FCL TRAVEL PACK 3P 60S X 32 MT	24	608088	608088	……	FACIAL

3.1. Preprocessing

Proses data cleaning dilakukan dengan memeriksa dataset yang akan digunakan, Proses ini dilakukan dengan pengecekan data kosong atau data ganda[10].

a) Pengecekan Data Kosong



Gambar 2, Pengecekan Data Kosong

Pengecekan data kosong dengan melihat kolom yang missing value pada dataset. Gambar 2 hasil pengecekan data missing value. Hasil pengecekan tidak ada data kosong.

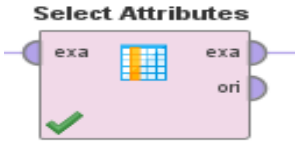
b) Pengecekan Data Yang Menyimpang Jauh.

Gambar 3, Pengecekan Data yang menyimpang jauh

Pengecekan data dilihat dari nilai minimal dan maksimal setiap attribute. Gambar 3 adalah nilai minimal dan masimal dari dataset. Dari gambar tersebut nilai minimal dan maksimal sesuai dengan yang diperlukan.

3.2. Transformation

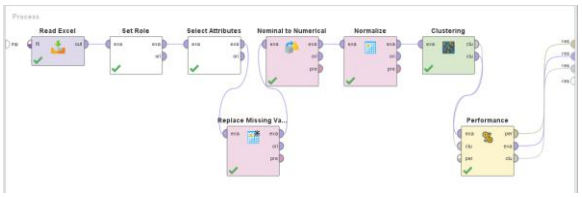
Proses ini dilakukan dengan seleksi pada attribute yang digunakan, tidak semua attribute yang ada di dataset digunakan. Proses pemilihan seleksi menggunakan operator select attribute.



Gambar 4, Atribut Select Atribute

3.3. Model Algoritma K-Means

Data mining merupakan salah satu proses untuk mempekerjakan satu atau lebih teknik pembelajaran komputer (machine learning) untuk menganalisis dan mengekstraksi pengetahuan (knowledge) secara otomatis.[13] Proses penelitian ini menggunakan algoritma k-means untuk mengelompokan hasil cek darah pada lansia yang mempunyai penyakit diabetes. Dalam penelitian ini dibantu rapidminer untuk algoritma k-means clustering. Tahapan dalam penelitian ini dimulai dengan tools rapidminer seperti gambar 5.



Gambar 5, Model Clustering

3.4. Evaluasi

Evaluasi model dilakukan pengelompokan data untuk menemukan cluster yang baik dalam proses pembuatan model. Proses ini dilakukan untuk mendapatkan cluster terbaik dengan melihat hasil evaluasi *Davies Bouldin Index* (DBI). Hasil nilai DBI adalah hasil nilai evaluasi paling kecil. Hasil dari model *k-means* yang digunakan, masing-masing algoritma *k-means* dari nilai k=2 sampai k= 10 bisa dilihat dan dievaluasi nilai performance.

4. HASIL DAN PEMBAHASAN

Pada bagian ini, hasil penerapan algoritma K-Means untuk mengelompokkan data transaksi penjualan di PT. Swiss Padma Jaya Cirebon dibahas secara mendalam. Analisis ini mengungkap wawasan penting terkait pola pembelian pelanggan dan karakteristik produk yang sering terjual. Evaluasi kualitas clustering, menggunakan Davies-Bouldin Index (DBI), menunjukkan efektivitas algoritma dalam memisahkan dan mengelompokkan data transaksi dengan baik. Selain itu, pembahasan mencakup faktor-faktor yang mempengaruhi akurasi model, seperti penambahan atribut relevan dan pemilihan jumlah klaster yang optimal. Hasil penelitian menunjukkan bahwa model clustering yang dikembangkan dapat mendukung perusahaan dalam merumuskan strategi pemasaran yang lebih terarah, meningkatkan efisiensi pengelolaan stok, serta merencanakan alokasi sumber daya berdasarkan pola permintaan yang teridentifikasi melalui klaster-klaster tersebut. Dengan demikian, hasil dan pembahasan ini memberikan kontribusi signifikan dalam memanfaatkan analisis data untuk pengambilan keputusan strategis di PT. Swiss Padma Jaya Cirebon.

4.1. Data Selection

Tahap *data selection* dalam metode KDD melibatkan pemilihan data yang relevan dari kumpulan data transaksi penjualan yang tersedia di PT. Swiss Padma Jaya Cirebon. Data yang dipilih harus sesuai dengan tujuan penelitian dan memiliki atribut-atribut yang penting untuk proses analisis klastering menggunakan algoritma K-Means. Data penjualan yang mencakup atribut seperti Cust.Code, Customer Name, Salesman, ProductCode, Product Name, Sales, sales Order, Sales order date, QTY SO, Brutto SO, DPP SO, Invoice No, Brand, Category2 dengan jumlah data 6429 record dapat dipilih untuk menyediakan informasi yang memadai bagi proses pengelompokan. Pada tahap ini, data yang kurang relevan atau memiliki kualitas rendah dapat diabaikan untuk memastikan hasil yang lebih akurat dan efektif. Dataset data transaksi penjualan barang hasil dari data selection seperti pada tabel 2

Tabel 2, Data Seletion.

