ANALISIS POLA PENJUALAN OBAT DI APOTEK AN-NAAFI MENGGUNAKAN METODE *K-MEANS CLUSTERING*

Miftahul Fajar, Nining Rahaningsih, Raditya Danar Dana
Teknik Informatika, Komputerisasi Akuntansi,
Manajemen Informatika, STMIK IKMI Cirebon
Jl. Perjuangan No.10B, Karyamulya, Kec. Kesambi, Kota Cirebon, Jawa Barat 45135, Indonesia
mfajarggs@gmail.com

ABSTRAK

Apotek An-Naafi mengalami minimnya informasi untuk pengembangan bisnis dan produk karena belum melakukan pengelompokan data penjualan obat berdasarkan perilaku konsumen. Ini menghambat analisis peluang pengembangan produk, layanan, dan strategi pemasaran yang tepat. Oleh karena itu, perlu melakukan pengelompokan data penjualan obat berdasarkan perilaku konsumen untuk mendukung pengembangan bisnis di Apotek An-Naafi. Penelitian ini menerapkan metode K-Means untuk menganalisis data penjualan obat di Apotek An-Naafi. Data obat dikelompokkan berdasarkan tingkat penjualan (rendah dan tinggi) untuk mengidentifikasi pola dan tren penjualan guna memberikan wawasan penting bagi apotek. Data penjualan melibatkan informasi seperti nama obat, penjualan, pendapatan dan laba. Analisis dimulai dengan *preprocessing* data, termasuk penanganan data yang hilang dan kategorisasi obat berdasarkan tingkat pemakaian. Metode K-Means diterapkan dengan inisialisasi *cluster* yang sesuai, membentuk 2 *cluster* utama mencerminkan tingkat penjualan: rendah dan tinggi. Setiap *cluster* menunjukkan karakteristik penjualan obat yang berbeda, memberikan wawasan tentang preferensi pelanggan dan potensi peningkatan penjualan. Hasil ini mendalam tentang tren penjualan obat, termasuk obat yang perlu perhatian lebih, peningkatan signifikan, atau penurunan dalam kategori tertentu. Diskusi hasil penelitian melibatkan rekomendasi untuk meningkatkan strategi penjualan obat di Apotek An-Naafi. Penerapan K-Means berhasil membentuk *cluster*, memberikan wawasan berharga untuk meningkatkan efisiensi dalam industri farmasi. Penggunaan *Euclidean distance* memengaruhi pembentukan hasil *cluster*, dengan cluster 1 menonjol dalam penjualan dan pendapatan tinggi, sedangkan *cluster* 0 lebih beragam. *Cluster* 0 mencakup 381 hasil penjualan dengan pendapatan sebesar Rp. 668.767, yang tergolong rendah. Sementara itu, *cluster* 1 menunjukkan hasil unggul dengan 521 penjualan dan pendapatan sebesar Rp. 3.353.880, yang dikategorikan sebagai tinggi.

Kata kunci : *Data Mining, Clustering, K-Means, Apotek, Obat*

1. PENDAHULUAN

Apotek Apotek an-naafi menghadapi kendala minimnya informasi untuk pengembangan bisnis dan produk. Pengelompokan data penjualan obat berdasarkan perilaku konsumen belum dilakukan. Metode k-means *clustering* diusulkan untuk mengelompokkan data tersebut, menciptakan *cluster* dengan pola pembelian dan karakteristik konsumen serupa. Hingga kini, kurangnya informasi mengenai segmentasi dan pola pembelian obat menjadi hambatan dalam pengembangan bisnis apotek an-naafi. Data tersebut memiliki potensi untuk menganalisis peluang pengembangan produk dan layanan sesuai dengan kebutuhan tiap segmen konsumen.

Penelitian sebelumnya oleh Gustientiedina [1] dengan judul "Penerapan Algoritma K-Means untuk *Clustering* Data Obat-Obatan pada RSUD Pekanbaru" menunjukkan hasil analisis cluster pada data obat-obatan. Kategori obat dengan tingkat penggunaan rendah memiliki permintaan rata-rata kurang dari 18.000 unit per tahun, sedangkan tingkat penggunaan menengah dan tinggi memiliki permintaan masing-masing antara 18.000 hingga 70.000 unit dan di atas 70.000 unit per tahun. Uji empiris oleh Izzah & Jananto [2] menunjukkan bahwa penerapan algoritma K-Means *Clustering* dalam proyeksi penggunaan obat di Klinik Citra Medika lebih mendekati realitas dibandingkan prediksi manual.

Tujuan utama penelitian ini adalah menerapkan metode K-Means pada data penjualan obat di Apotek An-Naafi untuk mengidentifikasi pola penjualan dan memberikan wawasan berharga. Signifikansinya terletak pada peningkatan efisiensi pengelolaan penjualan obat di industri farmasi dan kontribusi pada penggunaan analisis data dalam konteks praktis. Penelitian ini melibatkan aplikasi metode K-Means pada data penjualan obat yang dikelompokkan berdasarkan tingkat penjualan dan pendapatan (rendah dan tinggi), dengan proses analisis mencakup *preprocessing* data, penentuan jumlah *cluster* yang sesuai, dan evaluasi hasil. Data diperoleh dari ekspor aplikasi apotek.

