OPTIMASI ANALISIS DATA KEPUASAN PELANGGAN CV MEGA BAJA BINTARO DENGAN PENERAPAN ALGORITMA X-MEANS *CLUSTERING*

Royana Adniana, Dodi Solihudin, Riri Narasati
Teknik Informatika, STMIK IKMI Cirebon
Jalan Perjuangan No.10B, Karyamulya, Kec. Kesambi, Kota Cirebon, Jawa Barat 45135
royanaadniana5@gmail.com

ABSTRAK

Dalam menghadapi persaingan ketat di industri baja, CV Mega Baja Bintaro dituntut memahami preferensi dan perilaku konsumennya. Pemahaman mendalam tentang keragaman dan perubahan pola pelanggan penting untuk memberikan layanan terbaik dan mempertahankan loyalitas. Namun, keterbatasan dalam memahami dinamika konsumen kerap menjadi tantangan. Penelitian ini bertujuan menganalisis data kepuasan pelanggan di CV Mega Baja Bintaro dengan metode klasterisasi X-Means. Fokusnya adalah memanfaatkan X-Means untuk mengelompokkan data berdasarkan kesamaan karakteristik. X-Means mampu menentukan jumlah *cluster* secara otomatis berdasarkan pola data. Metode penelitian menerapkan algoritma X-Means pada data survei kepuasan pelanggan karena kemampuannya membagi *cluster* dengan *Bayesian Information Criterion* (BIC). Analisis *Davies Bouldin Index* (DBI) mengidentifikasi partisi *cluster* terbaik dan evaluasi karakteristiknya. Hasil penelitian diharapkan memberi wawasan tentang strategi peningkatan kepuasan pelanggan dan pengambilan keputusan oleh manajemen. Dari segi keilmuan, penelitian ini berkontribusi pada pengembangan metode X-Means. Hasil klasterisasi membentuk 19 *cluster* dengan tingkat kepuasan bervariasi dari 502 data survei.

Kata kunci: *Kepuasan pelanggan, Algoritma X-Means Clustering, Data mining, CV Mega Baja Bintaro, Analisis data pelanggan*

1. PENDAHULUAN

Perkembangan teknologi informasi dan transportasi yang pesat saat ini mendorong perusahaan untuk mengintegrasikan teknologi informasi guna mempertahankan daya saing [1]. CV Mega Baja Bintaro, distributor baja, menghadapi tantangan dalam menganalisis data kepuasan pelanggan untuk memahami perilaku konsumen dan menjaga loyalitas [2].

Untuk mengatasi situasi persaingan yang ketat, solusi yang potensial adalah penerapan data mining dan algoritma klasterisasi. Meskipun beberapa penelitian sebelumnya telah menggunakan teknik klasterisasi, keterbatasan metode masih menjadi kendala. Penelitian sebelumnya oleh [3] menggunakan Dimensi Servqual untuk mengukur kualitas pelayanan, dengan lima karakteristik utama. Sebaliknya, [4] menyatakan bahwa algoritma X-Means lebih unggul daripada K-Means dalam mengatasi keterbatasan pembelajaran tanpa pengawasan.

Dalam konteks lain, penelitian oleh [5] menunjukkan signifikansi penerapan algoritma X-Means dalam mengelompokkan tingkat kecanduan *games online*. Penelitian ini bertujuan menganalisis data kepuasan pelanggan CV Mega Baja Bintaro dengan menggunakan algoritma X-Means *Clustering*, yang menawarkan fleksibilitas dalam penentuan jumlah kelompok dan potensi mengatasi keterbatasan sebelumnya. Tujuan penelitian ini adalah membentuk *cluster* secara otomatis berdasarkan nilai *Bayesian Information Criterion* untuk menganalisis data kepuasan pelanggan. Hasil penelitian diharapkan dapat membantu perusahaan mengembangkan strategi peningkatan layanan dan loyalitas pelanggan, sementara secara akademis memberikan kontribusi pada pengembangan penerapan algoritma X-Means *Clustering* untuk analisis data konsumen. Pendekatan eksperimental digunakan dengan menerapkan X-Means *Clustering* pada data survei kepuasan dari 502 pelanggan CV Mega Baja Bintaro.

2. TINJAUAN PUSTAKA

2.1. Kepuasan Pelanggan

Penilaian kepuasan pelanggan setelah pembelian dianggap krusial dalam mendorong loyalitas pelanggan jika kebutuhan mereka terpenuhi [6]. Meningkatkan kualitas layanan perusahaan diidentifikasi sebagai strategi efektif untuk mencapai kepuasan pelanggan yang optimal [7]. Menurut penelitian oleh [3], alat yang digunakan untuk mengevaluasi kualitas layanan dan hubungannya dengan kepuasan pelanggan adalah Dimensi Servqual, yang mencakup lima karakteristik utama: 1) Fisik (*tangible*), melibatkan penampilan fisik staf yang rapi dan dukungan peralatan yang memadai untuk mendukung layanan, 2) Keandalan (*reliability*), berkaitan dengan kemampuan menangani dan menyelesaikan keluhan konsumen dengan akurat dan dapat diandalkan, 3) Responsif (*responsiveness*), melibatkan pemberian informasi kepada konsumen dengan jelas dan mudah dipahami, 4) Keamanan dan keyakinan (*assurance*), di mana karyawan mampu menumbuhkan kepercayaan dan keyakinan pelanggan terhadap janji atau perjanjian yang telah disepakati, dan 5) Kepedulian (*empathy*), yang mencakup pemberian perhatian atau pemahaman tentang masalah yang dihadapi oleh pelanggan.

2.2. *Data Mining*

Data mining, atau penambangan data, merujuk pada proses ekstraksi pola atau pengetahuan yang bermanfaat dari suatu kumpulan data. Pendekatan ini umumnya digunakan untuk mengidentifikasi contoh atau informasi menarik dalam dataset dengan menerapkan berbagai strategi. Sistem dan algoritma yang digunakan dalam data mining sangat beragam, dan pemilihan teknik yang optimal tergantung pada tujuan spesifik yang ingin dicapai serta tahapan dalam *Knowledge Discovery in Databases* (KDD). Dalam KDD, data mining berperan sebagai tahap penting di mana model atau pola yang relevan dihasilkan dari data yang telah dikumpulkan, membantu pengambilan keputusan dan pemahaman yang lebih baik terhadap informasi yang tersembunyi dalam dataset [8].