No	Product Name	Qty SO	Brutto SO	DPP SO		CATEGORY2
1	NC TLT CORE NON EMB 1 ROLL 238S X 120 GT	12	32618,16	32618,16	……	TOILET
2	PASEO SMART FACIAL TRAVEL PACK 50'S GT	12	20160	20160	……	FACIAL
3	PASEO SMART FACIAL TRAVEL PACK 50'S GT	40	67200	67200	……	FACIAL
4	PS SMRT FCL SOFT PACK 48X540PLY GT	1	8818,19	8818,19	……	FACIAL
5	PS SMRT FCL SOFT PACK 48X540PLY GT	1	8818,19	8818,19	……	FACIAL
6	PS SMRT FCL SOFT PACK 48X540PLY GT	1	8818,19	8818,19	……	FACIAL
7	NICE FACIAL SOFT PACK 60 X 360 HELAI- GT	1	5150	5150	……	FACIAL
8	PS SMRT HKY PERF 6 PACK 12S X 40	1	4160	4160	……	FACIAL
9	PASEO SMART FACIAL TRAVEL PACK 50'S GT	1	1680	1680	……	FACIAL
10	PASEO ANTI BACTERIAL WIPES 18X25S BOGOF	3	21000	21000	……	WET WIPES
11	PASEO SMART FACIAL TRAVEL PACK 50'S GT	30	50400	50400	……	FACIAL
12	PASEO BABY WIPES GAZETTE 50'S X 18	1	9009	9009	……	WET WIPES
…..	……	……..	……..	…….		……
6429	PS AROMA FCL TRAVEL PACK 3P 60S X 32 MT	24	608088	608088	……	FACIAL

Hasil dari uji k=2 sampai k=10 menghasilkan nilai DBI yang optimal menjadi 9 cluster dengan hasil nilai DBI 0,011 paling rendah dan anggota cluster pada gambar 6

Cluster Model

```
Cluster 0: 1194 items
Cluster 1: 42 items
Cluster 2: 28 items
Cluster 3: 4884 items
Cluster 4: 59 items
Cluster 5: 208 items
Cluster 6: 1 items
Cluster 7: 12 items
Cluster 8: 1 items
Total number of items: 6429
```

Gambar 6. Cluster Model

hasil k=2 dengan nilai DBI 0.028, Cluster 0: 1194 items dan Cluster 1: 5235 items. diatas hasil k=3 dengan nilai DBI 0,019 dengan anggota Cluster 0: 1194 items Cluster 1: 6 items dan Cluster 2: 5229 items. hasil k=4 dengan nilai DBI 0.015, dengan anggota Cluster 0: 7 items Cluster 1: 5215 items, Cluster 2: 1194 items dan Cluster 3: 13 items. hasil k=5 dengan nilai DBI 0.019, dengan anggota Cluster 0: 4993 items Cluster 1: 54 items, Cluster 2: 14 items, Cluster 3: 1192 items, Cluster 4: 176 items. hasil k=6 dengan nilai DBI 0,017, dengan anggota Cluster 0: 774 items, Cluster 1: 420 items, Cluster 2: 5086 items, Cluster 3: 1 items, Cluster 4: 59 items dan Cluster 5: 89 items. hasil k=7 dengan nilai DBI 0,019, dengan anggota cluster Cluster 0: 64 items Cluster 1: 4331 items, Cluster 2: 89 items, Cluster 3: 36 items, Cluster 4: 103 items, Cluster 5: 1789 items dan Cluster 6: 17 items. hasil k=8 dengan nilai DBI 0,012, dengan anggota cluster Cluster 0: 1194 items, Cluster 1: 9 items, Cluster 2: 49 items, Cluster 3: 10 items, Cluster 4: 87 items, Cluster 5: 5021 items, Cluster 6: 10 items, Cluster 7: 49 items. hasil k=9 dengan nilai DBI 0,011, dengan anggota cluster Cluster 0: 1194 items, Cluster 1: 42 items, Cluster 2: 28 items, Cluster 3: 4884 items, Cluster 4: 59 items, Cluster 5: 208 items, Cluster 6: 1 items, Cluster 7: 12 items, Cluster 8: 1 items. hasil k=10 dengan nilai DBI 0,014, dengan anggota cluster Cluster 0: 317 items, Cluster 1: 176 items, Cluster 2: 11 items, Cluster 3: 4455 items, Cluster 4: 89 items, Cluster 5: 42 items, Cluster 6: 53 items, Cluster 7: 88 items, Cluster 8: 1181 items, Cluster 9: 17 items. Dari hasil uji model algoritma K-Means mulai dari k = 2 sampai dengan k=20 pada gambar bahwa hasil yang optimal dan yang terbaik dilihat pada hasil nilai DBI 0,011, berada pada K=9 dengan nilai *Avg. within centroid distance*: 0.870. Hasil dari uji k=2 sampai k=10 menghasilkan nilai DBI yang optimal menjadi 9 cluster dengan hasil nilai DBI 0,011 paling rendah dan anggota cluster.