Hasil penelitian ini Berpotensi Memberikan Wawasan Strategis Untuk Penjualan Obat Di Apotek An-Naafi, Dengan Implikasi Praktis Yang Dapat Diadopsi Oleh Apotek lain Dalam Industri Farmasi. Selain Itu, Penelitian Ini Mencerminkan Potensi Penggunaan Analisis Data Dalam Berbagai Konteks Bisnis, Memberikan Pandangan Baru Bagi Praktisi,

Peneliti, Dan Pemangku Kepentingan Di Bidang Informatika dan Industri Farmasi.

2. TINJAUAN PUSTAKA

2.1. *Data mining*

Data Mining adalah proses penyaringan data secara implisit[3]. Untuk mendapatkan nilai dari data, dilakukan penggalian pola baru dari suatu *dataset* yang tersedia[4]. *Data mining* merupakan proses penggabungan informasi dan analisis yang digunakan untuk mengidentifikasi pola dan hubungan yang tersembunyi dalam kumpulan data yang besar[5].

2.2. *Clustering*

Clustering merupakan suatu proses pengelompokan data menjadi beberapa kelompok berdasarkan kemiripan setelah melalui proses pengujian[6]. *Clustering* adalah pembentukan kelompok atau *cluster* yang sebelumnya tidak diketahui oleh siapa pun, berdasarkan *dataset* yang digunakan[7].

2.3. K-Means

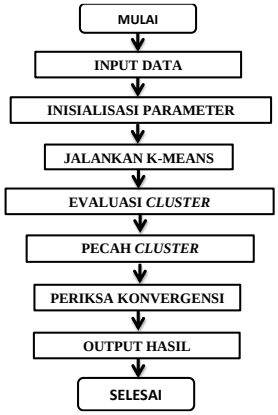
Algoritma K-Means adalah salah satu algoritma dengan teknik *clustering* berdasarkan pembagian jarak dalam *data mining*. Keuntungan menggunakan algoritma ini mudah dipahami, diterapkan, serta memiliki efek pengelompokan yang baik sehingga KMeans banyak digunakan dalam bidang penelitian[8].

2.4. *RapidMiner*

RapidMiner adalah perangkat lunak yang digunakan untuk melakukan proses pengolahan *data mining* dengan langkah awal berupa pemeriksaan dan analisis pola serta atribut yang akan digunakan selanjutnya[9]. *RapidMiner*, sebagai perangkat lunak, bersifat terbuka (*Open Source*) dan dapat digunakan oleh siapa pun, digunakan untuk berkolaborasi dalam pengolahan *data mining*[10].

2.5. Flowchart K-Means *Clustering*

Berikut adalah alur metode X-Means *Clustering* dalam *Flowchart*:



Gambar 1 Langkah- Langkah X-Means *Clustering* Menggunakan *Flowchart*

Gambar 1 menunjukkan langkah-langkah X-Means *Clustering* menggunakan flowchart. Pada flowchart ini, setiap simbol mewakili tahapan tertentu dalam algoritma X-Means *Clustering*, yang dimulai dari inisiasi dan input data hingga proses evaluasi *cluster*, pemecahan *cluster*, dan output hasil *clustering* akhir. Berikut penjelasan dari setiap proses X-Means *Clustering*:

1) **Mulai:** Menjalankan langkah awal algoritma.
2) **Input Data:** Memasukkan dataset yang akan dikelompokan.
3) **Inisialisasi Parameter:**
   a. Menetapkan nilai awal jumlah *Cluster* (k) pada nilai yang kecil.
   b. Menetapkan parameter lain seperti jumlah iterasi maksimal, kriteria konvergensi, dan sebagainya.
4) **Proses K-Means:** Melakukan algoritma *clustering* K-Means dengan nilai k saat ini untuk membagi data menjadi *cluster*.
5) **Evaluasi *Cluster*:** Mengevaluasi kualitas *cluster* menggunakan kriteria seperti *Bayesian Information Criterion* (BIC) atau metrik yang sesuai.
6) **Pemecahan *Cluster*:** Jika hasil *cluster* tidak memuaskan (berdasarkan evaluasi), melakukan pemecahan pada setiap *cluster* menjadi *subcluster*.
7) **Konvergensi:** Memeriksa apakah algoritma telah konvergen atau apakah jumlah iterasi maksimal telah tercapai. Jika telah konvergen, melanjutkan ke Langkah 8. Jika belum konvergen, kembali ke Langkah 4.
8) **Output Hasil:** Menampilkan hasil *clustering* akhir.
9) **Selesai:** Mengakhiri algoritma.