2.3. *X-Means*

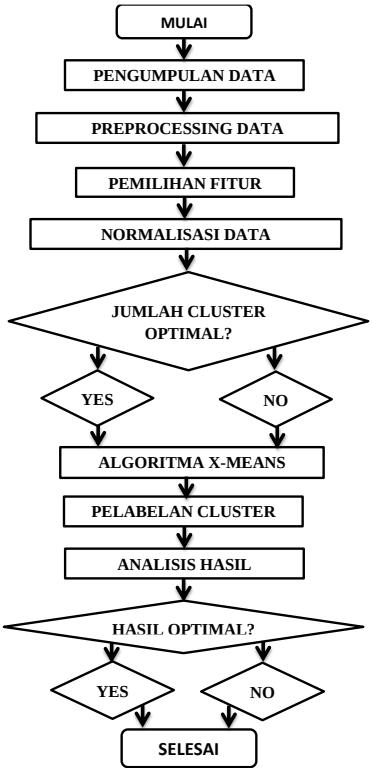
Pengertian X-Means adalah penyempurnaan dari metode K-Means dalam mengestimasi jumlah kelompok pada suatu data. K-Means telah berhasil mengurangi waktu yang diperlukan untuk proses klasterisasi, dan evaluasi terhadap pengelompokan K harus dilakukan secara serius oleh pengguna. X-Means termasuk dalam kategori pembelajaran tanpa pengawasan (*unsupervised learning*), di mana algoritma ini mampu mengelompokkan data tanpa memerlukan informasi sebelumnya mengenai kelas target. Peningkatan terhadap Harga *Bayesian Information Criterion* (BIC) terdapat pada metode X-Means. Secara dasar, estimasi harga BIC dari X-Means lebih bermanfaat daripada perhitungan K-Means, karena membutuhkan perhatian yang lebih intensif [4].

Berikut adalah langkah-langkah untuk mengimplementasikan Algoritma X-Means [9]:

1. Persiapkan data berdimensi-p dengan jumlah sampel n.
2. Tentukan jumlah kelompok awal k0 (biasanya 2) yang cukup sedikit.
3. Gunakan k-means pada seluruh data dengan k = k0. Beri nama pada kelompok klaster yang dihasilkan sebagai kelompok mendasar.
4. Iterasi Kan langkah-langkah 5 hingga 9 untuk setiap kelompok mendasar.
5. Terapkan k-implikasi pada setiap kelompok mendasar dengan k=2 untuk melakukan partisi. Beri nama pada sub-klaster yang terbentuk dan asumsikan distribusi normal p-dimensi data pada masing-masing sub-klaster.
6. Hitung batas-batas pengangkutan khas untuk setiap sub-klaster.
7. Jika nilai BIC split lebih besar dibandingkan sebelum dilakukan split, lanjutkan proses splitting klaster.
8. Jika nilai BIC split lebih kecil atau sama dengan nilai sebelum split, hentikan proses splitting klaster.
9. Setelah splitting selesai untuk satu kelompok mendasar, beri nomor unik pada setiap sub-klaster hasil split.
10. Ulangi proses splitting untuk semua kelompok mendasar hingga selesai, kemudian beri nomor unik pada semua klaster akhir.
11. Tampilkan detail dari semua klaster akhir beserta anggotanya.

2.4. *Flowchart* X-Means *Clustering*

Berikut penggambaran X-Means *Clustering* memakai Flowchart:



Gambar 1. Langkah-Langkah Umum dalam Pengelompokkan X-Means *Clustering*

Flowchart pada Gambar 1 ini mencerminkan langkah-langkah dalam proses analisis data dan pemilihan model *clustering* dengan menggunakan algoritma X-means. Berikut adalah penjelasan untuk setiap langkah:

1. Mulai
   Langkah awal dalam proses.
2. Pengumpulan Data
   Pada tahap ini, data dikumpulkan untuk analisis lebih lanjut.
3. Preprocessing Data
   Data yang dikumpulkan kemudian diproses untuk mengatasi masalah-masalah seperti *missing values*, *outliers*, atau data yang tidak sesuai.
4. Pemilihan Fitur
   Proses pemilihan fitur yang paling relevan dari dataset untuk digunakan dalam analisis selanjutnya.

5. Normalisasi Data
   Data dinormalisasi untuk memastikan bahwa semua fitur memiliki skala yang seragam, sehingga mempermudah dalam proses *clustering*.
6. Jumlah Cluster Optimal
   Tahap ini merupakan titik penting dalam proses *clustering*. Pertanyaan diajukan, apakah jumlah *cluster* yang optimal telah ditentukan atau belum.
   a. Yes: Jika jumlah cluster optimal sudah ditentukan sebelumnya, proses dilanjutkan ke tahap berikutnya.
   b. No: Jika belum, proses beralih ke tahap selanjutnya untuk menentukan jumlah *cluster* optimal menggunakan algoritma X-means.
7. Algoritma X-means
   Algoritma ini digunakan untuk menentukan jumlah *cluster* yang optimal secara otomatis.
8. Pelabelan Cluster
   Data kemudian diberi label sesuai dengan hasil *clustering* yang diperoleh dari algoritma X-means.
9. Analisis Hasil
   Dilakukan analisis terhadap hasil *clustering* untuk memahami pola atau karakteristik dari masing-masing *cluster*.
10. Hasil Optimal
    Pertanyaan diajukan lagi, apakah hasil *clustering* sudah optimal atau belum.
    a. Yes: Jika hasil *clustering* sudah optimal, proses dapat dianggap selesai.
    b. No: Jika belum optimal, mungkin perlu dilakukan tuning lebih lanjut atau pengulangan proses *clustering*.
11. Selesai
    Tahap akhir dari proses.