Gambar 7. visualisasi scatterplot

Gambar tersebut menunjukkan diagram sebar (scatter plot) yang memvisualisasikan data transaksi penjualan berdasarkan merek (*brand*) dan kuantitas penjualan (Qty SO) di PT. Swiss Padma Jaya Cirebon. Sumbu X menunjukkan kategori merek barang, yaitu "NICE," "PASEO," "JOLLY," "TOPLY," dan "PL," sedangkan sumbu Y menunjukkan kuantitas penjualan barang tersebut. Setiap titik dalam plot mewakili transaksi individu, dengan warna berbeda untuk setiap merek:

a. "NICE" diwakili oleh warna biru muda,
b. "PASEO" oleh warna hijau,
c. "JOLLY" oleh warna oranye,
d. "TOPLY" oleh warna hitam,
e. dan "PL" oleh warna ungu.

Pada contoh yang ditunjukkan, ada tooltip yang menampilkan informasi dari satu titik data "NICE," dengan koordinat X dan Y sebesar (NICE, 16.72), yang berarti transaksi tersebut terjadi dengan merek "NICE" dan kuantitas penjualan sekitar 16.72 unit.
Secara keseluruhan, grafik ini memberikan gambaran distribusi kuantitas penjualan per merek, yang dapat digunakan untuk mengidentifikasi tren penjualan berdasarkan merek tertentu di perusahaan.

Pada penelitian ini, algoritma K-Means digunakan untuk mengelompokkan data transaksi penjualan di PT. Swiss Padma Jaya Cirebon, dengan tujuan membagi data ke dalam klaster-klaster yang merepresentasikan karakteristik transaksi yang serupa. Melalui penggunaan Rapidminer dan penerapan metode KDD, dilakukan pemilihan data, praproses data, serta pemilihan jumlah klaster yang optimal menggunakan Davies-Bouldin Index (DBI). Hasilnya, model clustering menghasilkan sepuluh klaster dengan distribusi item yang bervariasi, dengan Klaster 3 menjadi klaster terbesar berisi 4.884 item, sementara beberapa klaster lainnya memiliki sedikit item, menunjukkan perbedaan signifikan dalam pola transaksi. Untuk meningkatkan akurasi model, identifikasi atribut tambahan seperti frekuensi transaksi, waktu pembelian, metode pembayaran, dan lokasi penjualan dilakukan, yang memperkaya informasi dalam proses clustering. Evaluasi model menggunakan DBI menunjukkan nilai 0,011, yang menandakan klastering yang sangat baik, dengan pemisahan antar-klaster yang optimal dan kesamaan tinggi dalam setiap klaster. Hasil ini menunjukkan bahwa K-Means efektif dalam mengidentifikasi pola

transaksi dan membantu PT. Swiss Padma Jaya Cirebon dalam merumuskan strategi pemasaran yang lebih terfokus serta mengelola stok berdasarkan pola permintaan dari setiap klaster pelanggan.

5. KESIMPULAN DAN SARAN

Penelitian ini berhasil mengembangkan model clustering menggunakan algoritma K-Means untuk menganalisis data transaksi penjualan di PT. Swiss Padma Jaya Cirebon. Dengan menggunakan metode KDD dan evaluasi menggunakan Davies-Bouldin Index (DBI),

Hasil dari penelitian ini dengan metode algoritma K-Means clustering dan mengunakan rapidminer dengan hasil nilai dari Davies Bouldin Index(DBI) diperoleh nilai optimal dan terbaik dari k=2 sapai dengan k=10 menghasilkan nilai optimal pada k=9 dengan nilai DBI 0,011 dengan anggota cluster, Cluster 0: 1194 items, Cluster 1: 42 items, Cluster 2: 28 items, Cluster 3: 4884 items, Cluster 4: 59 items, Cluster 5: 208 items, Cluster 6: 1 items, Cluster 7: 12 items, Cluster 8: 1 items. model clustering yang dihasilkan menunjukkan kualitas yang sangat baik, dengan nilai DBI 0,011 yang menunjukkan pemisahan antar-klaster yang optimal dan kesamaan tinggi dalam klaster. Klaster-klaster yang terbentuk merepresentasikan pola transaksi yang berbeda, yang dapat digunakan untuk merumuskan strategi pemasaran yang lebih efisien, meningkatkan pengelolaan stok, dan perencanaan alokasi sumber daya berdasarkan pola permintaan dari setiap klaster pelanggan.