3. METODE PENELITIAN

3.1. Jenis Penelitian

Penelitian ini menggunakan jenis pendekatan kuantitatif untuk mengumpulkan, menganalisis, dan menafsirkan data. Pendekatan kuantitatif memfokuskan pada pengumpulan data yang dapat diukur dalam bentuk angka atau numerik, memungkinkan peneliti untuk melakukan analisis statistik yang sistematis dan objektif. Metode-metode kuantitatif, seperti survei atau eksperimen, digunakan untuk memperoleh informasi yang dapat diolah secara matematis, sehingga memungkinkan pengambilan kesimpulan yang kuat berdasarkan bukti empiris. Pendekatan ini memberikan keunggulan dalam menggambarkan hubungan antarvariabel, mengukur frekuensi, dan mengidentifikasi pola atau tren dalam data.

3.2. Sumber Data

Data sekunder dalam penelitian ini berjumlah 1486 data penjualan bulan Januari sampai Oktober 2023 yang bersumber dari Apotek An-Naafi yang beralamat di Kertasemaya, Indramayu, dan menjadi

landasan utama dalam menganalisis pola penjualan obat serta mengimplementasikan Metode K-Means *Clustering* untuk mengelompokkan data tersebut.

3.3. Metode Penelitian

Metode penelitian dalam penelitian ini menggunakan K-Means *Clustering* sebagai teknik analisis data untuk mengelompokkan data penjualan obat. Selain itu, penelitian ini menerapkan tahapan *Knowledge Discovery in Databases* (KDD) sebagai kerangka kerja yang komprehensif, mencakup pengumpulan data, *preprocessing*, transformasi, pemodelan, evaluasi, dan interpretasi hasil. Kombinasi K-Means *Clustering* dan tahapan KDD memberikan pendekatan yang sistematis dalam mengidentifikasi pola dan struktur dalam data penjualan obat, sehingga memfasilitasi pemahaman yang lebih mendalam terhadap karakteristik penjualan di Apotek An-Naafi.

4. HASIL DAN PEMBAHASAN

Berikut ini rangkuman proses tahapan KDD (*Knowledge Discovery in Databases*) yang dilakukan dalam penelitian penerapan metode k-means *clustering* pada data penjualan obat Apotek An-Naafi:

4.1. Data Selection

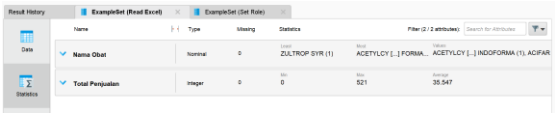
Data yang akan digunakan dalam penelitian ini adalah data penjualan obat di Apotek An-Naafi dalam kurun waktu tertentu 1 tahun terakhir sebanyak 1486 data. Data penjualan obat mencakup informasi seperti nama obat, kategori/jenis obat, jumlah terjual, harga jual, tanggal transaksi, dan lain-lain. Data diperoleh langsung dari sistem apotek An-Naafi, kemudian diekspor dan disiapkan untuk analisis. Data yang dipilih adalah data penjualan obat dalam rentang waktu 1 tahun terakhir yang lengkap dan akurat. Data yang tidak lengkap atau mengandung *noise* akan disaring terlebih dahulu sebelum analisis. Selanjutnya data akan di-*preprocess* dengan teknik seperti *handling missing values*, menstandarisasi format, dan lain-lain agar siap digunakan untuk analisis *cluster*. Atribut data yang akan digunakan antara lain jumlah terjual per obat, harga jual rata-rata, kategori obat, dan atribut lain yang relevan.

Tabel 1 *Data Sample* Penjualan Obat An-Naafi

No	Nama Obat	Penjualan	Pendapatan
1	CARBIDU 0,75 MG	521	3353880
2	TESPEK ONEMED	490	1297455
3	WIROS 20MG PIROXICAM KAPSUL	483	5015090
4	LACTO-B	438	3384440
5	ANDALAN BIRU PIL KB	387	5874123
6	VITACIMIN	381	668767
7	YUSIMOX SYR IFAR	369	1975964
8	TOLAK ANGIN BOX	361.5	4445570
....	.....	.....	.....
1485	ALBOTHYL OVULA	1	27000
1486	CITICOLINE 500 MG INFION	1	60000

4.2. Data Preprocessing

Tahap awal analisis data penjualan obat Apotek An-Naafi adalah melakukan *preprocessing* untuk membersihkan data sebelum proses data mining. *Preprocessing* yang dilakukan meliputi *cleaning* data, imputasi *missing values*, standarisasi format, integrasi data dari beberapa cabang, reduksi dimensi hingga pemilihan atribut penting, transformasi data, dan diskritisasi atribut numerik. Dengan *preprocessing* yang tepat, didapatkan data penjualan obat yang bersih dan siap dianalisis lebih lanjut menggunakan metode k-means *clustering*. Tahap ini penting agar diperoleh hasil *cluster* yang valid sesuai tujuan penelitian.