2.5. *Clustering*

Clustering adalah suatu teknik dalam analisis data yang bertujuan untuk mengelompokkan entitas data ke dalam sejumlah klaster berdasarkan kesamaan atau kemiripan karakteristik tertentu. Proses ini dilakukan melalui uji statistik dan metode pemisahan yang bertujuan agar data yang memiliki pola serupa dapat dikelompokkan bersama dalam satu klaster. Tujuan utama dari *clustering* adalah untuk merapihkan data yang awalnya tersebar dan sulit dipahami polanya, sehingga dapat diorganisir ke dalam struktur yang lebih teratur. Hal ini memungkinkan data yang memiliki kesamaan untuk dianalisis secara lebih mendalam, meningkatkan pemahaman, dan memberikan nilai tambah dalam konteks analisis data yang lebih lanjut [10].

2.6. *Euclidean Distance*

Euclidean distance merupakan suatu metode perhitungan yang digunakan untuk mengukur jarak antara dua titik dalam suatu ruang. Konsep ini menganalisis hubungan antara sudut dan jarak, menciptakan dasar matematis yang analog dengan perhitungan *Pythagoras* yang diterapkan pada pengukuran dua titik dalam satu dimensi. Proses penghitungan *Euclidean distance* dilakukan dengan menerapkan persamaan matematis yang terdefinisi sebagai Persamaan 1 [11].

$$dist_2(d_1, d_2) = \sqrt{\sum_{i-1}^{m} \left(d_1^i - d_2^i\right)^2} \quad (1)$$

Keterangan:
$dist_2(d_1, d_2)$ = *Euclidean distance* antara dua vektor *d*1 dan *d*2.
m = Jumlah dimensi atau komponen dalam vektor d_1 dan d_2.
i = Variabel indeks yang digunakan untuk merujuk ke setiap dimensi atau komponen vektor.
d_1^i = Komponen ke-*i* dari vektor d_1.
d_2^i: Komponen ke-*i* dari vektor d_2.
$\sum_{i-1}^{m}$ = Simbol sigma menunjukkan penjumlahan dari nilai-nilai dengan indeks *i* dari 1 hingga *m*.
$\left(d_1^i - d_2^i\right)^2$ = Selisih kuadrat antara komponen ke-*i* dari d_1 dan d_2.
$\sqrt{\sum_{i-1}^{m} \left(d_1^i - d_2^i\right)^2}$ = Akar kuadrat dari jumlah selisih kuadrat untuk setiap dimensi.

2.7. *Knowledge*

Penggalian informasi memiliki peran krusial dalam rangkaian Penggalian Pengetahuan dari Data (*Knowledge Discovery in Data*, KDD). KDD merupakan suatu metodologi yang digunakan untuk mengidentifikasi, menyaring, dan mengekstraksi informasi yang baru, lebih bermakna, serta lebih terstruktur dari sekumpulan data yang besar dan kompleks. Proses KDD melibatkan analisis menyeluruh yang mengintegrasikan konsep-konsep dari berbagai disiplin ilmu. Siklus KDD dimulai dengan penentuan tujuan dan diakhiri dengan evaluasi hasil yang telah ditemukan dari data [12].

3. METODE PENELITIAN

3.1. Jenis Penelitian

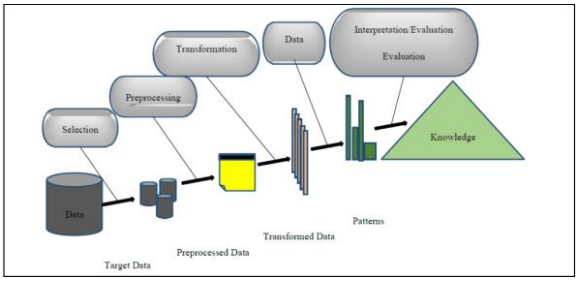
Penelitian ini bertujuan untuk mengevaluasi tingkat kepuasan pelanggan di CV Mega Baja Bintaro dengan menerapkan algoritma X-Mean *Clustering* pada analisis data kuantitatif. Metode penelitian ini mencakup pengumpulan data melalui survei terstruktur dan analisis statistik menggunakan algoritma X-Mean *Clustering* untuk mengidentifikasi pola kategori kepuasan pelanggan yang mungkin muncul.

3.2. Sumber Data

Terdapat 502 data kepuasan pelanggan yang digunakan dalam penelitian ini, mencakup periode tahun 2023. Sumber data berasal dari alamat Jl. Cemp. V No.6, RT.6/RW.11, Bintaro, Kec.

Pesanggrahan, Jakarta, Daerah Khusus Ibukota Jakarta 12330. Data tersebut menjadi landasan utama untuk menganalisis pola kepuasan pelanggan dan mengimplementasikan Metode X-Means *Clustering* guna mengelompokkan informasi tersebut.

3.3. Metode Penelitian



Gambar 2. Tahapan *Knowledge Discovery in Databases* (KDD)

Metodologi penelitian dalam proyek ini memanfaatkan teknik analisis data X-Means *Clustering* untuk mengelompokkan data kepuasan pelanggan di CV Mega Baja Bintaro. Penerapan X-Means *Clustering* bertujuan untuk secara adaptif dan otomatis mengenali pola serta struktur dalam data kepuasan pelanggan. Selain itu, penelitian ini menggunakan kerangka kerja *Knowledge Discovery in Databases* (KDD) yang holistik, melibatkan langkah-langkah seperti pengumpulan data, preprocessing, transformasi, pemodelan, evaluasi, dan interpretasi hasil. Integrasi X-Means *Clustering* dengan tahapan KDD memberikan pendekatan yang terstruktur untuk menggali wawasan mendalam terkait karakteristik kepuasan pelanggan di CV Mega Baja Bintaro. Tahapan pada KDD dilihat pada Gambar 1.