PT. Swiss Padma Jaya Cirebon disarankan untuk mempertimbangkan penambahan atribut tambahan, seperti waktu pembelian dan lokasi penjualan, dalam proses klastering selanjutnya guna memperkaya pemahaman terhadap pola transaksi serta meningkatkan akurasi model clustering. Selain itu, metode clustering lain, seperti DBSCAN atau hierarchical clustering, dapat dijadikan alternatif untuk membandingkan efektivitas metode dalam konteks data transaksi perusahaan. Penggunaan alat analitik lanjutan serta integrasi hasil klastering ke dalam sistem manajemen perusahaan juga akan mendukung optimalisasi strategi pemasaran dan pengelolaan stok secara dinamis, sesuai dengan kebutuhan pasar. Implementasi model clustering secara berkelanjutan dalam analisis perilaku pelanggan akan membantu perusahaan untuk tetap responsif terhadap perubahan kebutuhan konsumen, sekaligus meningkatkan kualitas layanan pelanggan secara keseluruhan.

DAFTAR PUSTAKA

[1] Jiawei Han 2022, *Data Mining: Concepts and Techniques. Morgan Kaufmann*. [Online]. Available: https://shop.elsevier.com/books/data-mining/han/978-0-12-811760-6

[2] A. Iqbal, A. Shil, M. J. M. Chowdhury, M. A. Moni, and I. H. Sarker, "An Improved K-means Clustering Algorithm Towards an," *Ann. Data Sci.*, vol. 11, no. 5, pp. 1525–1544, 2024, doi: 10.1007/s40745-022-00428-2.

[3] A. Wardhana, "Strategi Pemasaran Perusahaan," no. July, 2024.

[4] I. C. Gormley, T. B. Murphy, and A. E. Raftery, "Model-Based Clustering," pp. 573–595, 2023.

[5] B. I. Nugroho, A. Rafhina, P. S. Ananda, and G. Gunawan, "Customer segmentation in sales transaction data using k-means clustering algorithm," vol. 7, no. 2, pp. 130–136, 2024.

[6] A. Julian and H. S. R, "Optimizing Customer Segmentation through Machine Learning," in *2024 IEEE International Conference on Computing, Power and Communication Technologies (IC2PCT)*, 2024, pp. 413–416. doi: 10.1109/IC2PCT60090.2024.10486699.

[7] H. Xie *et al.*, "Improving K-means Clustering with Enhanced Firefly Algorithms," no. November, 2019, doi: 10.1016/j.asoc.2019.105763.

[8] D. Algoritma, S. Moving, and A. Sma, "Prediksi penjualan barang pada toko baby shop dengan algoritma single moving average (sma)," vol. 07, pp. 1189–1197, 2022.

[9] S. Kasus, C. V Arya, G. Usaha, Y. O. Tamba, and N. Ekawati, "KLASTERISASI PINJAMAN NASABAH KOPERASI MENGGUNAKAN ALGORITMA K-MEANS," vol. 8, no. 6, pp. 11107–11114, 2024.

[10] W. T. Pambudi and A. Witanti, "Penerapan Algoritma K-Means Clustering Untuk Menganalisis Penjualan Pada Toko Ayu Collection Barbasis Web," *J. Inform. Univ. Pamulang*, vol. 6, no. 3, pp. 645–650, 2021, [Online]. Available: https://www.neliti.com/publications/466160/penerapan-algoritma-k-means-clustering-untuk-menganalisis-penjualan-pada-toko-ay

[11] D. R. Ningrat, D. A. I. Maruddani, and T. Wuryandari, "Analisis cluster dengan algoritma k-means dan fuzzy c-means clustering untuk pengelompokan data obligasi korporasi," *J. Gaussian*, vol. 5, no. 4, pp. 641–650, 2016.

[12] T. Hartati, O. Nurdiawan, and E. Wiyandi, "Analisis Dan Penerapan Algoritma K-Means Dalam Strategi Promosi Kampus Akademi Maritim Suaka Bahari," *J. Sains Teknol. Transp. Marit.*, vol. 3, no. 1, pp. 1–7, 2021, doi: 10.51578/j.sitektransmar.v3i1.30.

[13] R. Supardi and I. Kanedi, "Implementasi Metode Algoritma K-Means Clustering pada Toko Eidelweis," *J. Teknol. Inf.*, vol. 4, no. 2, pp. 270–277, 2020, doi: 10.36294/jurti.v4i2.1444.