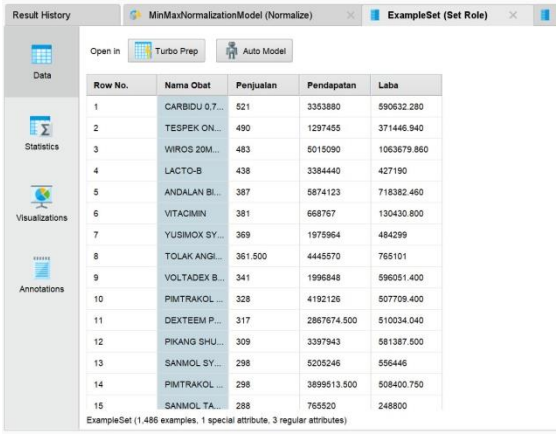


Gambar 2 Statistik *Dataset*

Gambar 2 Data penjualan obat Apotek An-Naafi telah melalui proses pembersihan data kosong dengan mengganti *error* dengan *missing values* yaitu dengan mengaktifkan *replace missing values* saat import ke *RapidMiner*. Setelah dilakukan pemeriksaan, tidak ditemukan adanya data kosong atau nilai yang hilang pada total 1486 data penjualan obat. Dengan demikian, data penjualan obat tersebut lolos tahap pembersihan data dan siap dilanjutkan ke proses *preprocessing* selanjutnya.

4.3. Data Transformation

Pada tahap ini dilakukan transformasi yaitu melakukan normalisasi.



Gambar 3 Data Hasil Normalisasi

Gambar 3 menampilkan data hasil transformasi yang telah dinormalisasi dengan metode min-max *normalization*. Normalisasi dilakukan pada atribut numerik dalam data penjualan obat, yaitu jumlah terjual dan harga jual rata-rata per obat. Nilai masing-masing atribut dinormalisasi ke rentang 0 hingga 1 agar semua atribut numerik memiliki skala yang sama. Hal ini dilakukan agar perbedaan skala antar atribut numerik tidak mempengaruhi proses *clustering*. Dengan normalisasi data, diharapkan hasil pengelompokan obat ke dalam *cluster* menjadi lebih akurat karena setiap atribut memberikan kontribusi yang seimbang dalam perhitungan jarak antar data.

4.4. Data Mining

Dalam proses *data mining*, dilakukan pengelompokan data penjualan obat Apotek An-Naafi menggunakan metode k-means *clustering*. Algoritma k-means *clustering* bekerja dengan mempartisi data ke dalam sejumlah *cluster* yang telah ditentukan sebelumnya (k *cluster*). Pembentukan *cluster* dilakukan melalui beberapa iterasi hingga diperoleh partisi optimum yang meminimalkan variasi dalam suatu *cluster* dan memaksimalkan variasi antar *cluster*.

Gambar 4 Proses *Rapidminer*

Gambar 4 menunjukkan proses *clustering* k-means yang dilakukan. Pertama, jumlah cluster optimum ditentukan dengan *Davies Bouldin Index (DBI)*, didapatkan 2 *cluster* sebagai jumlah optimal dengan *max* iterasi 5 kali. Kemudian, pusat *cluster* (*centroid*) awal diinisialisasi secara acak. Setiap data penjualan obat dikelompokkan ke *cluster* terdekat berdasarkan jarak *Euclidean* terhadap *centroid*. *Centroid cluster* diperbarui berdasarkan rata-rata data pada masing-masing *cluster*. Proses iterasi ini dilakukan hingga *centroid* tidak berubah lagi dan konvergen. Dengan metode k-means, data penjualan obat yang memiliki karakteristik serupa dikelompokkan ke dalam satu *cluster* yang sama.

Penentuan jumlah *cluster* optimum menggunakan *Davies-Bouldin Index* (DBI) melibatkan beberapa langkah. DBI mencoba mengukur seberapa baik cluster-cluster yang dihasilkan oleh algoritma k-means terpisah satu sama lain. Nilai DBI yang lebih rendah menunjukkan hasil clustering yang lebih baik. Berikut adalah langkah-langkahnya:

1. **Menjalankan K-Means untuk Berbagai Nilai K:** Pertama-tama, jalankan algoritma k-means untuk berbagai nilai K (jumlah *cluster*) yang mungkin. Inisialisasi pusat *cluster* dan iterasi k-means dilakukan untuk setiap nilai K.
2. **Menghitung Variabilitas *Intra-Cluster* (*Intra-Ci*):** Hitung variabilitas *intra-cluster* (Intra-Ci) untuk setiap *cluster*. Variabilitas ini dapat diukur dengan menggunakan metrik seperti varian atau jarak *Euclidean* antara titik data dalam suatu cluster dengan pusat *cluster*nya.
3. **Menghitung Keterpisahan Antar *Cluster* (dij):** Hitung keterpisahan antar *cluster* (dij) antara setiap pasangan *cluster*. Keterpisahan ini dapat diukur dengan menggunakan metrik seperti jarak *Euclidean* antara pusat *cluster*.
4. **Menghitung *Davies-Bouldin Index* (DBI):** DBI dihitung dengan rumus berikut:

$$\text{DBI} = \frac{1}{n} \sum_{i=1}^{n} \max(S_i, S_j) \quad (1)$$

Keterangan:
n : jumlah *cluster*
S_i : Rata-rata jarak data dengan pusat *cluster* pada cluster ke-i
S_j : Rata-rata jarak data dengan pusat *cluster* pada cluster ke-j.