4. HASIL DAN PEMBAHASAN

Berikut adalah ringkasan dari langkah-langkah dalam proses *Knowledge Discovery in Databases* (KDD) yang dilakukan dalam penelitian mengenai penggunaan metode X-Means *Clustering* pada data kepuasan pelanggan CV Mega Baja Bintaro:

4.1. *Data Selection*

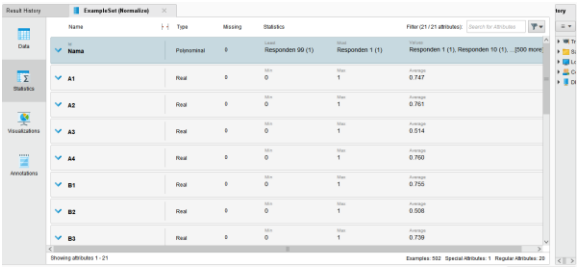
Data yang digunakan berasal dari survei kuesioner pada 502 pelanggan CV Mega Baja Bintaro tahun 2023. Kuesioner terdiri dari 5 kategori: Fisik, Keandalan, Responsif, Keamanan dan keyakinan, serta Kepedulian. Total 20 pertanyaan diajukan, dengan 4 pertanyaan untuk tiap kategori. Kuesioner bertujuan mengevaluasi pemahaman responden terkait kualitas layanan dan kepuasan pelanggan. Skala Likert digunakan untuk penilaian: (3) puas, (4) sangat puas, (2) cukup puas, (1) tidak puas.

Tabel .1 Dataset Kepuasan Pelanggan

Kode	Parameter	Atribut
A1	Berwujud (*tangible*)	Tata letak produk yang terpajang di rak.
A2		Kelengkapan produk yang dibutuhkan konsumen
A3		Kondisi kelayakan produk (bebas terhadap cacat)
A4		Kemudahan bagi pelanggan dalam menemukan produk yang mereka cari.
B1	Reliabilitas	Pemahaman karyawan mengenai produk layanan di CV Mega Baja Bintaro.
B2		Tarif umum yang dikenakan pada produk.
B3		Jumlah staf yang tersedia di bagian layanan kasir.
B4		Ketepatan Waktu Layanan (Ketepatan waktu layanan dalam mengirimkan produk ke konsumen)
C1	Ketanggapan (*responsiveness*)	Ekspresi rasa terima kasih yang disampaikan oleh karyawan pada akhir layanan.
C2		Respon terhadap keluhan (Kecepatan karyawan dalam menanggapi kebutuhan atau keluhan pelanggan).
C3		Tanggapan Terhadap Perubahan (Kemampuan karyawan perusahaan dalam merespon perubahan / perbaikan yang dikaukan)
C4		Ketersediaan Informasi (Kemudahan konsumen dalam mengakses informasi yang mereka butuhkan, Baik Berupa Brosur Ataupun Media Online)
D1	Jaminan dan kepastian (*assurance*)	Rasa aman dalam bertransaksi di CV Mega Baja Bintaro
D2		Privasi dan Keamanan Data (perusahaan menjaga privasi dan keamanan data konsumen dengan baik).
D3		Jaminan kelayakan produk dari bebas terhadap cacat (Jika ada cacat terhadap produk yang disebabkan oleh kelalaian pihak karyawan perusahaan maka produk akan diganti)
D4		Transparansi Harga (harga yang perusahaan tawarkan transparan dan sesuai dengan nilai produk atau layanan yang konsumen terima)
E1	Empati (*empathy*)	Perhatian personal yang diberikan oleh karyawan kepada pelanggan.
E2		Keterlibatan dan peran karyawan dalam memberikan bantuan kepada pelanggan dengan cepat ketika diperlukan.
E3		Kesopanan dan keramahan para karyawan.
E4		Keramahan petugas keamanan CV Mega Bintaro

4.2. *Data Preprocessing*

Pembersihan data merupakan langkah krusial yang harus dilakukan sebelum memasuki fase penambahan data saat membaca informasi dari Excel. Tujuan utama dari pembersihan ini adalah untuk mengatasi duplikasi data, memeriksa keberadaan nilai yang hilang, dan menangani kolom-kolom data yang kosong. Tindakan ini dilakukan dengan maksud agar hasil akhir pada tahap selanjutnya tidak terpengaruh secara negatif dalam hal akurasi.

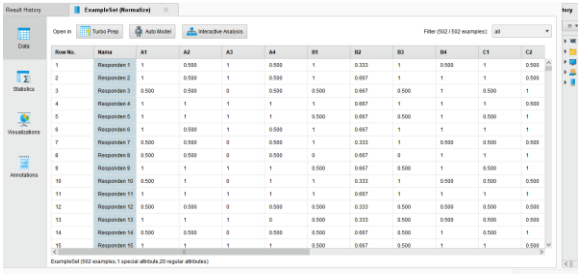


Gambar 3. *Result* dari statistik dataset

Gambar 3 menunjukkan hasil pembersihan data kosong (void) dengan mengaktifkan opsi *replace missing value* di *RapidMiner*. Metode ini berhasil mencegah nilai kosong pada dataset. Jumlah data kuesioner tetap 502, tanpa adanya perubahan atau hilangnya informasi. Dengan data yang lengkap dan tidak ada nilai kosong, proses preprocessing dapat dilanjutkan ke tahap berikutnya.

4.3. *Data Transformation*

Pada tahap transformasi data kepuasan pelanggan, dilakukan proses normalisasi.

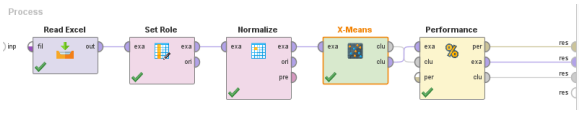


Gambar 4. Hasil Normalisasi

Gambar 4 menunjukkan normalisasi data menggunakan Operator *Normalize* di *RapidMiner*. Normalisasi berfungsi mengubah nilai numerik agar berada dalam rentang tertentu atau skala yang lebih standar. Hal ini penting agar perbedaan skala fitur tidak mempengaruhi analisis data. Metode yang digunakan adalah perubahan rentang dengan nilai minimum 0 dan maksimum 1.

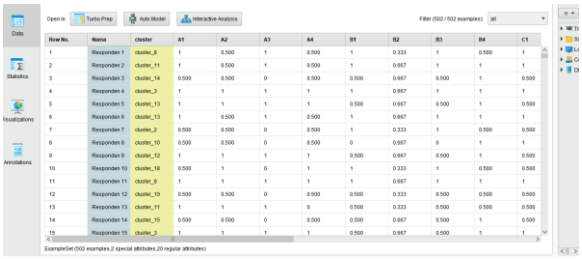
4.4. *Data Mining*

Dalam data mining, dilakukan pengelompokkan data kepuasan pelanggan CV Mega Baja Bintaro untuk mengekstraksi informasi berharga dan memahami pola perilaku pelanggan. Klasterisasi pelanggan penting agar perusahaan dapat menerapkan strategi pemasaran dan layanan yang spesifik untuk setiap segmen, meningkatkan efektivitas pelayanan, dan meningkatkan kepuasan pelanggan. Klasterisasi pelanggan juga memberikan wawasan tentang pelanggan setia, potensial, perlu perhatian, dan segmen lainnya, informasi ini penting untuk formulasi strategi bisnis dan pemasaran perusahaan.



Gambar 5. Model Proses X-Means *Clustering*

Gambar 5 menggambarkan model proses klasterisasi dengan menggunakan algoritma X-Means pada data kuesioner kepuasan pelanggan. X-Means adalah algoritma *clustering non-hierarkis* yang dapat secara otomatis menentukan jumlah *cluster* berdasarkan data. Dalam hasilnya, ditemukan bahwa terdapat 19 *cluster* dengan maksimal iterasi sebanyak 20 kali.

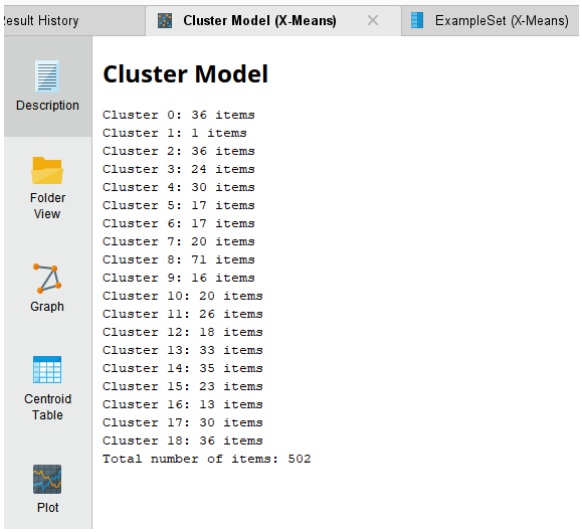


Gambar 6. Hasil X-Means *Clustering*

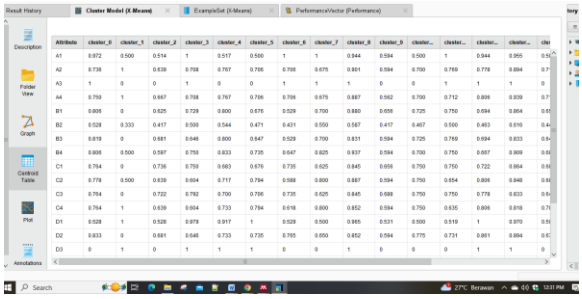
Gambar 6 menampilkan hasil pengelompokan terbaik dengan 19 *Cluster*. Jumlah partisipan dalam masing-masing *cluster* adalah sebagai berikut: *Cluster* 0 (36 anggota), *Cluster* 1 (1 anggota), *Cluster* 2 (36 anggota), *Cluster* 3 (24 anggota), *Cluster* 4 (30 anggota), *Cluster* 5 (17 anggota), *Cluster* 6 (17 anggota), *Cluster* 7 (20 anggota), *Cluster* 8 (71 anggota), *Cluster* 9 (16 anggota), *Cluster* 10 (20 anggota), *Cluster* 11 (26 anggota), *Cluster* 12 (18 anggota), *Cluster* 13 (33 anggota), *Cluster* 14 (35 anggota), *Cluster* 15 (23 anggota), *Cluster* 16 (13 anggota), *Cluster* 17 (30 anggota), dan *Cluster* 18 (36 anggota).

Gambar 7 mengilustrasikan hasil dari proses klasterisasi dalam bentuk Model *Cluster*, yang memvisualisasikan pengelompokan data ke dalam *cluster* atau kelompok yang serupa berdasarkan karakteristik tertentu.

Gambar 8 menampilkan Tabel Centroid, yang berisi informasi koordinat pusat klaster dalam proses klasterisasi. Tabel ini memainkan peran kunci dalam memahami distribusi dan representasi klaster-kelompok yang dihasilkan dari analisis data. Tabel 2 menampilkan hasil dari 20 percobaan variasi X-Means Clustering pada data kepuasan pelanggan CV Mega Baja Bintaro.



Gambar 7. *Cluster Model*



Gambar 8. Tabel *Centroid*

Tabel 2. Hasil Percobaan X-Means *Max Run* 20

No	Parameter		*Avg.vwithin centroidvdistance*	DBI
	K min	K max		
1	2	2	0.094	0.098
2	3	3	0.082	0.088
3	4	4	0.072	0.088
4	5	5	0.067	0.087
5	6	6	0.064	0.082
6	7	7	0.059	0.082
7	8	8	0.057	0.085
8	9	9	0.053	0.087
9	10	10	0.052	0.086
10	11	11	0.050	0.084
11	12	12	0.048	0.085
12	13	13	0.046	0.082
13	14	14	0.045	0.082
14	15	15	0.044	0.080
15	16	16	0.042	0.079
16	17	17	0.042	0.082
17	18	18	0.040	0.082
18	19	19	0.040	0.078
19	20	20	0.039	0.079

Setelah melakukan 20 percobaan dengan variasi nilai parameter *run max* dari k *min*=2 hingga k *max*=20, ditemukan bahwa nilai optimum terjadi pada k *min*=19. Pada konfigurasi tersebut, diperoleh nilai *Avg. within centroid distance* sebesar 0.040, *Davies Bouldin Index* (DBI) sebesar 0.078, dan terbentuk 19 *Cluster*.