DBI adalah rata-rata dari rasio antara variabilitas *intra-cluster* dengan keterpisahan antar cluster untuk setiap cluster. Semakin kecil nilai DBI, semakin baik kualitas clusteringnya.

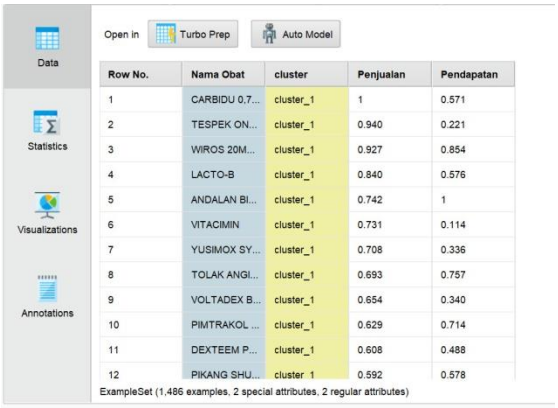
5. **Pemilihan Jumlah *Cluster* Optimum:** Pilih jumlah *cluster* yang memberikan nilai DBI terendah. Jumlah *cluster* yang menghasilkan DBI minimum dianggap sebagai jumlah *cluster* optimum.

Penggunaan *Euclidean distance* dalam perhitungan Variabilitas *Intra-Cluster (Intra-Ci)* pada langkah-langkah penentuan jumlah cluster optimum dengan *Davies-Bouldin Index* (DBI) sangat mempengaruhi pembentukan *cluster*. Saat menjalankan algoritma k-means untuk berbagai nilai K, perhitungan jarak *Euclidean* antara titik data dalam suatu *cluster* dengan pusat *clusternya* menjadi kunci untuk mengukur variabilitas *intra-cluster (Intra-Ci)*. Jarak *Euclidean* digunakan sebagai metrik untuk menentukan seberapa tersebar atau dekat titik-titik data dalam suatu *cluster* terhadap pusat *cluster*nya.

Variabilitas *intra-cluster* yang diukur menggunakan *Euclidean distance* ini memberikan gambaran tentang sejauh mana titik-titik data dalam suatu *cluster* memiliki kesamaan di antara mereka. Semakin kecil nilai variabilitas *intra-cluster*, semakin padat dan seragam titik-titik data dalam suatu *cluster*. Sebaliknya, nilai variabilitas *intra-cluster* yang besar menandakan bahwa titik-titik data dalam *cluster* tersebut lebih tersebar atau heterogen.

Selain itu, *Euclidean distance* juga digunakan dalam menghitung keterpisahan antar *cluster* (dij) pada langkah-langkah selanjutnya. Keterpisahan antar cluster diukur dengan jarak *Euclidean* antara pusat-pusat *cluster*. Hal ini membantu menentukan sejauh mana *cluster-cluster* berbeda satu sama lain dalam ruang fitur. Penggunaan *Euclidean distance* dalam perhitungan ini secara efektif memberikan informasi tentang seberapa terpisah atau berbeda *cluster-cluster* tersebut.

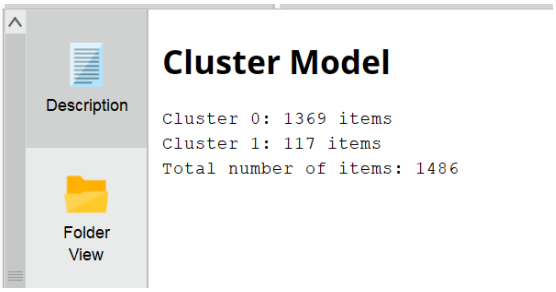
Dengan demikian, *Euclidean distance* tidak hanya menjadi alat pengukur variabilitas *intra-cluster*, tetapi juga menjadi metrik yang penting dalam evaluasi keterpisahan antar *cluster*. Kontribusi *Euclidean distance* dalam DBI membantu menghasilkan nilai indeks yang mencerminkan sejauh mana *cluster-cluster* terpisah satu sama lain, dan nilai DBI yang lebih rendah tetap menjadi indikasi clustering yang lebih baik.