Dalam konteks evaluasi *cluster*, *Davies Bouldin Index* (DBI) menjadi ukuran yang relevan untuk menilai kualitas *cluster*. DBI menggabungkan rata-rata jarak intra-*cluster* dan jarak antara *cluster* untuk mengevaluasi kehomogenan dalam *cluster* dan keberagaman antar *cluster*. *Davies-Bouldin Index* (DBI) merupakan salah satu metrik evaluasi *cluster* yang digunakan untuk mengukur kualitas suatu klasterisasi. Secara umum, DBI dihitung dengan rumus:

$$\text{DBI}=\frac{1}{n}\sum_{i=1}^{n}\max(S_i,S_j) \quad (2)$$

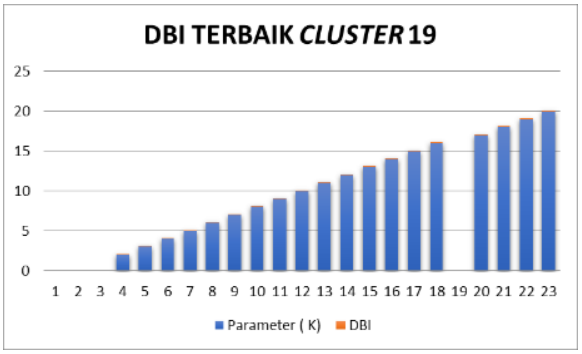
Keterangan:
n : jumlah *cluster*
S_i : Rata-rata jarak data dengan pusat *cluster* pada klaster ke-i
S_j : Rata-rata jarak data dengan pusat *cluster* pada klaster ke-j.

Dalam rumus tersebut, *Euclidean distance* memainkan peran penting. *Euclidean distance* digunakan untuk menghitung rata-rata jarak *intra-cluster* (*avg intra-cluster distance*) yang mencerminkan seberapa rapat titik-titik dalam suatu *cluster*. Juga, *Euclidean distance* digunakan untuk mengukur jarak antara *cluster* (*distance between cluster_ij*), yang mencerminkan seberapa terpisah *cluster* satu dengan yang lain.

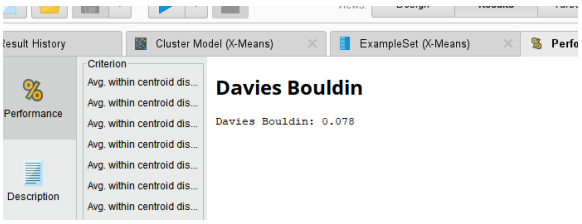
Penggunaan dalam konteks ini sangat relevan karena memberikan representasi geometris yang baik dari jarak antar titik dalam ruang berdimensi tinggi. Dalam hal ini, *Euclidean distance* memberikan ukuran yang intuitif tentang kedekatan atau kejauhan antar titik, yang diperlukan untuk menghitung homogenitas dan keberagaman dalam hasil klasterisasi. Oleh karena itu, nilai *Euclidean distance* memainkan peran kunci dalam mengukur kualitas *cluster* menggunakan *Davies Bouldin Index* (DBI).

Hasil evaluasi dengan DBI sebesar 0.078 menunjukkan bahwa konfigurasi *cluster* dengan k min=19 memberikan klasterisasi yang baik, dengan rata-rata jarak *intra-cluster* yang kecil dibandingkan dengan jarak antar *cluster*. Nilai DBI semakin kecil semakin baik, menunjukkan klasterisasi yang kompak dan terpisah dengan baik. Nilai terendah 0.078 pada 19 *cluster* menunjukkan bahwa jumlah *cluster* optimal untuk data tersebut adalah 19, menandakan kualitas klasterisasi yang optimal dengan menggunakan *Euclidean distance*.

Gambar 9 menunjukkan diagram *cluster* terbaik dari 20 percobaan, yang menghasilkan 19 *cluster* sebagai *cluster* optimal.



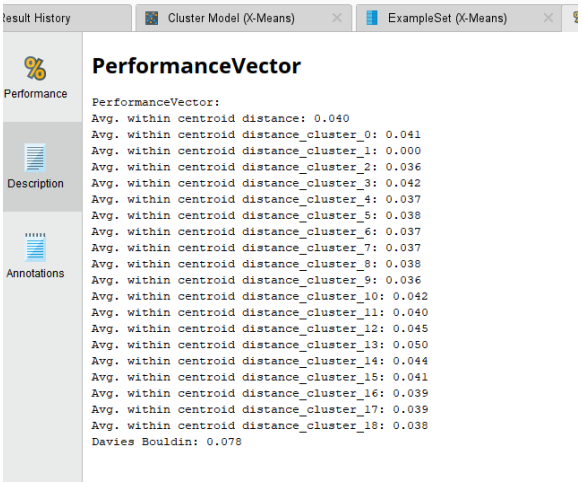
Gambar 9. Diagram *Cluster* Terbaik



Gambar 10. Hasil Davies Bouldin Terbaik

Gambar 10 menampilkan Hasil Davies Bouldin Terbaik, yang mencerminkan evaluasi kinerja klasterisasi dengan menggunakan indeks *Davies-Bouldin* untuk menentukan seberapa baik pemisahan dan kesamaan antar klaster dalam analisis data. Diperoleh nilai DBI terbaik sebesar 0.078.

4.5. *Interpretation/Evaluation*



Gambar 11. Performance Vector

Gambar 11 menyajikan hasil evaluasi klasterisasi kepuasan pelanggan CV Mega Baja Bintaro dengan X-Means *Clustering*. Nilai-nilai *Avg. within centroid distance* dan *Avg. within centroid distance_cluster*_X mencerminkan homogenitas *cluster*, dengan nilai yang lebih kecil menunjukkan klasterisasi yang lebih baik. Indeks *Davies Bouldin* (DBI) dengan nilai 0.078 menunjukkan efektivitas pemisahan dan keseragaman *cluster*. Dengan penjelasan terperinci untuk setiap *cluster* (*cluster* 0 hingga *cluster* 18), Performance Vector memberikan wawasan singkat tentang kesuksesan klasterisasi dan pemahaman pola data kepuasan pelanggan.

Dalam rangka eksperimen dengan variasi nilai k min dari 2 hingga k max 20, ditemukan bahwa nilai *Davies-Bouldin* mencapai optimum pada 0.078 saat k = 19 dalam perhitungan X-Means. Semakin rendah nilai indeksnya, hasil pengelompokannya menjadi semakin baik. Dari jarak *centroid* rata-rata, diferensial dapat dihitung untuk mengidentifikasi anggota kelompok berdasarkan nilai k, sebagaimana terlihat dalam tabel berikut:

Tabel 3. Perhitungan Selisih *Centroid*

Cluster	**Jumlah Anggota**	***Avg. within centroid distance***	***Avg.within centroid distance cluster***	**Selisih *Centroid***
Cluster 0	36	0.040	0.041	-0.001
Cluster 1	1	0.040	0.000	0.04
Cluster 2	36	0.040	0.036	0.004
Cluster 3	24	0.040	0.042	-0.002
Cluster 4	30	0.040	0.037	0.003
Cluster 5	17	0.040	0.038	0.002
Cluster 6	17	0.040	0.037	0.003
Cluster 7	20	0.040	0.037	0.003
Cluster 8	71	0.040	0.038	0.002
Cluster 9	16	0.040	0.036	0.004
Cluster 10	20	0.040	0.042	-0.002
Cluster 11	26	0.040	0.040	0
Cluster 12	18	0.040	0.045	-0.005
Cluster 13	33	0.040	0.050	-0.01
Cluster 14	35	0.040	0.044	-0.004
Cluster 15	23	0.040	0.041	-0.001
Cluster 16	13	0.040	0.039	0.001
Cluster 17	30	0.040	0.039	0.001
Cluster 18	36	0.040	0.038	0.002

Dari analisis selisih *centroid*, terlihat bahwa *cluster* 1 memiliki selisih centroid tertinggi sebesar 0.04, menunjukkan tingkat kepuasan pelanggan yang paling tinggi. Sebaliknya, *cluster* 13 dengan selisih terendah, yaitu -0.01, menandakan batas bawah tingkat kepuasan. Batas atas dan batas bawah awalnya ditetapkan oleh peneliti masing-masing sebesar 0.04 dan -0.01. Selanjutnya, dengan menghitung nilai tengah sebesar 0.02, ditentukan sebagai batas label tingkatan kepuasan. Label kepuasan pelanggan kemudian diatributkan berdasarkan perbandingan selisih *centroid cluster* dengan ketiga batas nilai tersebut, memungkinkan klasifikasi tingkat kepuasan ke dalam label yang sesuai pada setiap *cluster*.

Tabel 4. Rentan Label Berdasarkan Selisih *Centroid*

Label	**Batasan Label**
Sangat Puas	0.02 < Selisih *centroid* ≤ 0.04
Puas	0.01 ≤ Selisih *centroid* ≤ 0.02
Cukup Puas	-0.01 ≤ Selisih *centroid* < 0.01
Tidak Puas	Selisih *centroid* < -0.01

Tabel 4 mengkategorikan tingkat kepuasan pelanggan berdasarkan perbedaan *centroid* yang

dihasilkan dari analisis *cluster*. Label "Sangat Puas" diberikan untuk perbedaan 0.02 hingga kurang dari 0.04, "Puas" untuk perbedaan 0.01 hingga kurang dari 0.02, "Cukup Puas" untuk perbedaan -0.01 hingga kurang dari 0.01, dan "Tidak Puas" untuk perbedaan kurang dari -0.01. Panduan ini berguna dalam mengevaluasi tingkat kepuasan pelanggan di dalam *cluster*.

Tabel 5. Segmentasi Pelanggan Menggunakan X-Means

Cluster	**Jumlah Anggota**	**Tingkatan kepuasan**
Cluster 0	36	Cukup Puas
Cluster 1	1	Sangat Puas
Cluster2	36	Sangat Puas
Cluster 3	24	Cukup Puas
Cluster 4	30	Sangat Puas
Cluster 5	17	Sangat Puas
Cluster 6	17	Sangat Puas
Cluster 7	20	Sangat Puas
Cluster 8	71	Sangat Puas
Cluster 9	16	Sangat Puas
Cluster 10	20	Cukup Puas
Cluster 11	26	Cukup Puas
Cluster 12	18	Cukup Puas
Clusteri 13	33	Tidak Puas
Cluster 14	35	Cukup Puas
Cluster 15	23	Cukup Puas
Cluster 16	13	Sangat Puas
Cluster 17	30	Sangat Puas
Cluster 18	36	Sangat Puas

Dari hasil studi, ditemukan bahwa terdapat 19 *cluster*. *cluster* 0, yang terdiri dari 36 item, mewakili pelanggan dengan tingkat kepuasan "Cukup Puas". *cluster* 1 hanya memiliki 1 item dengan pelanggan yang berada dalam kategori "Sangat Puas". *cluster* 2 (36 item) dan *cluster* 4 (30 item) juga melibatkan pelanggan dengan tingkat kepuasan "Sangat Puas". *cluster* 3 (24 item) dan *cluster* 13 (33 item) masing-masing mencakup pelanggan dengan tingkat kepuasan "Cukup Puas" dan "Tidak Puas". *cluster* 5, 6, 7, 9, dan 16 masing-masing terdiri dari 17, 17, 20, 16, dan 13 item dengan pelanggan yang merasakan kepuasan yang tinggi ("Sangat Puas"). *cluster* 8 (71 item) mendominasi dengan pelanggan yang memiliki tingkat kepuasan "Sangat Puas". *cluster* 10 (20 item) dan *cluster* 11 (26 item) masing-masing mencakup pelanggan dengan tingkat kepuasan "Cukup Puas". *cluster* 12 (18 item), *cluster* 14 (35 item), dan *cluster* 15 (23 item) juga mencakup pelanggan dengan tingkat kepuasan "Cukup Puas". Sementara itu, *cluster* 17 (30 item) dan *cluster* 18 (36 item) juga melibatkan pelanggan dengan tingkat kepuasan "Sangat Puas".