Row No.	Nama Obat	cluster	Penjualan	Pendapatan
1	CARBIDU 0,7...	cluster_1	1	0.571
2	TESPEK ON...	cluster_1	0.940	0.221
3	WIROS 20M...	cluster_1	0.927	0.854
4	LACTO-B	cluster_1	0.840	0.576
5	ANDALAN BI...	cluster_1	0.742	1
6	VITACIMIN	cluster_1	0.731	0.114
7	YUSIMOX SY...	cluster_1	0.708	0.336
8	TOLAK ANGI...	cluster_1	0.693	0.757
9	VOLTADEX B...	cluster_1	0.654	0.340
10	PIMTRAKOL ...	cluster_1	0.629	0.714
11	DEXTEEM P...	cluster_1	0.608	0.488
12	PIKANG SHU...	cluster_1	0.592	0.578

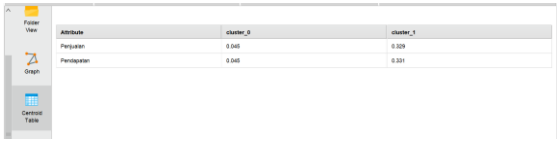
Gambar 5 Hasil *Clustering* K-Means

Gambar 5 menampilkan hasil *clustering* data penjualan obat Apotek An-Naafi menggunakan metode k-means dengan jumlah *cluster* sebanyak 2. Terlihat bahwa *cluster* 0 berisi obat dengan jumlah yaitu 1369 data dan *Cluster* 1 merupakan obat dengan berisi obat dengan jumlah yaitu 117 data.



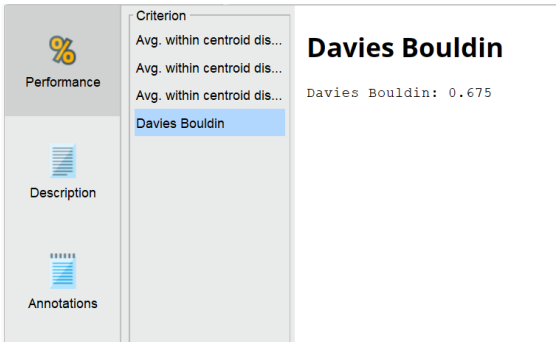
Gambar 6 *Cluster Model*

Gambar 6 menampilkan visualisasi hasil *clustering* data penjualan obat Apotek An-Naafi menggunakan model K-Means. Terlihat bahwa data penjualan obat telah terkelompok ke dalam 2 *cluster* yang direpresentasikan dengan warna yang berbeda. *Cluster* 1 berjumlah 1369 data dan *Cluster* 2 berjumlah 117 data dengan total data berjumlah 1486 data.



Gambar 7 Tabel *Centroid*

Gambar 7 menampilkan tabel *centroid* atau nilai rata-rata dari setiap atribut numerik pada masing-masing *cluster* hasil k-means *clustering*. Tabel *centroid* ini menjadi acuan untuk memahami karakteristik dan kedekatan antar data dalam suatu *cluster*.



Gambar 8 Nilai *Davies Buildin Index* Terbaik

Gambar 8 menampilkan nilai *Davies-Bouldin Index* (DBI) terbaik yang diperoleh dari proses *clustering* k-means pada data penjualan obat Apotek An-Naafi. Semakin rendah nilai DBI mengindikasikan bahwa *clustering* semakin baik karena jarak antar *cluster* lebih jauh dan jarak data dalam satu *cluster* lebih dekat. Dari gambar terlihat bahwa model k-means dengan 2 *cluster* menghasilkan DBI paling minimum yaitu 0.675.

Tabel 2 menyajikan hasil percobaan k-means *clustering* pada data penjualan obat Apotek An-Naafi dengan parameter jumlah *cluster* (k) yang divariasikan dari 2 hingga 5 *cluster*. Proses *clustering* dilakukan dengan batasan maksimum iterasi sebanyak 5 kali.

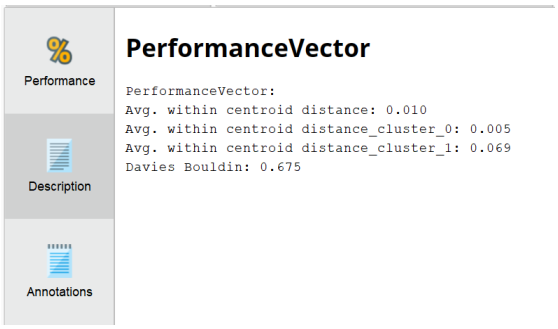
Tabel 2 Hasil Percobaan dengan parameter *Max* Iterasi = 5

K	***Avg. within centroid distance***	**DBI**	***Cluster***
2	0.010	0.675	*Cluster* 0: 1369 *items* *Cluster* 1: 117 *items*

3	0.006	0.712	*Cluster* 0: 1184 *items* *Cluster* 1: 258 *items* *Cluster* 2: 44 *items*
4	0.004	0.810	*Cluster* 0: 1127 *items* *Cluster* 1: 15 *items* *Cluster* 2: 56 *items* *Cluster* 3: 288 *items*
5	0.004	0.902	*Cluster* 0: 969 *items* *Cluster* 1: 36 *items* *Cluster* 2: 15 *items* *Cluster* 3: 340 *items* *Cluster* 4: 126 *items*

Terlihat bahwa semakin banyak jumlah *cluster*, rata-rata jarak antar *centroid cluster* semakin kecil. Ini menunjukkan *cluster* yang terbentuk semakin kompak. Namun, nilai DBI (*Davies-Bouldin Index*) justru semakin membesar, yang berarti pemisahan antar *cluster* semakin buruk. Nilai DBI terendah 0,675 didapatkan saat k=2 *cluster*. Ini mengindikasikan bahwa 2 *cluster* merupakan jumlah optimal untuk pengelompokan data penjualan obat, karena menghasilkan *cluster* yang kompak dan terpisah dengan baik. Dari tabel juga terlihat bahwa masing-masing *cluster* memiliki jumlah anggota yang seimbang pada k=2. Dengan demikian, Tabel 1 memberikan bukti bahwa 2 *cluster* adalah pilihan terbaik untuk pengelompokan data penjualan obat berdasarkan evaluasi jarak antar *centroid* dan nilai DBI optimum.

4.5. Interpretation/Evaluation



Gambar 9 Deskripsi *Performace Vector*

Performance vector pada proses *clustering* k-means berisi beberapa metrik yang menunjukkan kualitas *clustering* yang dilakukan. *Avg. within centroid distance dengan nilai* 0,010 mengindikasikan *cluster* cukup kompak secara keseluruhan. Nilai *Avg. within centroid distance* yang lebih kecil pada *cluster* 0 dibanding *cluster* 1 menunjukkan *cluster* 0 lebih kompak. Selain itu *Davies Bouldin index* bernilai 0,675 mengindikasikan pemisahan antar *cluster* cukup baik. Secara keseluruhan, metrik pada *performance vector* menunjukkan kualitas *clustering* yang baik dengan *cluster* yang kompak dan terpisah dengan baik. Nilai DBI sesuai dengan hasil optimasi sebelumnya yang menunjukkan 2 *cluster* optimal. Dengan demikian *performance vector* secara kuantitatif memvalidasi bahwa proses k-means *clustering* data penjualan obat telah berjalan dengan optimal. Anggota masing-masing kelompok pada k *cluster* dapat diketahui dengan melakukan perhitungan selisih rata-rata jarak *centroid*. Perhitungan ini ditunjukkan pada tabel di bawah yang menampilkan rata-rata jarak *centroid* untuk setiap nilai k yang diuji.

Tabel 3 Selisih *Centroid*

K-Means	0	1
Avg. within centroid distance	0.010	0.010
Avg. within centroid distance_cluster	0.005	0.069
Selisih *Centroid*	0.005	-0.59

Tabel 3 mencerminkan hasil evaluasi *clustering* menggunakan algoritma K-Means dengan dua *cluster* (0 dan 1). *Avg. within centroid distance* mengindikasikan bahwa titik data dalam suatu *cluster* cenderung berdekatan satu sama lain, dan *Avg. within centroid distance_cluster* memberikan gambaran tentang seberapa baik kedua *cluster* terpisah. Selisih *Centroid* merupakan perbedaan antara *centroid-cluster*, di mana nilai positif pada *Cluster* 0 menandakan bahwa *centroid Cluster* 0 lebih baik dari yang diharapkan, sementara nilai negatif pada *Cluster* 1 menunjukkan bahwa *centroid Cluster* 1 tidak optimal.

Tabel 4 Evaluasi *Cluster* Penjualan Obat Berdasarkan Jumlah Penjualan dan Pendapatan

Cluster	Jumlah Penjualan	Pendapatan	Tingkatan
0	381	Rp. 668.767	Rendah
1	521	Rp. 3.353.880	Tinggi

Dari hasil penelitian memberikan evaluasi *cluster* penjualan obat berdasarkan dua variabel utama, yaitu jumlah penjualan dan pendapatan, dengan dua *cluster* yang diidentifikasi sebagai *Cluster* 0 dan *Cluster* 1. *Cluster* 0 memiliki jumlah penjualan sebanyak 381 dengan pendapatan sekitar Rp. 668.767, dan secara umum dikategorikan dengan tingkatan "Rendah." Di sisi lain, *Cluster* 1 menunjukkan performa yang lebih tinggi, dengan jumlah penjualan sebanyak 521 dan pendapatan mencapai Rp. 3.353.880, dan diklasifikasikan sebagai tingkatan "Tinggi." Analisis ini memberikan gambaran tentang perbedaan dalam pola penjualan obat antara dua *cluster* tersebut, dengan *Cluster* 1 menonjol sebagai *cluster* dengan penjualan dan pendapatan yang lebih tinggi.