5. KESIMPULAN DAN SARAN

Berdasarkan hasil studi peneliti pada CV Mega Baja Bintaro, dapat disimpulkan bahwa analisis data kepuasan pelanggan menggunakan Algoritma X-Means Clustering dengan *Tools RapidMiner* menghasilkan 19 kelompok optimal. Nilai *Davies-Bouldin Index* (DBI) yang diperoleh sebesar 0.078. Pengelompokkan ini mencerminkan variasi tingkat kepuasan pelanggan dalam berbagai *Cluster* dengan total 502 dataset. Klaster-klasternya terbagi ke dalam *Cluster* 0 (Cukup Puas) dengan 36 anggota, *Cluster* 1 (Sangat Puas) dengan 1 anggota, *Cluster* 2 (Sangat Puas) dengan 36 anggota, *Cluster* 3 (Cukup Puas) dengan 24 anggota, *Cluster* 4 (Sangat Puas) dengan 30 anggota, *Cluster* 5 (Sangat Puas) dengan 17 anggota, *Cluster* 6 (Sangat Puas) dengan 17 anggota, *Cluster* 7 (Sangat Puas) dengan 20 anggota, *Cluster* 8 (Sangat Puas) dengan 71 anggota, *Cluster* 9 (Sangat Puas) dengan 16 anggota, *Cluster* 10 (Cukup Puas) dengan 20 anggota, *Cluster* 11 (Cukup Puas) dengan 26 anggota, *Cluster* 12 (Cukup Puas) dengan 18 anggota, *Cluster* 13 (Tidak Puas) dengan 33 anggota, *Cluster* 14 (Cukup Puas) dengan 35 anggota, *Cluster* 15 (Cukup Puas) dengan 23 anggota, *Cluster* 16 (Sangat Puas) dengan 13 anggota, *Cluster* 17 (Sangat Puas) dengan 30 anggota, dan *Cluster* 18 (Sangat Puas) dengan 36 anggota.

Adapun saran yang dapat diajukan untuk penelitian selanjutnya adalah pertama, mengembangkan Algoritma X-Means *Clustering* dengan mempertimbangkan perbandingan kelompok terbaik menggunakan 2 atau lebih *measure tape* selain dari *Euclidean Distance*. Kedua, penelitian dapat diperluas dengan membandingkan *output* segmentasi menggunakan metode *Clustering* lain atau dengan menggabungkan metode X-Means dengan teknik *Clustering* alternatif lainnya. Peningkatan dalam hal ini diharapkan dapat memberikan pemahaman yang lebih mendalam mengenai faktor-faktor yang mempengaruhi kepuasan pelanggan di CV Mega Baja Bintaro.

DAFTAR PUSTAKA

[1] N. D. Sari, “Penerapan kelasifikasi kepuasan pelanggan go-jek menggunakan metode algoritma naïve bayes,” p. 60, 2018.

[2] H. Indrayani, “Penerapan Teknologi Informasi Dalam Peningkatan Efektivitas, Efisiensi Dan Produktivitas Perusahaan Oleh : Henni Indrayani Abstraksi,” *J. El-Riyasah*, vol. 3, no. 1, pp. 48–56, 2017.

[3] S. Z. M. Lilik Trianah, Diah Pranitasari, “Pengaruh Kualitas Layanan, Produk Dan Nilai Pelanggan Terhadap Kepuasan Pelanggan Dan Loyalitas Pelanggan,” *J. Bina Bangsa Ekon.*, vol. 14, no. 1, pp. 230–239, 2021, doi: 10.46306/jbbe.v14i1.79.

[4] A. Wijayanto, “Penggunaan X-Means Clustering Method untuk Mengelompokkan Potensi Sekolah Menengah Unggul di Kabupaten Banyumas,” *J. Informatics, Inf. Syst. Softw. Eng. Appl.*, vol. 2, no. 1, pp. 80–88, 2019, doi: 10.20895/inista.v2i1.99.

[5] Doli Fancius, Elvia Budianita, Iwan Iskandar,

and Fadhilah Syafria, "Tingkat Kecanduan Game Online Menggunakan Algoritma X-Means Clustering," pp. 135–145, 2022.

[6] I. K. Tampanguma, J. A. F. Kalangi, and O. Walangitan, "Pengaruh Kualitas Pelayanan Terhadap Kepuasan Pelanggan Rumah Es Miangas Bahu Kota Manado," *Productivivty*, vol. 3, no. 1, pp. 7–12, 2022.

[7] T. Ismail and R. Yusuf, "Pengaruh Kualitas Pelayanan Terhadap Kepuasan Pelanggan Kantor Indihome Gegerkalong Di Kota Bandung," *J. Ilm. MEA (Manajemen, Ekon. dan Akuntansi)*, vol. 5, no. 3, pp. 413–423, 2021.

[8] Y. Mardi, "Data Mining : Klasifikasi Menggunakan Algoritma C4.5," *Edik Inform.*, vol. 2, no. 2, pp. 213–219, 2017, doi: 10.22202/ei.2016.v2i2.1465.

[9] S. Anwar, N. D. Nuris, and Y. A. Wijaya, "Pengelompokkan Tingkat Pemahaman Kurikulum Berbasis KKNI Menggunakan Metode X-Means Clustering," *J. Inform. J. Pengemb. IT*, vol. 4, no. 2–2, pp. 187–190, 2019, doi: 10.30591/jpit.v4i2-2.1869.

[10] D. P. Tri Ananda Widyaningsih, "Penerapan Metode K-Means Clustering Dalam Mengelompokkan Jumlah Peserta Bpjs Kesehatan Jkn/Kis Di Kabupaten Cirebon," *E-Link J. Tek. Elektro dan Inform.*, vol. 18, no. 1, p. 17, 2023, doi: 10.30587/e-link.v18i1.5330.

[11] M. Mughnyanti and S. Hafiz Nanda Ginting, "Data Mining Manhattan Distance dan Euclidean Distance Pada Algoritma X-Means Dalam Klasifikasi Minat dan Bakat Siswa," *Remik*, vol. 7, no. 1, pp. 835–842, 2023, doi: 10.33395/remik.v7i1.12162.

[12] S. Widaningsih, "Perbandingan Metode Data Mining Untuk Prediksi Nilai Dan Waktu Kelulusan Mahasiswa Prodi Teknik Informatika Dengan Algoritma C4,5, Naïve Bayes, Knn Dan Svm," *J. Tekno Insentif*, vol. 13, no. 1, pp. 16–25, 2019, doi: 10.36787/jti.v13i1.78.