5. KESIMPULAN DAN SARAN

Hasil penelitian menunjukkan keberhasilan Metode K-Means *Clustering* dalam mengelompokkan data penjualan obat di Apotek An-Naafi. Pemilihan jumlah cluster optimal (K) dengan dua kelompok, berdasarkan *Davies-Bouldin Index* (DBI), menghasilkan segmentasi informatif. Penggunaan

Euclidean distance memengaruhi pembentukan cluster, dengan *Cluster* 1 menonjol dalam penjualan dan pendapatan tinggi, sedangkan *Cluster* 0 lebih beragam. *Cluster* 0 mencakup 381 penjualan dengan pendapatan sebesar Rp. 668.767, yang tergolong rendah. Sementara itu, *Cluster* 1 menunjukkan performa unggul dengan 521 penjualan dan pendapatan sebesar Rp. 3.353.880, yang dikategorikan sebagai tinggi. Temuan ini berkontribusi pada pemahaman pola penjualan dan memberikan peluang perbaikan, terutama dalam mengoptimalkan strategi pemasaran dan manajemen stok obat di Apotek An-Naafi. Saran untuk penelitian selanjutnya mencakup eksplorasi Algoritma K-Means dengan *measure type* lain, perbandingan dengan metode *clustering* alternatif, dan integrasi K-Means dengan teknik *clustering* lainnya untuk wawasan yang lebih komprehensif.

DAFTAR PUSTAKA

[1] G. Gustientiedina, M. H. Adiya, and Y. Desnelita, "Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan Pada RSUD Pekanbaru," *J. Nas. Teknol. dan Sist. Inf.*, vol. 5, no. 1, pp. 17–24, Apr. 2019, doi: 10.25077/TEKNOSI.v5i1.2019.17-24.

[2] L. ' Izzah and A. Jananto, "Penerapan Algoritma K-Means Clustering Untuk Perencanaan Kebutuhan Obat Di Klinik Citra Medika," *Progresif J. Ilm. Komput.*, vol. 18, no. 1, p. 69, Jan. 2022, doi: 10.35889/progresif.v18i1.769.

[3] R. W. Nasution, I. O. Kirana, I. Gunawan, and I. P. Sari, "Penerapan Data Mining Untuk Pengelompokan Minat Konsumen Terhadap Pengguna Jasa Pengiriman Pada PT . Jalur Nugraha Ekakurir (JNE) Pematangsiantar," *Resolusi*, vol. 1, no. 4, pp. 274–281, 2021.

[4] S. M. Hutabarat and A. Sindar, "Data Mining Penjualan Suku Cadang Sepeda Motor Menggunakan Algoritma K-Means," *J. Nas. Komputasi dan Teknol. Inf.*, vol. 2, no. 2, p. 126, 2019, doi: 10.32672/jnkti.v2i2.1555.

[5] B. I. Nugroho, N. P. Lestari, R. D. Kurniawan, and G. Gunawan, "Tinjauan Pustaka Sistematis: Data Mining Dalam Bidang Kesehatan," *J. Ekon. Teknol. dan Bisnis*, vol. 1, no. 1, pp. 14–27, 2022, doi: 10.57185/jetbis.v1i1.2.

[6] V. Herlinda, D. Darwis, and Dartono, "Analisis Clustering Untuk Recredesialing Fasilitas Kesehatan Menggunakan Metode Fuzzy C-Means," *JTSI J. Teknol. dan Sist. Inf.*, vol. 2, no. 2, pp. 94–99, 2021, [Online]. Available: http://jim.teknokrat.ac.id/index.php/JTSI

[7] H. Priyatman, F. Sajid, and D. Haldivany, "Klasterisasi Menggunakan Algoritma K-Means Clustering untuk Memprediksi Waktu Kelulusan Mahasiswa," *J. Edukasi dan Penelit. Inform.*, vol. 5, no. 1, p. 62, 2019, doi: 10.26418/jp.v5i1.29611.

[8] Sekar Setyaningtyas, B. Indarmawan Nugroho, and Z. Arif, "Tinjauan Pustaka Sistematis: Penerapan Data Mining Teknik Clustering Algoritma K-Means," *J. Teknoif Tek. Inform. Inst. Teknol. Padang*, vol. 10, no. 2, pp. 52–61, 2022, doi: 10.21063/jtif.2022.v10.2.52-61.

[9] M. Faid, M. Jasri, and T. Rahmawati, "Perbandingan Kinerja Tool Data Mining Weka dan Rapidminer Dalam Algoritma Klasifikasi," *Teknika*, vol. 8, no. 1, pp. 11–16, 2019, doi: 10.34148/teknika.v8i1.95.

[10] Y. R. Sari, A. Sudewa, D. A. Lestari, and T. I. Jaya, "Penerapan Algoritma K-Means Untuk Clustering Data Kemiskinan Provinsi Banten Menggunakan Rapidminer," *CESS (Journal Comput. Eng. Syst. Sci.*, vol. 5, no. 2, p. 192, 2020, doi: 10.24114/cess.v5i2.18519.